Human Disease Variation in the Light of Population Genomics

Ana Prohaska,[1,2,10] Fernando Racimo,[2,10] Andrew J. Schork,[2,3,4,10] Martin Sikora,[2] Aaron J. Stern,[5,6,11] Melissa Ilardo,[5,6,7,11] Morten Erik Allentoft,[2] Lasse Folkersen,[2,3,4] Alfonso Buil,[2,3,4] J. Víctor Moreno-Mayar,[2] Thorfinn Korneliussen,[2] Daniel Geschwind,[2,8] Andrés Ingason,[2,3,4] Thomas Werge,[2,3,4,9,*] Rasmus Nielsen,[2,5,6,*] and Eske Willerslev[1,2,*]

[1]Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK
[2]Lundbeck Foundation GeoGenetics Centre, Natural History Museum of Denmark, University of Copenhagen, 1350 Copenhagen, Denmark
[3]Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Copenhagen, Denmark
[4]Institute of Biological Psychiatry, Mental Health Services Capital Region of Denmark, Copenhagen, Denmark
[5]Department of Integrative Biology, UC Berkeley, Berkeley, CA 94720, USA
[6]Department of Statistics, UC Berkeley, Berkeley, CA 94720, USA
[7]The University of Utah Molecular Medicine Program, University of Utah, Salt Lake City, UT 84132, USA
[8]Brain Research Institute, University of California, Los Angeles, Los Angeles, CA 90095, USA
[9]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
[10]These authors contributed equally
[11]These authors contributed equally
*Correspondence: Thomas.Werge@regionh.dk (T.W.), rasmus_nielsen@berkeley.edu (R.N.), ewillerslev@snm.ku.dk (E.W.)
https://doi.org/10.1016/j.cell.2019.01.052

Identifying the causes of similarities and differences in genetic disease prevalence among humans is central to understanding disease etiology. While present-day humans are not strongly differentiated, vast amounts of genomic data now make it possible to study subtle patterns of genetic variation. This allows us to trace our genomic history thousands of years into the past and its implications for the distribution of disease-associated variants today. Genomic analyses have shown that demographic processes shaped the distribution and frequency of disease-associated variants over time. Furthermore, local adaptation to new environmental conditions—including pathogens—has generated strong patterns of differentiation at particular loci. Researchers are also beginning to uncover the genetic architecture of complex diseases, affected by many variants of small effect. The field of population genomics thus holds great potential for providing further insights into the evolution of human disease.

Introduction

Disease prevalence in humans varies considerably across the globe. The reasons for this are numerous and complex, from social (e.g., varying health-related infrastructure and medical practices), environmental (e.g., locally to regionally specific environmental risks), and genetic factors (e.g., spatial patterning of genetic variants) to their intricate interplay over time. For example, country-specific social mores and legislative regulations surrounding smoking behavior have large influences on the prevalence of lung cancer (Jha and Peto, 2014; GBD 2015 Tobacco Collaborators, 2017), which also has an important genetic component (Bossé and Amos, 2018). Another example is the prevalence of malaria in Africa, which depends on the prevalence of HBB gene variants (Ackerman et al., 2005) but also on the climate-driven geographical distribution of pathogen-carrying mosquitoes and on access to protective measures, such as mosquito nets or chemoprevention (World Health Organization, 2017). Additionally, environmental factors themselves are, in some cases, causes of particular patterns of diversity in the human genome, which can, in turn, affect the prevalence of disease. Recently, researchers have begun integrating medical and evolutionary approaches to study the trajectories of genetic variants over time and space and exploring ways in which population genomics can complement the fields of statistical genetics, molecular biology, physiology, psychology, and anthropology in elucidating the origin and evolution of human disease.

Human groups around the world share a large number of genetic variants, and most common variants found in a specific region are typically found in other regions. Genetic measures of differentiation and population structure are quite low when computed among human populations, and vastly larger amounts of genetic variation in our species are due to differences between individuals within populations than between populations (Elhaik, 2012). For example, genomes obtained from present-day humans in different corners of the world are as differentiated from each other as genomes obtained from two African chimpanzee populations living only a few kilometers apart (Prado-Martinez et al., 2013). This can be explained by the fact that most of our collective ancestry can be traced to a metapopulation that was confined to a single continent until only 60–100 thousand years ago (kya) (Malaspinas et al., 2016; Mallick et al., 2016; Pagani et al., 2016) and that groups of humans are highly mobile and have been continuously splitting and merging with each other (Pickrell and Reich, 2014). Nonetheless, these past processes
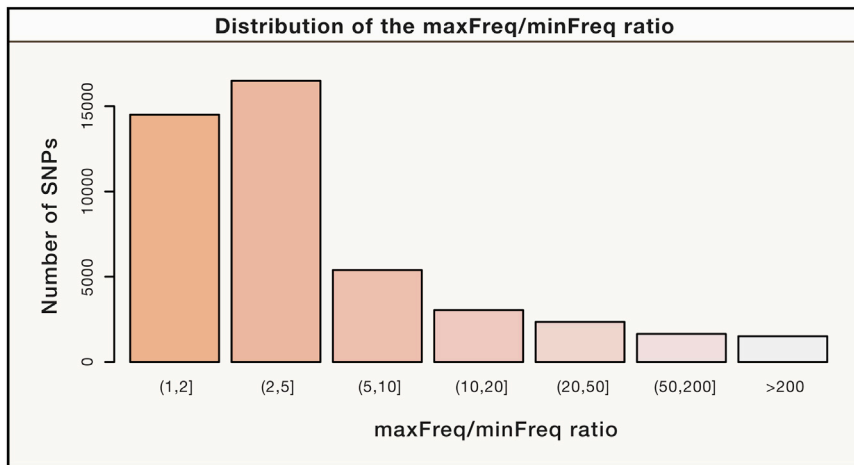
Distribution of the maxFreq/minFreq ratio

Figure 1. Ratio of Allele Frequencies for Disease-Associated SNPs
Distribution of the ratio of maximum and minimum allele frequency for all the SNPs with minor allele frequency larger than 0.05 in any population associated with a trait in the NIH GWAS catalog (https://www.ebi.ac.uk/gwas/). Allele frequencies are taken from five major population panels of the 1000 Genomes Project (African, East Asian, South Asian, European, and American). Each bar represents a range of ratios between maximum and minimum allele frequencies of a SNP. The height of the bar is the number of SNPs in that range.

of divergence and admixture have left subtle patterns of variation in our genomes that we can now detect thanks to the vast amounts of sequencing data generated by the genomic revolution.

The number of present-day human genomes sequenced to date has risen to the hundreds of thousands, providing catalogs of human genetic variability. Studies such as the 1000 Genomes Project (1000 Genomes Project Consortium, 2015) have shown that many common variants within a population are also observed across the globe, although subtle clines in allele frequencies are observed both within and across geographical divides. In addition, over 1,000 ancient genome sequences have been obtained from human remains, such as hair, teeth, or bones. These ancient genomes provide clear anchor points to discern the state of local variation in a population before and after it expanded, diverged, or intermixed with other populations. The joint analysis of ancient and present-day genomes has enabled unprecedentedly detailed reconstruction of the evolutionary history of our species and its implications for human health and disease today. Specifically, it has allowed the investigation of processes driving genetic differentiation at particular loci (including disease-associated variants) among populations, such as genetic drift and natural selection.

The genomic revolution has also been instrumental in uncovering the genetic basis of complex human diseases (Visscher et al., 2017), including metabolic and psychiatric diseases (Atanasovska et al., 2015; Geschwind and Flint, 2015). Over the last decade, researchers have identified thousands of loci that are significantly associated with complex traits, which may be causal or linked to causal variants that are important for health and disease (Visscher et al., 2017). Large-scale genome-wide association studies (GWASs) regularly aggregate genomic data from tens to hundreds of thousands of individuals in an attempt to associate the genotypes of millions of variants with disease. Arguably, the most important trend emerging from these studies is that most complex disease traits are highly polygenic, affected by many thousands of variants with individually small effects (Visscher et al., 2017), often only indirectly related to core disease pathways (Boyle et al., 2017), and that trait-associated variants naturally vary in frequency across the globe—due to

both neutral and selective processes (Figure 1). Thus, allele-frequency differences contribute, along with environmental and social differences, to varying prevalence of human diseases around the world. This poses important challenges, but also opportunities, for the study of the genetic basis of disease.

In this Review, we cover advances in our understanding of the impacts of neutral demographic processes, like bottlenecks, migrations, and admixture, and of adaptive processes, such as selective sweeps and polygenic adaptation, on human health and disease. We also suggest possible approaches for integrating evolutionary genomics and medical genomics and discuss the implications of this integration for future research practices.

Neutral Genetic Processes
Genetic Drift
Genomic studies of human populations have shown that "genetic drift" has played an important role in the distribution of human genomic variation, through demographic events such as population bottlenecks, divergence, and isolation (Ramachandran et al., 2005; Coop et al., 2009). Genetic drift is the random change in allele frequency over time, caused by the stochastic effects of reproduction and segregation of alleles in populations of finite size. Over time, this process leads to a loss of genetic diversity as alleles move to low or high frequencies and are eventually lost or fixed in the population. At the same time, new alleles are introduced into the population by mutation.

In population genetic models, both the number of new mutations introduced into the population and the rate of loss or fixation of existing alleles depend on a parameter called the "effective population size," which is related to the number of breeding individuals in a population at a particular time, under idealized conditions (see Ewens [2012]). The amount of genetic variability observed in a population, therefore, is directly linked to its demographic history.

Strong bottlenecks in the history of a population lead to a reduction in genetic diversity as well as increased relatedness between individuals. Any disease-associated alleles that were common in the founder individuals will likely be found at high frequency in their descendants. The impact of this process is particularly important for recessive disorders: an increase in the frequency of previously rare disease alleles due to a bottleneck can strongly increase the probability of homozygosity. As a consequence, populations that underwent recent bottlenecks

such as the Finnish, Amish, or Quebecois exhibit a high prevalence of certain autosomal recessive diseases (Payne et al., 2011; Scriver, 2001). For example, in a recent study of 5,685 Ashkenazi Jewish exomes, Rivas et al. (2018) reported a higher prevalence of alleles responsible for rare Mendelian disorders such as Tay-Sachs disease (HEXA gene) or Gaucher disease (GBA gene) relative to other populations. This is likely a consequence of the strong bottleneck in the recent history of Ashkenazi Jews, which has been estimated to be as low as a mere ~350 individuals in terms of effective population size (Carmi et al., 2014).

Greenlandic Inuit provide another example of the impact of an extreme bottleneck on patterns of genetic diversity. By analyzing exome data from 18 individuals, Pedersen et al. (2017) inferred a pronounced population reduction affecting the Greenlandic Inuit over the past ~15,000 years. As a consequence, alleles that are rare in European or East Asian populations were found to be common in Greenlandic Inuit. These included loss-of-function variants that disrupt protein function and have potentially large phenotypic impacts. The authors identified a total of 20 such variants, including previously described examples such as a variant in the gene TBC1D4 increasing the risk of type 2 diabetes as well as an indel in the gene SEMA4C involved in sucrase-isomaltase deficiency.

The aforementioned examples highlight the utility of founder populations for mapping both Mendelian and complex (polygenic) traits (Peltonen et al., 2000; Southam et al., 2017; Xue et al., 2017). In an elegant study, Belbin et al. (2017) demonstrated that with the help of population genomic data, such founder populations can even be identified in a large urban area. Using a dataset from a multiethnic biobank in New York City, they used patterns of haplotype sharing to identify a cohort of individuals of Puerto Rican descent with a signature of a recent bottleneck. Analyzing phenotypic data from electronic health records, they then identified a missense variant in a collagen gene (COL27A1) associated with extreme short stature, affecting up to 2% of people of Puerto Rican descent. Further population genetic analyses revealed that this variant is largely absent in other populations and likely arose in an ancestral Native American population. A strong bottleneck in the recent past, likely associated with European colonization of the island, then caused the rise in frequency of the allele, highlighting the potential impact of genetic drift on the distribution of disease-causing mutations.

Ancient genomes recovered from human remains have greatly increased our understanding of the patterning of genetic diversity in human prehistory. Although the spatiotemporal sampling of ancient individuals is skewed toward northern latitudes and more recent times, the data nevertheless suggest small populations affected by high levels of genetic drift were common in early humans across Eurasia. Genomic data from Stone Age foragers have consistently shown low levels of genetic diversity compared to present-day populations, as expected for small, mobile societies affected by repeated bottlenecks (Allentoft et al., 2015; Sikora et al., 2017; Skoglund et al., 2014). As a consequence, genetic differentiation (measured by Wright's fixation index, F_{ST}) between different Stone Age groups in Europe has been estimated to range from 0.05 to 0.08, which is more

than five times higher than the value observed among present-day European populations (F_{ST} ~0.01) (Allentoft et al., 2015; Skoglund et al., 2014). Interestingly, a high frequency of developmental anomalies has been observed in the archaeological record of Paleolithic human remains (Trinkaus, 2018), raising the possibility that higher prevalence of genetic disorders, as observed in present-day founder populations, might have been a common occurrence among prehistoric groups. To what extent the repeated strong bottlenecks in Eurasian prehistory have shaped the prevalence of alleles associated with health and disease in contemporary populations remains an open question and a fruitful avenue for future research.

Admixture

"Admixture between populations," by which previously isolated and diverged populations exchange genetic material, is another key historical process shaping human genetic diversity (Winkler et al., 2010). The general outcome of this process is a homogenization of the gene pools of the affected populations. This leads to lower differentiation, counteracting the effects of isolation and genetic drift. The myriad recent DNA studies on ancient humans have documented that migration and admixture were pervasive throughout human history (Nielsen et al., 2017; Skoglund and Mathieson, 2018). Overall, these studies highlight that the present-day occupants of a particular region are often poor representatives of the genetic diversity present just a few thousand years prior (Pickrell and Reich, 2014).

An example is the recent population history of western Eurasia, for which at least two major events of admixture and genetic turnover over the last 10,000 years were evinced from ancient genomes. The first one occurred ~8,000 years ago, when early farmers migrated north from the Near East during the Neolithic transition, replacing and admixing with the indigenous hunter-gatherer groups that occupied Europe previously (Sikora et al., 2014; Skoglund et al., 2012). Later on during the Bronze Age, large-scale migrations of pastoralists from the Eurasian steppe likely introduced Indo-European languages into Europe (Allentoft et al., 2015; Haak et al., 2015). Subsequent gene flow led to a homogenization of the European gene pool, eventually producing the comparably low amounts of genetic differentiation observed among present-day Europeans.

A similar picture has emerged from recent studies on the prehistory of Southeast Asia, where ancient genomes document a complex history involving a minimum of four genetically distinct ancestral groups (McColl et al., 2018). Much of the region was, until around 4,000 years ago, occupied by Hoabinhian hunter-gatherers, whose ancestry is today shared with only a few geographically isolated groups (e.g., Andamanese Onge). Later waves of migration, including those associated with the expansion of rice and millet farming, had a similarly profound impact on the genetic structure of human populations as those taking place in western Eurasia. It is only by around 2,000 years ago that the ancient genetic variation starts to resemble what is found today in the region.

The pervasiveness of admixture in human history has also had a profound effect on the distribution of trait-associated alleles across human populations. Admixture results in a mosaic of chromosomal segments originating from the parental populations within admixed individuals. This process can be an efficient

way to contribute novel phenotypic variation, if trait-associated alleles that were previously absent are introduced into a population. This can be seen in the distribution of lactase persistence *LCT* gene variants (further discussed below) in South Africa. Genome sequencing of a diverse set of South African populations revealed that lactase persistence was introduced in two separate admixture events (Breton et al., 2014; Macholdt et al., 2014). Strict pastoralist groups such as the Nama carry a persistence allele common in East African pastoralists, consistent with archaeological and genetic evidence for a pastoralist migration to South Africa within the last 2,000 years. Individuals from the mixed-ancestry population, on the other hand, carry the "European" allele at high frequencies—a consequence of recent admixture with European colonists.

The introduction of genetic susceptibility to malaria in Madagascar serves to also highlight the impact of admixture on the evolution of disease. The Malagasy people are predominantly of Eastern African descent and as such carry the Duffy-negative blood group, which confers complete protection from *Plasmodium vivax* malaria. However, admixture with Austronesian peoples that have reached the island from East Asia over the past 2,000 years have resulted in the introduction of Duffy-positive blood groups and, with it, susceptibility to *P. vivax* malaria. Interestingly, this process seems to have established a sufficiently high prevalence of malaria parasites exposed to Duffy-negative erythrocytes for the parasites to evolve a novel erythrocyte invasion mechanism (Ménard et al., 2010). As a result, clinical malaria is now also commonly observed in Duffy-positive Malagasy individuals, despite carrying the previously protective allele.

Adaptive Genetic Processes

As humans expanded across the globe, they encountered new environments and adapted to novel physical and cultural conditions. The genetic diversity of present-day humans is thus reflective of past selection events, particularly in isolated populations experiencing unique lifestyles and challenges. Some of these adaptations have important implications for human health.

Adaptation to New Environments

Among the most extreme examples of novel environments inhabited by humans are high-altitude regions. Previous research on high-altitude populations (reviewed in Ilardo and Nielsen [2018]) has provided new insights into the physiology of chronic hypoxia and related health disorders (Grocott et al., 2007). Examples of high-altitude human habitation include the Tibetan plateau, the highlands of Ethiopia, and the Andes mountain range. At high elevations, the low partial pressure of oxygen triggers various altitude acclimatization responses in humans. One component of these responses is orchestrated by hypoxia-inducible factors (HIFs), triggering the release of erythropoietin. This hormone stimulates erythropoiesis in the bone marrow, which results in an excess of red blood cells or polycythemia (reviewed in West [2006]). This physiological response, which is detectable as elevated hemoglobin levels in acclimatized individuals, carries with it the benefit of increased oxygen delivery. However, it has a number of negative consequences, such as increased blood viscosity, which can negatively affect cardiac function and complicate pregnancies.

At least three human populations appear to have independently evolved adaptive mechanisms to counteract the detrimental effects of the altitude-acclimatization response. In the Tibetan plateau highlanders and the Amhara high-altitude dwellers of Ethiopia, natural selection appears to have blunted the HIF pathway. Tibetans have reduced HIF signaling due to natural selection on components of the HIF pathway (EPAS1 and EGLN1), thus eliciting lower-than-expected hemoglobin levels (Beall et al., 2010; Bigham et al., 2010; Yi et al., 2010). Similarly, the Amhara appear to have a blunted HIF response that has been linked to BHLHE41, a gene encoding an upstream regulator of HIF signaling, although additional candidate genes have been proposed (Alkorta-Aranburu et al., 2012; Beall et al., 2002; Hoit et al., 2011; Huerta-Sánchez et al., 2013; Scheinfeldt et al., 2012). In contrast, the high-altitude populations of the Andes seem to have undergone selection for phenotypes related to cardiovascular health (Crawford et al., 2017), the reduction of blood pressure (Cowburn et al., 2013), and the reduction of oxidative stress (Bigham et al., 2009).

Extreme temperatures also present a major physiological stress; even at moderately cold temperatures, the human body initiates a vasoconstrictive response that increases cardiac load (Rintamäki, 2007). It was recently observed that allele frequencies of a variant upstream of TRPM8 increase along a latitudinal cline, corresponding with a decrease in temperature (Key et al., 2018). TRPM8 encodes a cation channel involved in detecting and reacting to a cold stimulus. The derived allele, which is at higher frequency at higher latitudes, appears to blunt the endogenous, metabolically costly response to cold temperatures. It therefore potentially reflects positive (directional) selection on the response to cold temperatures. Interestingly, however, the ancestral allele has been demonstrated to be protective against migraine; thus, migraine prevalence among present-day populations may be linked to unrelated, past selective events (Key et al., 2018).

The presence of environmental toxins can also act as a selective pressure. For example, the Atacama Desert is not only the most arid environment on Earth, but what little water is available is often contaminated by high levels of arsenic. Arsenic is highly toxic and exposure to inorganic arsenic is associated with severe health complications including cancer, cardiovascular and liver toxicity, and increased mortality later in life (Apata et al., 2017). Natural variation in the ability to process and excrete ingested arsenic has been linked to genetic variation. Specifically, polymorphisms in the gene AS3MT (an arsenite methyltransferase) have been implicated in arsenic methylation in a number of populations worldwide (Agusa et al., 2011; Chung et al., 2009; Engström et al., 2015; Gomez-Rubio et al., 2010; Pierce et al., 2013). In populations living in the Atacama Desert, natural selection is argued to have acted on genetic variation in AS3MT, allowing them to more efficiently metabolize the arsenic to which they are chronically exposed (Apata et al., 2017).

Adaptation to Cultural and Technological Changes

Past technological and cultural innovations have also altered the modes of human subsistence. Thus, humans have created new "self-made" ecological niches that, in the long run, have resulted in new adaptive pressures. A well-known example of this is the evolution of lactase persistence after the advent of agriculture,

which enabled regular consumption of dairy (extensively reviewed in Ségurel and Bon [2017]). Following the domestication of cattle, dairy became one of the most ubiquitous and important nutritional sources in the recent history of our species. Dairy from cows contains the sugar lactose, and most human infants are able to efficiently break down lactose into glucose and galactose via the enzyme lactase. However, the production of lactase abates after weaning, resulting in adult lactase deficiency. Lactose thus passes intact into the colon, where it is consumed and fermented by gas-producing microbes. Consequently, the consumption of dairy causes gastrointestinal stress to humans not carrying adaptive alleles. The ability to utilize dairy as a food source presented such a strong selective advantage that lactose tolerance evolved independently at least twice: in European (Enattah et al., 2002) and East African pastoralist populations (Tishkoff et al., 2007). In both instances, variants near the LCT gene allow individuals to maintain elevated levels of the lactase enzyme through adulthood (also known as lactase persistence).

The development of new hunting technologies also affected human diets, for example in Arctic populations. The traditional diet of many Arctic people is predominantly marine based, including large quantities of marine mammals that are rich in fatty acids. Fatty acids, when consumed in excess, can lead to health problems including high cholesterol and heart disease. In response, two separate adaptations to this dietary stressor have been observed in Arctic populations. CPT1A encodes a catalyst essential to the uptake of long-chain fatty acids and harbors signatures of selection in Northeastern Siberians (Clemente et al., 2014; Skotte et al., 2017). In turn, in the Greenlandic Inuit, the fatty acid desaturase (FADS) genes also harbor signatures of selection. These genes encode enzymes that regulate the rate of long-chain polyunsaturated fatty acid synthesis and seem to have allowed the Inuit to better process the particular fatty acids that came about with a rich marine-based diet (Fumagalli et al., 2015). Selection on the FADS genes have also occurred in other populations multiple times, particularly during transitions to a more or less vegetarian diet, including in Africa (Ameur et al., 2012), South Asia (Kothapalli et al., 2016), Europe (Buckley et al., 2017; Mathieson and Mathieson, 2018; Ye et al., 2017), and East Asia (Liu et al., 2018).

Adoption of extreme lifestyles has also led to adaptive responses in humans. For instance, the Sea Nomads of southeast Asia have a marine-based lifestyle that is heavily reliant on free diving. Breath-hold diving elicits a strong physiological response. This response includes bradycardia and peripheral vasoconstriction, to reduce the consumption of oxygen and preserve oxygen-rich blood for the most sensitive organs, and contraction of the spleen to provide an oxygen boost via stored, oxygenated red blood cells. Ilardo et al. (2018) demonstrated that the Bajau have larger spleens than a neighboring population living on land. This physiological trait is believed to endow them with superior breath-holding abilities. The large-spleen phenotype was shown to be associated with genetic variants in the gene PDE10A. This gene encodes a phosphodiesterase that affects the thyroid-hormone-production pathway. As thyroid hormones have been shown to affect spleen size, it is likely that the Bajau have larger spleens due to increased thyroid-

hormone production. A selection scan identified additional candidate genes with potential relevance for diving, including the bradykinin receptor BDKRB2. Previous work has shown that bradykinin levels affect peripheral vasoconstriction in response to the diving stimulus, and BDKRB2 is the only other gene previously argued to influence human dive capacity (Baranova et al., 2017). The Bajau study was the first to investigate human adaptation to breath-hold diving, and future studies may uncover further adaptations to this unique lifestyle in genetically distinct diving populations. This has potentially important implications for understanding human response to acute hypoxia, which can be often seen in critical care incidents (McKenna and Martin, 2016).

Adaptation to Pathogens

Genetic variants affecting the response to pathogen infections are likely to be under strong natural selection because of the intimate relationship between fitness and the ability to survive infectious diseases. In fact, the first genetic variants discovered to be under selection were associated with the infectious-disease response (Allison, 1954). As pathogens vary among geographic locations, it has been argued that much local adaptation leading to genetic differentiation among humans has been driven by pathogens. For instance, Fumagalli et al. (2011) compared how closely pathogen, climate, and diet-related environmental factors predict differentiation of allele frequencies among human populations. They showed that pathogenic diversity is a better predictor of allele-frequency differences among populations in genic SNPs than other environmental factors. Similarly, Daub et al. (2013) reported evidence that gene sets related to the immune response are enriched for signals of population differentiation.

Perhaps one of the most prominent examples of human adaptation to pathogens, the HbS allele of hemoglobin-beta (HBB), alters the shape of red blood cells, thereby conferring the carrier with strong resistance to malaria infection from P. falciparum (Ackerman et al., 2005). Across different regions of Africa, HbS and other HBB alleles vary dramatically in frequency. In Equatorial African populations with a high incidence of malaria infection from P. falciparum, HbS is at unusually high frequencies. Despite conferring some protection against certain forms of malaria, the HbS allele has not gone to fixation in any population, as people homozygous for this variant suffer from sickle-cell disease. Thus, HbS is not only a prime example of adaptation to pathogens, it is also an example of heterozygote advantage (overdominance) in humans.

The FY*O mutation of the DARC/ACKR1 (Duffy) gene, another important example of adaptation to malaria resistance, illustrates a common molecular path toward the evolution of resistance to specific pathogens: the deactivation of cell-surface receptors, or other proteins, that are otherwise employed by the pathogens during infection (Figure 2A). This locus exhibits some of the most extreme geographical differentiation in the human genome. The FY*O mutation is nearly fixed in sub-Saharan Africa—with notable exceptions, such as ‡Khomani San and Zulu populations, where FY*O segregates at 22% and 79% frequencies, respectively—while it is nearly non-existent in other populations. DARC codes for an erythroid surface protein necessary for infection from P. vivax, and the FY*O allele is a mutation in the promoter region that disrupts expression of these surface
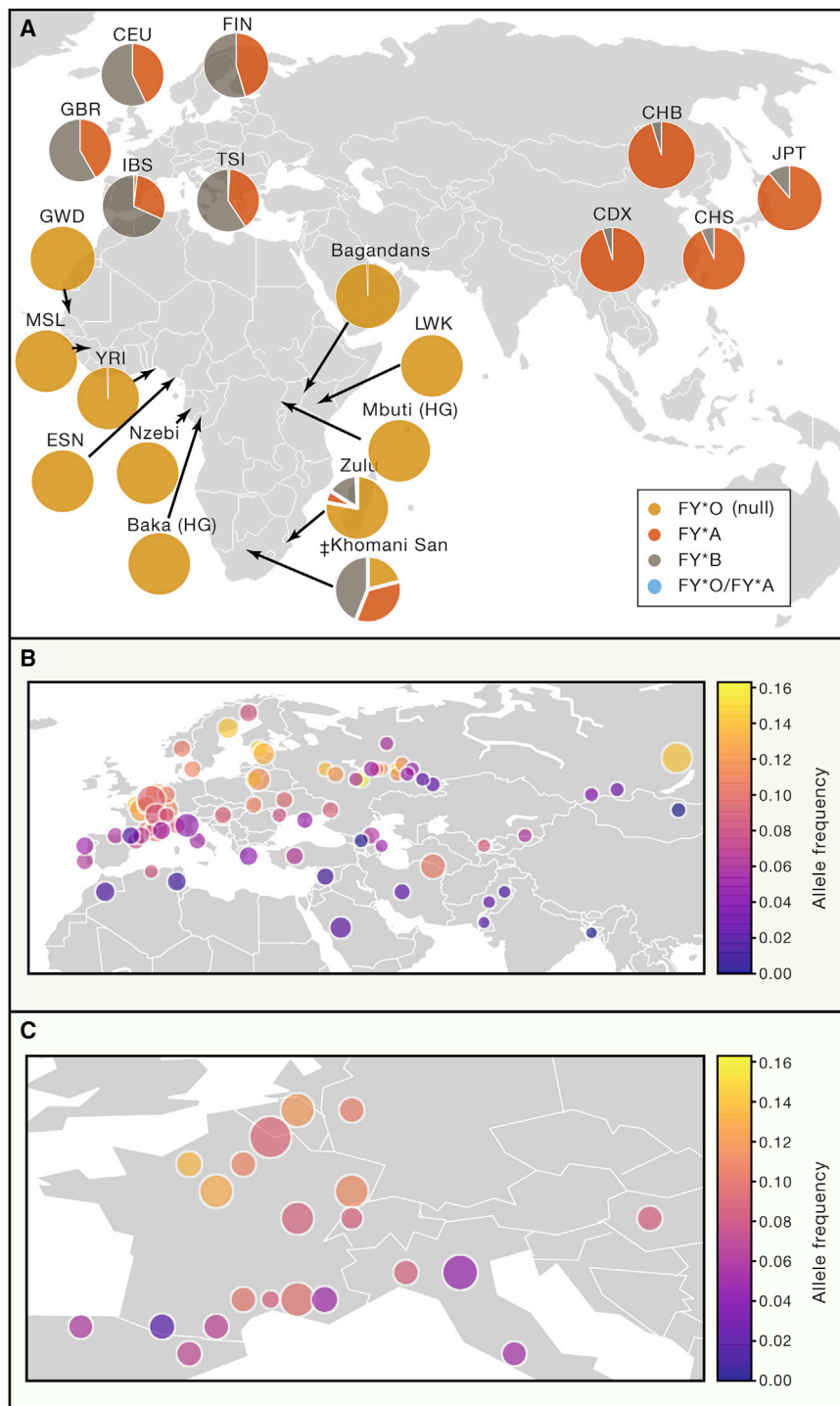
Figure 2. Differentiation of Pathogen-Response Alleles among Human Populations
(A) Geographical distribution of Duffy alleles. Frequencies were calculated by McManus et al. (2017) from the following datasets: African Genome Variation Project, Human Genome Diversity Project, and the 1000 Genomes Project (Phase 3).
(B) The geographical distribution of the CCR5-delta32 allele.
(C) Close-up of the Central European geographical distribution of the CCR5-delta32 allele.
Allele frequencies in (B) and (C) were both calculated from the POPRES data by Novembre et al. (2005). Scatter points sizes are proportional to sample size.

conferred a fitness advantage of over 4%—one of the highest estimated selection coefficients for any common variant in humans.

Another example of resistance via gene disabling is a deletion in the CCR5 gene (CCR5-D32) (Figures 2B and 2C). This deletion destroys the protein, which is a co-factor used for HIV cell entry, thereby leaving the carriers largely resistant to HIV infection (Samson et al., 1996). CCR5-D32 frequency exhibits a north–south cline in Europe, West Asia, and North Africa (Libert et al., 1998). Some have speculated that this cline might reflect a northern origin for the allele and a subsequent dispersal from the north, while others speculate that the cline reflects intense selection on CCR5-D32, citing also the long-range linkage disequilibrium patterns on D32-carrying haplotypes. Novembre et al. (2005) and Sabeti et al. (2005) modeled allelic dispersal in a geographically structured population and showed that the pattern of differentiation at CCR5 is consistent with an intense fitness advantage for the deletion (~10%). However, revised recombination rate estimates in this region as well as ancient DNA evidence suggest that the age of CCR5-D32 may be not as recent as previously suggested, making it difficult to rule out the possibility that the allele evolved altogether neutrally (Novembre and Han, 2012; Sabeti et al., 2005).

proteins in erythrocytes. Recently, McManus et al. (2017) analyzed the DARC gene in an assortment of diverse populations and showed that haplotypes carrying the FY*O mutation are consistent with very recent positive selection on the locus (TMRCA [time to the most recent ancestor] = 42 kya, 95% confidence interval [CI]: 34–49 kya). They estimate that FY*O has

In contrast to innate immunity—which encompasses mechanisms of immunity that are inherited and do not change throughout life, often forming the first line of defense against pathogens (Barreiro and Quintana-Murci, 2010)—adaptive immunity consists in immune mechanisms that are capable of changing substantially during life history by acquiring a

"memory" of exposure to pathogens. Perhaps the most canonical example of pathogen-driven selection is the human leukocyte antigen (HLA) complex, a cluster of genes that encode the major histocompatibility complex (MHC) in humans. The role of MHC genes is to bind antigens and present them on the cell surface to T cells. The MHC genes, therefore, play a crucial role in the recognition of foreign antigens. The HLA region exhibits an unusually high level of genetic diversity that is thought to be maintained by various forms of balancing selection, including selection in a fluctuating environment, frequency-dependent selection, and overdominance (Andrés, 2011). Recently, Brandt et al. (2018) showed that overdominance only drives population differentiation when levels of overdominance differ between populations. Their result explains why, quizzically, HLA genes exhibit low levels of F_{ST} between divergent populations (i.e., samples obtained from different continents) yet high levels of F_{ST} between recently split populations.

The evidence of high differentiation of pathogen-response genes due to recent local selective pressures has been further strengthened by a recent finding that the differential immune response in human cell lines depends on genetic ancestry (Nédélec et al., 2016). Specifically, macrophages derived from self-identified European-American and African-American individuals show different transcriptional responses when infected with pathogens such as *Listeria* or *Salmonella*, and these responses correlate significantly with ancestry. They also harbor signatures of positive selection. This suggests that genetic variation at immune response expression quantitative trait loci (eQTLs) are consistent with recent population-specific directional selection.

Recently, ancient genomics have started to provide novel insights into adaptation to pathogens. Lindo et al. (2016) reported a remarkable case of fluctuating selection on pathogen response. Using modern and ancient sequencing transects from a Native American population (Tsimshians), they identified a SNP in an HLA gene (HLA-DQA1) with exceptionally high differentiation from European and Han Chinese outgroup populations. This signal is consistent with positive selection on HLA-DQA1 in the Tsimshian lineage. Though the frequency of the selected allele is high in the ancient samples, it is surprisingly low in the present-day population. The authors argued that only a model with a change from positive to negative (purifying) selection following European contact is consistent with the observed frequency change. This finding supports the idea that contact with other populations can be a driver of adaptation to pathogens.

Ancient genomes recovered from archaic humans such as Neanderthals and Denisovans have provided direct evidence for several admixture events involving modern and archaic humans. All human populations outside of Africa derive ~2% of their ancestry from Neanderthals (Prüfer et al., 2017), and groups from Oceania carry an additional ~1%–3% from Denisovans (Sankararaman et al., 2016; Vernot et al., 2016). While the overall amount of introgressed sequences is low, multiple lines of evidence nevertheless suggest that archaic DNA plays an important role in human phenotypic diversity and adaptation (Dannemann and Kelso, 2017; Huerta-Sánchez et al., 2014; Quach et al., 2016; Racimo et al., 2017). For example, a recent study by Enard and Petrov (2018) showed that Neanderthal introgressed segments are enriched for genes coding for proteins that interact with viruses and specifically with RNA viruses in present-day Europeans. This suggests that ancient virus epidemics may have favored archaic adaptively introgressed variants that provided resistance to this type of pathogens.

Selection and Genetic Architecture of Complex Traits

Many of the above-mentioned cases of adaptation occurred via large allele frequency changes at a few loci of large effect, leaving distinct signatures on the genome that are easy to detect. These are termed "selective sweeps" (Berry et al., 1991; Smith and Haigh, 1974). However, most human traits are highly polygenic, with no single allele contributing disproportionately to a trait. Selection operating to change or preserve these traits need not have occurred via selective sweeps but likely operated via much subtler allele frequency shifts at multiple loci across the genome (Pritchard et al., 2010). GWASs can be a roadway into studying these processes, as they have become a mainstay in the identification of genomic loci associated with polygenic traits, including complex diseases. This type of study has successfully identified over 10,000 trait-associated variants (Visscher et al., 2017) in humans, which may be causal or linked to causal variants affecting a trait, opening new research horizons for the study of disease.

Purifying and Stabilizing Selection

GWASs have made it possible to examine how negative or purifying selection has impacted the genetic architecture of polygenic traits. Purifying selection acts by pushing the frequencies of deleterious alleles to low frequencies and removing them from the population. For example, loss-of-function nonsynonymous variants tend to be rarer than other nonsynonymous variants (MacArthur et al., 2012). A negative relationship between effect-size estimates for a trait and allele-frequency estimates could suggest that purifying selection is keeping new mutations from rising to high frequency. In fact, studies of GWAS results have found this pattern for several sets of variants, including those associated with height, body mass index (BMI), prostate cancer risk, and many other complex traits (Yang et al., 2015; Mancuso et al., 2016; Zeng et al., 2018). The proportion of low-frequency variants contributing to a trait's heritability can thus be used to understand how strongly negative selection is acting on sets of trait-associated variants, if the trait has pleiotropic effects on fitness (Eyre-Walker 2010; Gazal et al., 2018). Negative selection has also resulted in younger SNPs with low levels of linkage disequilibrium, explaining a larger proportion of trait heritability than older SNPs (Gazal et al., 2017).

Population genetic models have also informed the design and interpretation of gene-disease association studies. This is because models of the relationship between different forms of natural selection and the distribution of trait-associated variants are key to understanding how selection shapes the genetic architecture of a trait. Lohmueller (2014) showed that recent population growth can influence the proportion of mutations in a population that influence fitness. When these mutations also have a large effect on a studied trait, they can increase the proportion of the trait's heritability that is composed of low-frequency variants, thereby reducing the power of association tests. Sanjak et al. (2017) showed that types of gene action

relating genotypes to the value of a trait impact how genetic variance for that trait is partitioned across the allele-frequency spectrum. This, in turn, impacts the power of statistical tests of association.

More recently, Simons et al. (2018) developed a comprehensive model to study "stabilizing selection" acting in a multidimensional trait space. Stabilizing selection acts by keeping the mean trait value of a population at an optimum. It generally leads to negative selection on deleterious variants that might push the trait away from the optimum. Simons et al.'s model allowed the authors to trace the allele dynamics of a trait under stabilizing selection, while accounting for pleiotropic effects on other traits. They were able to relate the genetic architecture of a trait to the strength of selection acting to stabilize the trait and the amount of pleiotropy and to derive useful predictions about how large an association study must be to recover a particular fraction of a trait's heritability.

Polygenic Adaptation

GWAS data have also enabled the study of "polygenic adaptation" on a wide variety of trait-associated variants (Pritchard et al., 2010; Guo et al., 2018). Polygenic adaptation is the process by which positive selection acts to concertedly shift the frequency of alleles at many variants that affect a trait, moving the average phenotype of a population to a new optimum, without necessarily causing large allele-frequency changes at any one variant. Polygenic adaptation has been suggested to be the reason why population geneticists have found relatively few highly differentiated alleles across human populations (Pritchard et al., 2010). Studies of polygenic adaptation (Berg and Coop, 2014; Racimo et al., 2018; Robinson et al., 2015) aim to assess whether trait-associated variants have behaved in a way that is inconsistent with genetic drift by studying their frequencies within and between populations. Researchers have found evidence for polygenic adaptation on a number of complex traits and disorders, including skin pigmentation, height, BMI, and inflammatory bowel disease (Berg and Coop, 2014). Recently, evidence for polygenic adaptation on particular traits like age at first birth in females and BMI in males has also been recovered by performing a regression between individual lifetime reproductive success and the aforementioned traits (Sanjak et al., 2018), based on classic selection gradient theory by Lande and Arnold (1983).

However, some of these results have been recently called into question (Berg et al., 2018; Sohail et al., 2018). Most of these tests rely on obtaining accurate and unbiased estimates of the effect sizes of particular variants on a trait or disorder of interest. These estimates are often obtained from a GWAS on that trait. It is thus important to note that, if the GWAS has not properly controlled for "population stratification," these tests will also be biased (Berg et al., 2018; Sohail et al., 2018). Population stratification occurs when cases and controls in a GWAS are not well matched for genetic ancestry. For example, if the case group in a GWAS contains a larger proportion of individuals with European versus non-European ancestry than the control group, then the estimates of association between a specific variant and the trait will be confounded by the fact that that particular variant may have different frequencies in individuals of European and non-European ancestry. Moreover, these tests often rely on a com-

parison between trait-associated variants and some set of variants that are used as the neutral stand-in. These two sets of variants must be well controlled for in terms of allele frequencies, recombination rates, levels of background selection, or other properties that may confound the particular test used, which can be difficult to achieve. Furthermore, even if population stratification and other potential confounders have been fully accounted for, some care is needed in interpreting studies of selection based on GWAS. For an extensive review of additional sources of bias and caveats in interpretation, see Novembre and Barton (2018).

Alternatively, polygenic adaptation can be studied using pathway enrichment analyses, which do not rely on GWAS effect-size estimates. For example, Daub et al. (2013) found evidence for widespread positive selection on pathways involved in the response to pathogens. Another recent study searched biological pathways containing genes that harbor signatures of selection and that directly interact with each other in a biological network (Gouy et al., 2017). This way, the authors found evidence for polygenic adaptation on a pathway involved in the response to hypoxia in high-altitude Tibetan populations, again highlighting the importance of genetic adaptations in these groups. Another recent study based on pathway enrichment found signatures of convergent polygenic adaptation in multiple populations of rainforest hunter-gatherers in Africa and Asia, on pathways related to growth-factor binding and cardiac development (Bergey et al., 2018).

Impact of Human Diversity on Disease Gene Mapping and Prediction

Challenges for Genetic Disease Risk Prediction

Regardless of whether differences in the polygenic architecture of diseases are due to natural selection or neutral causes, these differences play an important role in our ability to predict health outcomes from genomic data. For example, a recent study showed via simulations that natural selection and population growth can strongly increase the impact of rare alleles on trait variance, decreasing the ability of GWASs, which rely on common variants, to identify trait-associated loci (Uricchio et al., 2016).

Additionally, demographic and adaptive forces limit the ability of polygenic risk scores (PRSs) produced from a GWAS in one population to predict risk in other, distantly related populations. PRSs (Torkamani et al., 2018) are predictors of disease liability in undiagnosed subjects, computed as the sum of risk alleles carried by an individual, weighted by risk effect estimates obtained from a GWAS. These scores have been used to identify individuals who carry clinically appreciable increases in disease liability, for example for coronary artery disease or type 2 diabetes (Khera et al., 2018). However, the risk predicted by these scores is a function of the specific patterns of linkage disequilibrium, allele frequencies, and allelic effect sizes within the population studied in the original GWAS (Martin et al., 2017).

Recent studies have shown via simulations that PRSs derived from European GWASs are poorer predictors of disease risk in individuals from distantly related populations (Martin et al., 2017). As an example, PRSs for height derived from European GWASs tend to be lower for West Africans than for Europeans, even though the former are on average as tall as the latter (Martin

et al., 2017). This could occur because variants affecting height in West Africans may be less common in Europeans and missed in European-centric GWASs, because causative variants have different effect sizes in West African populations than in European populations, because marker genotypes may be differently linked to underlying causative variants in the two populations, or due to unaccounted gene-by-environment interactions, among other reasons. Leveraging the full predictive potential of PRSs will thus require large-scale GWASs to be carried out in non-European individuals (e.g., Hilton et al. [2010]; Liu et al. [2018]; Nagai et al. [2017]; Wojcik et al. [2018]).

Leveraging Diversity for Disease Gene Mapping

Differences in allele frequencies, patterns of linkage disequilibrium, and/or disease prevalence between populations can actually be an advantage under specific study paradigms. For example, "admixture mapping" (Winkler et al., 2010) uses individuals from a population that resulted from recent admixture between two differentiated parental populations. This method relies on looking at sites where allele frequencies vary between the two parental populations and is based on the assumption that the differences in the rate of disease between these two populations are partly due to frequency differences at disease-causing variants. Due to the process of recombination after the initial admixture event, different individuals from the admixed population will have different parental ancestries at these variants. Admixture mapping consists of finding disease-associated loci by finding correlations between the parental ancestry in a particular region of the genome and the presence or absence of the disease in each individual. This approach has been used, for example, in studies of African-Americans, who can trace ancestry to European and West African populations, to identify loci associated with complex diseases such as hypertension, obesity, and prostate cancer (Winkler et al., 2010).

In a similar vein, populations that have undergone strong bottlenecks can be leveraged in gene-mapping studies (Peltonen et al., 2000; Southam et al., 2017; Xue et al., 2017). The power to detect a causal variant in an association study is a function of the proportion of variance in susceptibility it explains in a studied population, which itself increases with variant frequency. The strong genetic drift that results from a population bottleneck can lead to large increases in frequency for disease-associated variants, which may have been kept at low frequencies in the original population due to the stronger efficiency of negative selection. This, in effect, increases the power to detect these variants in association studies. Examples include mapping genes for Hirschsprung disease in Mennonites or non-syndromic hearing loss in Bedouins, among others (reviewed in Peltonen et al. [2000]).

Another approach, "trans-ethnic fine mapping" (Asimit et al., 2016), leverages population differences in patterns of linkage disequilibrium among variants within a locus to improve the identification of causative variants among a set of correlated markers. When marker genotypes within a disease-associated locus are highly correlated, both among each other and with a putative causal variant, it can be difficult to attribute the correlated effect estimates to the causal variant correctly. The differentiation of human populations provides a natural experiment that randomizes the dependencies among markers and causal variants, and a variety of statistical approaches can leverage this to predict which are the most likely causal variants. This approach is powerful when the differences in patterns of linkage disequilibrium are large and thus is most effective when large association studies are performed in highly differentiated populations.

Complex Gene-by-Environment Interactions in Disease Etiology

Given that present-day human populations differ in their genetic makeup and that many genetic variants impact human biology, it is expected that vulnerability to many diseases will also differ across populations, as most human diseases have a substantial genetic component. However, different populations are also exposed to different environments, which can lead to further differences in disease prevalence among populations as well. The interdependence between population differences and disease risk is often complicated by behavioral and socioeconomic factors that change over short time periods. This may become particularly important in the case of lifestyle diseases where differences in environment and genetic makeup between populations are confounded with cultural and socioeconomic differences, often to the point where it becomes impossible to distinguish between causal relationships and non-causal co-occurrences.

The prevalence of alcohol use disorder, for example, differs more than 10-fold among geographical regions, from 0.8% in Eastern Mediterranean regions to more than 8% in European and American Regions (World Health Organization, 2018). These differences are in part caused by genetic factors, such as alleles in alcohol-catabolic genes common in East and Southeast Asians that interact to increase the levels of acetaldehyde following ingestion of alcohol. While these alleles could hypothetically confer a protection against alcoholism, because of the erythema and physical discomfort that follows from increases in acetaldehyde levels (Edenberg, 2007), alcohol consumption is also a highly socially and culturally conditioned activity. Thus, it is difficult to separate the genetic, cultural, and environmental factors that could potentially affect the rates of alcohol use disorder over time and across cultures.

The genetic architecture of a behavioral trait may have also changed over time in response to societal changes. For example, a recent study found that as the proportion of smokers in a population decreased over recent decades, the genetic correlation between smoking and schizophrenia increased (Reginsson et al., 2018). This implies that the genetic correlation between human traits depends on exposure to environmental risk factors that in some cases is largely determined by societal norms and behavior.

Another example of the importance of the interplay between environment and genetics in relation to disease is the impact of a protein-disrupting variant of the TBC1D4 gene on the risk of type 2 diabetes (T2D) in Greenlandic Inuits. The variant seems to confer high risk of diabetes by decreasing glucose tolerance and is considered to be unique to the native population of Greenland, a historically isolated founder population. However, until very recently, the diet of Greenlanders was very low in sugars and carbohydrates. Thus, the variant had likely little negative physiological effects. This may have allowed it to drift upward in frequency or possibly to have been subject to positive selection (Moltke et al., 2014).
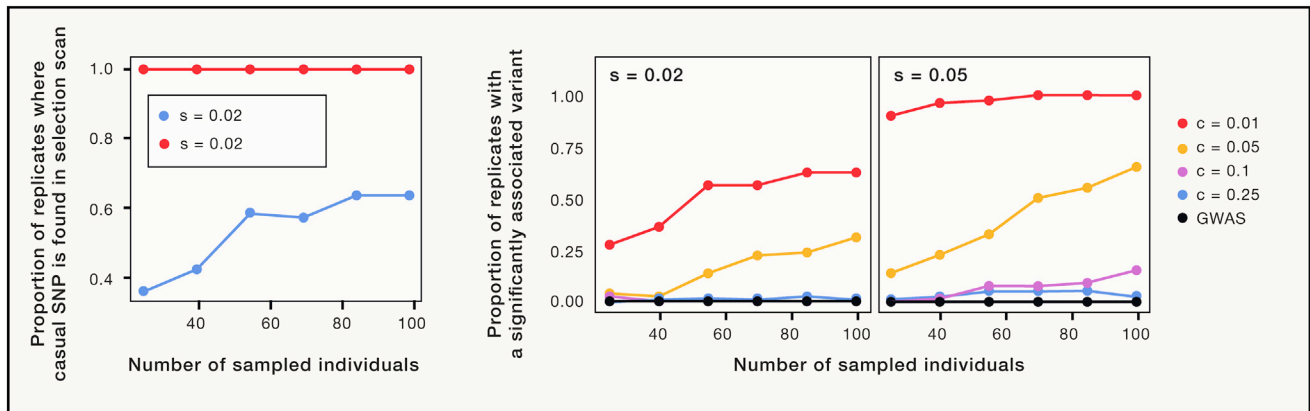
Figure 3. Power of Association Tests Guided by Selection Scans
Simulation study to assess the power to detect a causal variant with a selection scan followed by association mapping with small sample sizes. We simulated 100 whole-genome datasets (3,000 1 Mb regions) for a single European population under the model in (Gravel et al., 2011) using SLiM 2 (Haller and Messer, 2017). We assumed a single causal variant, uniformly chosen among all variants that had ~10% allele frequency 200 generations ago (5,000 years ago), and then simulated positive selection acting on this variant, with different selection coefficients (s), for the last 200 generations. We sampled 25–100 diploid individuals at two time points: the present and 201 generations ago (one generation before the onset of selection), respectively. For each replicate, we compared the ancient and present-day populations using the weighted average F_{ST} (Bhatia et al., 2013) computed over 50 kb blocks and kept the 20 most differentiated blocks (each represented by the site with the highest F_{ST}) for association mapping using a linear model (Purcell et al., 2007). Phenotypes for each individual were simulated as a Gaussian trait where the effect size on the standard variance scaling is defined by s/c (different values of c were explored). Left: Power to find the causal variant in the selection scan, as a function of the selection coefficient (s) and the number of sampled individuals. Right: Power of association mapping if the top 20 selection candidates are considered for further association testing with a Bonferroni correction. Notice that due to multiple testing, a traditional GWAS would have essentially zero power to detect associations for these sample sizes. In addition, we note that the false-positive rate for all sample–effect size combinations was below 10%, except for the two largest sample sizes with the largest effect size (both <14%).

Future Directions
Positive Selection Studies as a Roadway to Phenotypically Important Loci
As we have seen throughout this Review, people in many areas of the world have experienced drastic environmental changes, including changes in diet and lifestyle. Similarly, fast-evolving pathogens that challenge the immune system in populations around the world have likely led to "red-queen" dynamics, in which the immune system is constantly challenged to evolve to maintain its efficacy (e.g., Woolhouse et al. [2002]).

Evolutionary theory suggests that there likely will be genetic trade-offs for many traits that evolve as a consequence of environmental perturbations (Neel, 1962; Stearns, 1989). For example, "thrifty" alleles that previously were beneficial may, in the context of a modern diet, become deleterious because they predispose the carriers to diseases such as type 2 diabetes, obesity, and metabolic syndrome (Neel, 1999). The mutation in TBC1D4 in the Inuit population in Greenland, which strongly increases the risk of type 2 diabetes, may be an example of such a thrifty allele. Another particularly well-publicized hypothetical shift in equilibrium imposed by novel environmental conditions is the "hygiene hypothesis" (Strachan, 1989), which posits that there is a trade-off between an immune system that is too active incurring an increased risk of allergies and autoimmune disease and an immune system that is too passive and that provides insufficient defense against new pathogens. Finally, trade-offs have also been claimed to help explain the existence of cancer (Jacqueline et al., 2016). One way to identify these disease-related trade-offs may be via genomic studies of natural selection. Identifying genes under positive selection in humans during the last millennia might allow us not only to understand the evolu-

tionary dynamics of disease but also to find important variants for human health. Variants that are deleterious in an environment are likely to segregate at low frequencies. However, if, due to environmental changes, variants with a large effect on human health are positively selected, they will begin to segregate at high frequencies.

Selection scans may thus have the potential to aid in association mapping, in particular when sample sizes are small and the selected variants are associated with a trait or disease of interest (Figure 3). To show how this could be possible, we simulated whole-genome data for a population with a European-like demographic history and introduced a causal variant that was also under positive selection. We sampled 25–100 diploid individuals at two time points: the present and 201 generations ago (1 generation before the onset of selection). We then compared them using weighted average F_{ST} (Bhatia et al., 2013) computed over 50 kb blocks. We kept the 20 most differentiated blocks (each represented by the site with the highest F_{ST}) for association mapping using a linear model (Purcell et al., 2007). Phenotypes for each individual were simulated as a Gaussian trait. We find that, due to multiple testing, a traditional GWAS would have essentially zero power to detect associations for these sample sizes, but the power of a "selection-boosted" GWAS is much higher (Figure 3). In addition, the false-positive rate under most explored combinations of parameters was below 10%, except for the two largest sample sizes with the largest effect size (both <14%).

While most complex diseases have a polygenic basis, there are many cases of mutations segregating at high frequency with large effect sizes. Such mutations are arguably often a consequence of strong selection acting on a trait, perhaps
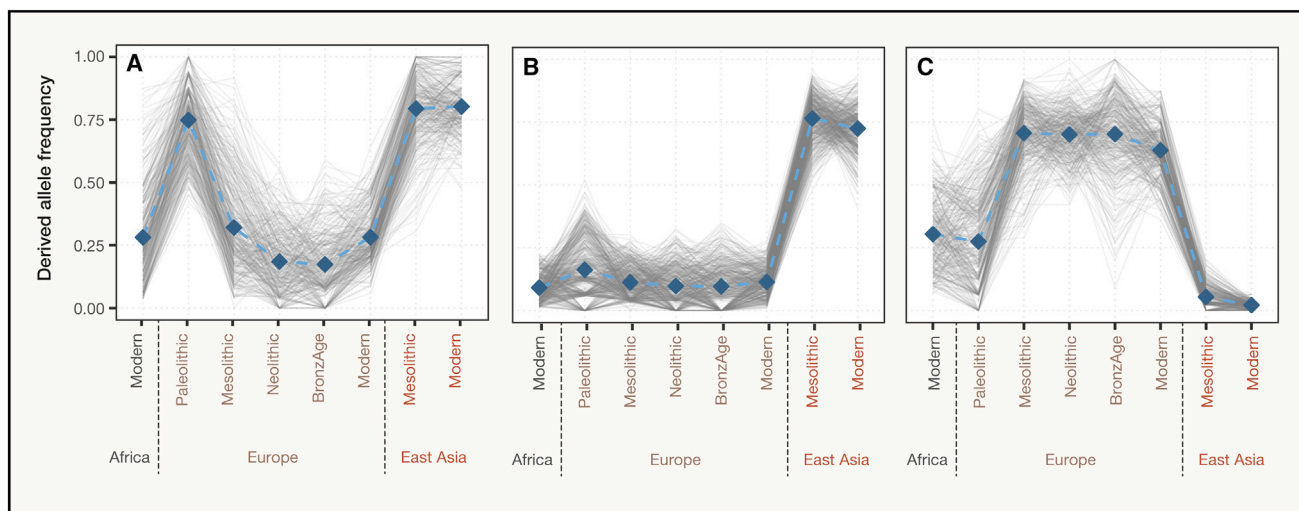
Figure 4. Example Allele Frequency Trajectories of SNPs with High Temporal and Spatial Differentiation

Each panel shows individual (gray lines) and mean (blue lines and symbols)-derived allele-frequency trajectories of SNPs among the top 0.01% of most extremely differentiated SNPs between 174 ancient individuals and 15,496 present-day whole-genome sequences from the Genome Aggregation Database (gnomAD) (Lek et al., 2016). Panels show examples of patterns of highly differentiated SNPs, with high-derived allele frequencies specific to (A) European Paleolithic and ancient and present-day East Asians, (B) ancient and present-day East Asians, and (C) post-Paleolithic ancient and present-day Europeans.

related to shifting environmental conditions and/or evolutionary trade-offs. In particular, selection acting in heterogeneous environments tends to favor genetic architectures with fewer loci (Yeaman and Whitlock 2011). Indeed, the previously discussed EPAS1 mutation in Tibetans might be an example of this phenomenon. Detecting variants under selection can, therefore, be a valuable complementary tool to other types of studies in human disease genetics.

Previous studies have already shown that selection scans can be a powerful way to identify candidate regions for further association testing. These methods have, for example, identified regions affecting differential regulation of red blood cell production in Tibetans (Huerta-Sánchez et al., 2014; Peng et al., 2011) and important variation associated with endogenous production of omega-3 fatty acids in the Greenlandic Inuit (Fumagalli et al., 2015). Other studies have shown that most signatures of recent positive selection in modern humans have been found in loci associated with pathogens (Enard and Petrov, 2018; Fumagalli et al., 2011), but very little work has been done to leverage this insight for discovery of associations between genetic variation in humans and susceptibility to infectious diseases.

An Ancient Genomics Approach to Understand the Genetic Architecture of Disease

Ancient genomics holds great potential to identify genetic variants that were subjected to negative and positive selection and to better understand how these processes have shaped the genetic architecture of disease. They provide a way to build detailed time series of allele frequency trajectories, so that selection can be caught "in the act" (Figure 4). The application of ancient genomics to the study of human diseases, however, is in its infancy due to poor spatial and temporal resolution of available records. Recently, the number of genome sequences produced in a single ancient genomics study became similar to

what is regularly generated in studies on extant populations (e.g., Damgaard et al. [2018]). However, most of these large-scale ancient genomic studies have generated either low-coverage genome sequences (~1X) or targeted capture of common genomic variants. Although these approaches are well suited for the analysis of deep demographic history at broader, continental-level scales, they do not allow for the detection of rare mutations potentially involved in disease susceptibility. An important step in facilitating the application of ancient genomics to elucidating human disease is to produce a large number of higher-coverage genomes that span longer timescales.

Ancient DNA from skeletal remains also provides genomic information of blood-borne microorganisms (e.g., DNA viruses, bacteria, parasites) that infected the ancient individuals at the time of death, thereby providing the means to directly investigate the prevalence of ancient diseases and the genetic makeup of their causal pathogens. Several studies have demonstrated that shotgun sequencing datasets of ancient individuals with no prior evidence of disease are suitable for detecting pathogens responsible for a variety of severe diseases (Mühlemann et al., 2018; Rasmussen et al., 2015; Schuenemann et al., 2013). This approach enables the detection of pathogen lineages that impacted prehistoric human populations, including those that are rare or unknown today. It also allows for determining when and where these lineages and their pathogenicity-associated variants emerged and how they spread among these populations. For instance, Rasmussen et al. (2015) demonstrated that *Yersinia pestis*, the etiological agent of plague, was widespread in Eurasia at least 3,000 years earlier than any historical records of the disease, and they identified a temporal sequence of genetic changes that led to its increased virulence and the emergence of the bubonic plague. Extinct strains of the hepatitis B virus and parvovirus B19 in humans from central Eurasia and

western Europe up to 7,000 BP (before present) have also been recovered (Mühlemann et al., 2018). Bringing together information about genomic variation targeted by natural selection in humans and pathogens present in the deceased individuals could allow for direct investigation into association between specific human genetic variants and pathogenic infections. This is particularly relevant considering that most adaptations targeting immune-related genes seem to have taken place in the last 6–13 thousand years (Deschamps et al., 2016). However, little is known about the selection trajectories of the beneficial variants in the host genome, the identity of the pathogens, or the nature of their interaction with the host.

Genetic variants involved in the immune response (e.g., MHC, TCR, TLR) are expected to have a particularly strong effect on fitness in the presence of epidemics. Thus, even a minor improvement in the immune response to epidemic pathogens such as Y. pestis may translate into strong identifiable selection signals using ancient genomic data from before, during, and after the epidemics. If a selection episode occurred close in time to a known epidemic like plague or smallpox, and the population under study did not undergo a major replacement over the same time period, it should be possible to focus on the particular disease or response to disease that the variant is associated with and further validate the proposed associations using large-scale present-day genome datasets. While selection scans on ancient genomic datasets in itself are not sufficient for identifying functional relationships between the host and the pathogens, they can greatly help improve the power of future association studies.

Of course, one must be aware that the present-day occupants of a region need not be the direct descendants of the people that lived in the same location hundreds or thousands of years ago. As we discussed above, the genetic history of human populations is characterized by multiple migration, admixture, and replacement events. For this reason, selection scans that assume complete population continuity are likely to lead to false positives: allele frequencies at a particular locus may not have necessarily increased over time because of positive selection on that locus but simply because incoming populations into a region may have had higher allele frequencies at that locus than previous occupants. Thus, methods that can account for complex migration histories will need to be deployed in order to address this issue (Günther and Coop 2013; Lee and Coop 2017; Refoyo-Martínez et al., 2018).

Large-Scale Genotyping and Phenotyping of Underrepresented Present-Day Populations

Though the large majority of genetic diversity today is found in Africa, most large-scale human sequencing and phenotyping projects have been restricted to individuals of overwhelmingly European ancestry. Non-European populations from across the globe remain vastly understudied at the genomic level. This state of affairs limits the transferability of gene-trait association studies and prevents accurate disease risk prediction in these populations, exacerbating global health disparities (Martin et al., 2017, 2018).

Our understanding of the genetic basis of disease and the evolution of disease-causing variants over time will benefit tremendously from efforts to retrieve genomic and biomedical data from these understudied communities. Not only are we bound to find many more disease-causing variants by studying diverse populations across the world, but the genomes of people with historically small effective population sizes can provide information that is sometimes irretrievable from large homogeneous populations. This, in turn, can affect trait—and disease—prediction beyond the study population. For example, the TBC1D4 mutation mentioned above was discovered because it exists at intermediate frequencies in the isolated Greenlandic population, where the variant has likely risen in frequency due to drift or positive selection. In Europeans, this mutation only exists at low frequencies. Consequently, it was missed in earlier GWASs, though its effect on glucose tolerance and type 2 diabetes was several times larger than any previously found loci associated with these traits (Moltke et al., 2014).

Importantly, these efforts will need to be conducted in a concerted dialog with the local communities that will be most affected by the sampling. Informed consent forms should no longer be the bare minimum to which population genomic projects aspire: sequencing and phenotyping should involve native groups at both an educational and research level, with the goal of building up the scientific infrastructure that will allow local communities to thrive and take the maximum possible benefits from recovering their own genomic and biomedical information. In many corners of the world, where genomic infrastructure is limited or entirely absent, this will be a Herculean effort, but it is an effort that the scientific community must help facilitate.

Over the last decade, population genomics and paleogenomics have provided a wealth of new insights into the origin and evolution of human disease, many of which have important implications for healthcare strategies and disease treatment. Further advancement of this research area requires sustained efforts to produce and analyze genomes from ancient and present-day individuals, especially from understudied groups, and in continuous dialog with the communities that they aim to benefit. Moreover, a deeper understanding of disease-related phenomena will also require that evolutionary and population geneticists build bridges with other research disciplines that study disease and its implications for society, including psychology, neurobiology, sociology, anthropology, ecology, physiology, molecular and cellular biology, and statistical and medical genetics. Only via a truly multidisciplinary approach can we hope to discern the social, environmental, and biotic underpinnings of human disease and, ultimately, devise ways to curb them.

REFERENCES

1000 Genomes Project Consortium (2015). A global reference for human genetic variation. Nature *526*, 68–74.

Ackerman, H., Usen, S., Jallow, M., Sisay-Joof, F., Pinder, M., and Kwiatkowski, D.P. (2005). A comparison of case-control and family-based association methods: the example of sickle-cell and malaria. Ann. Hum. Genet. *69*, 559–565.

Agusa, T., Fujihara, J., Takeshita, H., and Iwata, H. (2011). Individual variations in inorganic arsenic metabolism associated with AS3MT genetic polymorphisms. Int. J. Mol. Sci. *12*, 2351–2382.

Alkorta-Aranburu, G., Beall, C.M., Witonsky, D.B., Gebremedhin, A., Pritchard, J.K., and Di Rienzo, A. (2012). The genetic architecture of adaptations to high altitude in Ethiopia. PLoS Genet. *8*, e1003110.

Allentoft, M.E., Sikora, M., Sjögren, K.-G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P.B., Schroeder, H., Ahlström, T., Vinner, L., et al. (2015). Population genomics of Bronze Age Eurasia. Nature *522*, 167–172.

Allison, A.C. (1954). Protection afforded by sickle-cell trait against subtertian malareal infection. BMJ *1*, 290–294.

Ameur, A., Enroth, S., Johansson, A., Zaboli, G., Igl, W., Johansson, A.C.V., Rivas, M.A., Daly, M.J., Schmitz, G., Hicks, A.A., et al. (2012). Genetic adaptation of fatty-acid metabolism: a human-specific haplotype increasing the biosynthesis of long-chain omega-3 and omega-6 fatty acids. Am. J. Hum. Genet. *90*, 809–820.

Andrés, A.M. (2011). Balancing Selection in the Human Genome. In eLS (John Wiley & Sons).

Apata, M., Arriaza, B., Llop, E., and Moraga, M. (2017). Human adaptation to arsenic in Andean populations of the Atacama Desert. Am. J. Phys. Anthropol. *163*, 192–199.

Asimit, J.L., Hatzikotoulas, K., McCarthy, M., Morris, A.P., and Zeggini, E. (2016). Trans-ethnic study design approaches for fine-mapping. Eur. J. Hum. Genet. *24*, 1330–1336.

Atanasovska, B., Kumar, V., Fu, J., Wijmenga, C., and Hofker, M.H. (2015). GWAS as a Driver of Gene Discovery in Cardiometabolic Diseases. Trends Endocrinol. Metab. *26*, 722–732.

Baranova, T.I., Berlov, D.N., Glotov, O.S., Korf, E.A., Minigalin, A.D., Mitrofanova, A.V., Ahmetov, I.I., and Glotov, A.S. (2017). Genetic determination of the vascular reactions in humans in response to the diving reflex. Am. J. Physiol. Heart Circ. Physiol. *312*, H622–H631.

Barreiro, L.B., and Quintana-Murci, L. (2010). From evolutionary genetics to human immunology: how selection shapes host defence genes. Nat. Rev. Genet. *11*, 17–30.

Beall, C.M., Decker, M.J., Brittenham, G.M., Kushner, I., Gebremedhin, A., and Strohl, K.P. (2002). An Ethiopian pattern of human adaptation to high-altitude hypoxia. Proc. Natl. Acad. Sci. USA *99*, 17215–17218.

Beall, C.M., Cavalleri, G.L., Deng, L., Elston, R.C., Gao, Y., Knight, J., Li, C., Li, J.C., Liang, Y., McCormack, M., et al. (2010). Natural selection on EPAS1 (HIF2alpha) associated with low hemoglobin concentration in Tibetan highlanders. Proc. Natl. Acad. Sci. USA *107*, 11459–11464.

Belbin, G.M., Odgis, J., Sorokin, E.P., Yee, M.-C., Kohli, S., Glicksberg, B.S., Gignoux, C.R., Wojcik, G.L., Van Vleck, T., Jeff, J.M., et al. (2017). Genetic identification of a common collagen disease in Puerto Ricans via identity-by-descent mapping in a health system. eLife *6*, e25060.

Berg, J.J., and Coop, G. (2014). A population genetic signal of polygenic adaptation. PLoS Genet. *10*, e1004412.

Berg, J.J., Harpak, A., Sinnott-Armstrong, N., Joergensen, A.M., Mostafavi, H., Field, Y., Boyle, E.A., Zhang, X., Racimo, F., Pritchard, J.K., et al. (2018). Reduced signal for polygenic adaptation of height in UK Biobank. bioRxiv. https://doi.org/10.1101/354951.

Bergey, C.M., Lopez, M., Harrison, G.F., Patin, E., Cohen, J.A., Quintana-Murci, L., Barreiro, L.B., and Perry, G.H. (2018). Polygenic adaptation and convergent evolution on growth and cardiac genetic pathways in African

and Asian rainforest hunter-gatherers. Proc. Natl. Acad. Sci. USA *115*, E11256–E11263.

Berry, A.J., Ajioka, J.W., and Kreitman, M. (1991). Lack of polymorphism on the Drosophila fourth chromosome resulting from selection. Genetics *129*, 1111–1117.

Bhatia, G., Patterson, N., Sankararaman, S., and Price, A.L. (2013). Estimating and interpreting F_{ST}: the impact of rare variants. Genome Res. *23*, 1514–1521.

Bigham, A., Bauchet, M., Pinto, D., Mao, X., Akey, J.M., Mei, R., Scherer, S.W., Julian, C.G., Wilson, M.J., López Herráez, D., et al. (2010). Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. PLoS Genet. *6*, e1001116.

Bigham, A.W., Mao, X., Mei, R., Brutsaert, T., Wilson, M.J., Julian, C.G., Parra, E.J., Akey, J.M., Moore, L.G., and Shriver, M.D. (2009). Identifying positive selection candidate loci for high-altitude adaptation in Andean populations. Hum. Genomics *4*, 79–90.

Bossé, Y., and Amos, C.I. (2018). A Decade of GWAS Results in Lung Cancer. Cancer Epidemiol. Biomarkers Prev. *27*, 363–379.

Boyle, E.A., Li, Y.I., and Pritchard, J.K. (2017). An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell *169*, 1177–1186.

Brandt, D.Y.C., César, J., Goudet, J., and Meyer, D. (2018). The Effect of Balancing Selection on Population Differentiation: A Study with HLA Genes. G3 (Bethesda) *8*, 2805–2815.

Breton, G., Schlebusch, C.M., Lombard, M., Sjödin, P., Soodyall, H., and Jakobsson, M. (2014). Lactase persistence alleles reveal partial East African ancestry of southern African Khoe pastoralists. Curr. Biol. *24*, 852–858.

Buckley, M.T., Racimo, F., Allentoft, M.E., Jensen, M.K., Jonsson, A., Huang, H., Hormozdiari, F., Sikora, M., Marnetto, D., Eskin, E., et al. (2017). Selection in Europeans on Fatty Acid Desaturases Associated with Dietary Changes. Mol. Biol. Evol. *34*, 1307–1318.

Carmi, S., Hui, K.Y., Kochav, E., Liu, X., Xue, J., Grady, F., Guha, S., Upadhyay, K., Ben-Avraham, D., Mukherjee, S., et al. (2014). Sequencing an Ashkenazi reference panel supports population-targeted personal genomics and illuminates Jewish and European origins. Nat. Commun. *5*, 4835.

Chung, C.-J., Hsueh, Y.-M., Bai, C.-H., Huang, Y.-K., Huang, Y.-L., Yang, M.-H., and Chen, C.-J. (2009). Polymorphisms in arsenic metabolism genes, urinary arsenic methylation profile and cancer. Cancer Causes Control *20*, 1653–1661.

Clemente, F.J., Cardona, A., Inchley, C.E., Peter, B.M., Jacobs, G., Pagani, L., Lawson, D.J., Antão, T., Vicente, M., Mitt, M., et al. (2014). A Selective Sweep on a Deleterious Mutation in CPT1A in Arctic Populations. Am. J. Hum. Genet. *95*, 584–589.

Coop, G., Pickrell, J.K., Novembre, J., Kudaravalli, S., Li, J., Absher, D., Myers, R.M., Cavalli-Sforza, L.L., Feldman, M.W., and Pritchard, J.K. (2009). The role of geography in human adaptation. PLoS Genet. *5*, e1000500.

Cowburn, A.S., Takeda, N., Boutin, A.T., Kim, J.-W., Sterling, J.C., Nakasaki, M., Southwood, M., Goldrath, A.W., Jamora, C., Nizet, V., et al. (2013). HIF isoforms in the skin differentially regulate systemic arterial pressure. Proc. Natl. Acad. Sci. USA *110*, 17570–17575.

Crawford, J.E., Amaru, R., Song, J., Julian, C.G., Racimo, F., Cheng, J.Y., Guo, X., Yao, J., Ambale-Venkatesh, B., Lima, J.A., et al. (2017). Natural selection on genes related to cardiovascular health in high-altitude adapted Andeans. Am J Hum Genet. *101*, 752–767.

Damgaard, P.B., Marchi, N., Rasmussen, S., Peyrot, M., Renaud, G., Korneliussen, T., Moreno-Mayar, J.V., Pedersen, M.W., Goldberg, A., Usmanova, E., et al. (2018). 137 ancient human genomes from across the Eurasian steppes. Nature *557*, 369–374.

Dannemann, M., and Kelso, J. (2017). The Contribution of Neanderthals to Phenotypic Variation in Modern Humans. Am. J. Hum. Genet. *101*, 578–589.

Daub, J.T., Hofer, T., Cutivet, E., Dupanloup, I., Quintana-Murci, L., Robinson-Rechavi, M., and Excoffier, L. (2013). Evidence for polygenic adaptation to pathogens in the human genome. Mol. Biol. Evol. *30*, 1544–1558.

Deschamps, M., Laval, G., Fagny, M., Itan, Y., Abel, L., Casanova, J.-L., Patin, E., and Quintana-Murci, L. (2016). Genomic Signatures of Selective Pressures and Introgression from Archaic Hominins at Human Innate Immunity Genes. Am. J. Hum. Genet. *98*, 5–21.

Edenberg, H.J. (2007). The genetics of alcohol metabolism: role of alcohol dehydrogenase and aldehyde dehydrogenase variants. Alcohol Res. Health *30*, 5–13.

Elhaik, E. (2012). Empirical distributions of F(ST) from large-scale human polymorphism data. PLoS ONE *7*, e49837.

Enard, D., and Petrov, D.A. (2018). Evidence that RNA Viruses Drove Adaptive Introgression between Neanderthals and Modern Humans. Cell *175*, 360–371.e13.

Enattah, N.S., Sahi, T., Savilahti, E., Terwilliger, J.D., Peltonen, L., and Järvelä, I. (2002). Identification of a variant associated with adult-type hypolactasia. Nat. Genet. *30*, 233–237.

Engström, K.S., Vahter, M., Fletcher, T., Leonardi, G., Goessler, W., Gurzau, E., Koppova, K., Rudnai, P., Kumar, R., and Broberg, K. (2015). Genetic variation in arsenic (+3 oxidation state) methyltransferase (AS3MT), arsenic metabolism and risk of basal cell carcinoma in a European population. Environ. Mol. Mutagen. *56*, 60–69.

Ewens, W.J. (2012). Mathematical Population Genetics 1: Theoretical Introduction (Springer Science & Business Media).

Eyre-Walker, A. (2010). Evolution in health and medicine Sackler colloquium: Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. Proc. Natl. Acad. Sci. USA *107* (*Suppl 1*), 1752–1756.

Fumagalli, M., Sironi, M., Pozzoli, U., Ferrer-Admetlla, A., Pattini, L., and Nielsen, R. (2011). Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. PLoS Genet. *7*, e1002355.

Fumagalli, M., Moltke, I., Grarup, N., Racimo, F., Bjerregaard, P., Jørgensen, M.E., Korneliussen, T.S., Gerbault, P., Skotte, L., Linneberg, A., et al. (2015). Greenlandic Inuit show genetic signatures of diet and climate adaptation. Science *349*, 1343–1347.

Gazal, S., Finucane, H.K., Furlotte, N.A., Loh, P.-R., Palamara, P.F., Liu, X., Schoech, A., Bulik-Sullivan, B., Neale, B.M., Gusev, A., and Price, A.L. (2017). Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. Nat. Genet. *49*, 1421–1427.

Gazal, S., Loh, P.-R., Finucane, H.K., Ganna, A., Schoech, A., Sunyaev, S., and Price, A.L. (2018). Functional architecture of low-frequency variants highlights strength of negative selection across coding and non-coding annotations. Nat. Genet. *50*, 1600–1607.

GBD 2015 Tobacco Collaborators (2017). Smoking prevalence and attributable disease burden in 195 countries and territories, 1990-2015: a systematic analysis from the Global Burden of Disease Study 2015. Lancet *389*, 1885–1906.

Geschwind, D.H., and Flint, J. (2015). Genetics and genomics of psychiatric disease. Science *349*, 1489–1494.

Gomez-Rubio, P., Meza-Montenegro, M.M., Cantu-Soto, E., and Klimecki, W.T. (2010). Genetic association between intronic variants in AS3MT and arsenic methylation efficiency is focused on a large linkage disequilibrium cluster in chromosome 10. J. Appl. Toxicol. *30*, 260–270.

Gouy, A., Daub, J.T., and Excoffier, L. (2017). Detecting gene subnetworks under selection in biological pathways. Nucleic Acids Res. *45*, e149.

Gravel, S., Henn, B.M., Gutenkunst, R.N., Indap, A.R., Marth, G.T., Clark, A.G., Yu, F., Gibbs, R.A., and Bustamante, C.D.; 1000 Genomes Project (2011). Demographic history and rare allele sharing among human populations. Proc Natl Acad Sci USA *108*, 11983–11988.

Grocott, M., Montgomery, H., and Vercueil, A. (2007). High-altitude physiology and pathophysiology: implications and relevance for intensive care medicine. Crit. Care *11*, 203.

Günther, T., and Coop, G. (2013). Robust identification of local adaptation from allele frequencies. Genetics *195*, 205–220.

Guo, J., Yang, J., and Visscher, P.M. (2018). Leveraging GWAS for complex traits to detect signatures of natural selection in humans. Curr. Opin. Genet. Dev. *53*, 9–14.

Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., et al. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. Nature *522*, 207–211.

Haller, B.C., and Messer, P.W. (2017). SLiM 2: Flexible, Interactive Forward Genetic Simulations. Mol. Biol. Evol. *34*, 230–240.

Hilton, C.L., Fitzgerald, R.T., Jackson, K.M., Maxim, R.A., Bosworth, C.C., Shattuck, P.T., Geschwind, D.H., and Constantino, J.N. (2010). Brief report: Under-representation of African americans in autism genetic research: a rationale for inclusion of subjects representing diverse family structures. J. Autism Dev. Disord. *40*, 633–639.

Hoit, B.D., Dalton, N.D., Gebremedhin, A., Janocha, A., Zimmerman, P.A., Zimmerman, A.M., Strohl, K.P., Erzurum, S.C., and Beall, C.M. (2011). Elevated pulmonary artery pressure among Amhara highlanders in Ethiopia. Am. J. Hum. Biol. *23*, 168–176.

Huerta-Sánchez, E., Degiorgio, M., Pagani, L., Tarekegn, A., Ekong, R., Antao, T., Cardona, A., Montgomery, H.E., Cavalleri, G.L., Robbins, P.A., et al. (2013). Genetic signatures reveal high-altitude adaptation in a set of ethiopian populations. Mol. Biol. Evol. *30*, 1877–1888.

Huerta-Sánchez, E., Jin, X., Asan, Bianba, Z., Peter, B.M., Vinckenbosch, N., Liang, Y., Yi, X., He, M., Somel, M., et al. (2014). Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature *512*, 194–197.

Ilardo, M., and Nielsen, R. (2018). Human adaptation to extreme environmental conditions. Curr. Opin. Genet. Dev. *53*, 77–82.

Ilardo, M.A., Moltke, I., Korneliussen, T.S., Cheng, J., Stern, A.J., Racimo, F., de Barros Damgaard, P., Sikora, M., Seguin-Orlando, A., Rasmussen, S., et al. (2018). Physiological and Genetic Adaptations to Diving in Sea Nomads. Cell *173*, 569–580.e15.

Jacqueline, C., Biro, P.A., Beckmann, C., Moller, A.P., Renaud, F., Sorci, G., Tasiemski, A., Ujvari, B., and Thomas, F. (2016). Cancer: A disease at the crossroads of trade-offs. Evol. Appl. *10*, 215–225.

Jha, P., and Peto, R. (2014). Global effects of smoking, of quitting, and of taxing tobacco. N. Engl. J. Med. *370*, 60–68.

Key, F.M., Abdul-Aziz, M.A., Mundry, R., Peter, B.M., Sekar, A., D'Amato, M., Dennis, M.Y., Schmidt, J.M., and Andrés, A.M. (2018). Human local adaptation of the TRPM8 cold receptor along a latitudinal cline. PLoS Genet. *14*, e1007298.

Khera, A.V., Chaffin, M., Aragam, K.G., Haas, M.E., Roselli, C., Choi, S.H., Natarajan, P., Lander, E.S., Lubitz, S.A., Ellinor, P.T., and Kathiresan, S. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat. Genet. *50*, 1219–1224.

Kothapalli, K.S.D., Ye, K., Gadgil, M.S., Carlson, S.E., O'Brien, K.O., Zhang, J.Y., Park, H.G., Ojukwu, K., Zou, J., Hyon, S.S., et al. (2016). Positive Selection on a Regulatory Insertion-Deletion Polymorphism in FADS2 Influences Apparent Endogenous Synthesis of Arachidonic Acid. Mol. Biol. Evol. *33*, 1726–1739.

Lande, R., and Arnold, S.J. (1983). THE MEASUREMENT OF SELECTION ON CORRELATED CHARACTERS. Evolution *37*, 1210–1226.

Lee, K.M., and Coop, G. (2017). Distinguishing Among Modes of Convergent Adaptation Using Population Genomic Data. Genetics *207*, 1591–1619.

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al.; Exome Aggregation Consortium (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291.

Libert, F., Cochaux, P., Beckman, G., Samson, M., Aksenova, M., Cao, A., Czeizel, A., Claustres, M., de la Rúa, C., Ferrari, M., et al. (1998). The deltaccr5 mutation conferring protection against HIV-1 in Caucasian populations has a single and recent origin in Northeastern Europe. Hum. Mol. Genet. *7*, 399–406.

Lindo, J., Huerta-Sánchez, E., Nakagome, S., Rasmussen, M., Petzelt, B., Mitchell, J., Cybulski, J.S., Willerslev, E., DeGiorgio, M., and Malhi, R.S.

(2016). A time transect of exomes from a Native American population before and after European contact. Nat. Commun. 7, 13175.

Liu, S., Huang, S., Chen, F., Zhao, L., Yuan, Y., Francis, S.S., Fang, L., Li, Z., Lin, L., Liu, R., et al. (2018). Genomic Analyses from Non-invasive Prenatal Testing Reveal Genetic Associations, Patterns of Viral Infections, and Chinese Population History. Cell 175, 347–359.e14.

Lohmueller, K.E. (2014). The impact of population demography and selection on the genetic architecture of complex traits. PLoS Genet. 10, e1004379.

MacArthur, D.G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., Jostins, L., Habegger, L., Pickrell, J.K., Montgomery, S.B., et al.; 1000 Genomes Project Consortium (2012). A systematic survey of loss-of-function variants in human protein-coding genes. Science 335, 823–828.

Macholdt, E., Lede, V., Barbieri, C., Mpoloka, S.W., Chen, H., Slatkin, M., Pakendorf, B., and Stoneking, M. (2014). Tracing pastoralist migrations to southern Africa with lactase persistence alleles. Curr. Biol. 24, 875–879.

Malaspinas, A.-S., Westaway, M.C., Muller, C., Sousa, V.C., Lao, O., Alves, I., Bergström, A., Athanasiadis, G., Cheng, J.Y., Crawford, J.E., et al. (2016). A genomic history of Aboriginal Australia. Nature 538, 207–214.

Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., Tandon, A., et al. (2016). The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature 538, 201–206.

Mancuso, N., Rohland, N., Rand, K.A., Tandon, A., Allen, A., Quinque, D., Mallick, S., Li, H., Stram, A., Sheng, X., et al.; PRACTICAL consortium (2016). The contribution of rare variation to prostate cancer heritability. Nat. Genet. 48, 30–35.

Martin, A.R., Gignoux, C.R., Walters, R.K., Wojcik, G.L., Neale, B.M., Gravel, S., Daly, M.J., Bustamante, C.D., and Kenny, E.E. (2017). Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. Am. J. Hum. Genet. 100, 635–649.

Martin, A.R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B.M., and Daly, M.J. (2018). Hidden "risk" in polygenic scores: clinical use today could exacerbate health disparities. bioRxiv. https://doi.org/10.1101/441261.

Mathieson, S., and Mathieson, I. (2018). FADS1 and the timing of human adaptation to agriculture. Mol. Biol. Evol. 35, 2957–2970.

McColl, H., Racimo, F., Vinner, L., Demeter, F., Gakuhari, T., Moreno-Mayar, J.V., van Driem, G., Gram Wilken, U., Seguin-Orlando, A., de la Fuente Castro, C., et al. (2018). The prehistoric peopling of Southeast Asia. Science 361, 88–92.

McKenna, H., and Martin, D. (2016). Surviving physiological stress: Can insights into human adaptation to austere environments be applied to the critical care unit? Trends in Anaesthesia and Critical Care 11, 6–13.

McManus, K.F., Taravella, A.M., Henn, B.M., Bustamante, C.D., Sikora, M., and Cornejo, O.E. (2017). Population genetic analysis of the DARC locus (Duffy) reveals adaptation from standing variation associated with malaria resistance in humans. PLoS Genet. 13, e1006560.

Ménard, D., Barnadas, C., Bouchier, C., Henry-Halldin, C., Gray, L.R., Ratsimbasoa, A., Thonier, V., Carod, J.-F., Domarle, O., Colin, Y., et al. (2010). Plasmodium vivax clinical malaria is commonly observed in Duffy-negative Malagasy people. Proc. Natl. Acad. Sci. USA 107, 5967–5971.

Moltke, I., Grarup, N., Jørgensen, M.E., Bjerregaard, P., Treebak, J.T., Fumagalli, M., Korneliussen, T.S., Andersen, M.A., Nielsen, T.S., Krarup, N.T., et al. (2014). A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes. Nature 512, 190–193.

Mühlemann, B., Jones, T.C., Damgaard, P.B., Allentoft, M.E., Shevnina, I., Logvin, A., Usmanova, E., Panyushkina, I.P., Boldgiv, B., Bazartseren, T., et al. (2018). Ancient hepatitis B viruses from the Bronze Age to the Medieval period. Nature 557, 418–423.

Nagai, A., Hirata, M., Kamatani, Y., Muto, K., Matsuda, K., Kiyohara, Y., Ninomiya, T., Tamakoshi, A., Yamagata, Z., Mushiroda, T., et al.; BioBank Japan Cooperative Hospital Group (2017). Overview of the BioBank Japan Project: Study design and profile. J. Epidemiol. 27 (3S), S2–S8.

Nédélec, Y., Sanz, J., Baharian, G., Szpiech, Z.A., Pacis, A., Dumaine, A., Grenier, J.-C., Freiman, A., Sams, A.J., Hebert, S., et al. (2016). Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. Cell 167, 657–669.e21.

Neel, J.V. (1962). Diabetes mellitus: a "thrifty" genotype rendered detrimental by "progress"? Am. J. Hum. Genet. 14, 353–362.

Neel, J.V. (1999). The "thrifty genotype" in 1998. Nutr. Rev. 57, S2–S9.

Nielsen, R., Akey, J.M., Jakobsson, M., Pritchard, J.K., Tishkoff, S., and Willerslev, E. (2017). Tracing the peopling of the world through genomics. Nature 541, 302–310.

Novembre, J., and Barton, N.H. (2018). Tread Lightly Interpreting Polygenic Tests of Selection. Genetics 208, 1351–1355.

Novembre, J., and Han, E. (2012). Human population structure and the adaptive response to pathogen-induced selection pressures. Philos. Trans. R. Soc. Lond. B Biol. Sci. 367, 878–886.

Novembre, J., Galvani, A.P., and Slatkin, M. (2005). The geographic spread of the CCR5 Δ32 HIV-resistance allele. PLoS Biol. 3, e339.

Pagani, L., Lawson, D.J., Jagoda, E., Mörseburg, A., Eriksson, A., Mitt, M., Clemente, F., Hudjashov, G., DeGiorgio, M., Saag, L., et al. (2016). Genomic analyses inform on migration events during the peopling of Eurasia. Nature 538, 238–242.

Payne, M., Rupar, C.A., Siu, G.M., and Siu, V.M. (2011). Amish, mennonite, and hutterite genetic disorder database. Paediatr. Child Health 16, e23–e24.

Pedersen, C.T., Lohmueller, K.E., Grarup, N., Bjerregaard, P., Hansen, T., Siegismund, H.R., Moltke, I., and Albrechtsen, A. (2017). The Effect of an Extreme and Prolonged Population Bottleneck on Patterns of Deleterious Variation: Insights from the Greenlandic Inuit. Genetics 205, 787–801.

Peltonen, L., Palotie, A., and Lange, K. (2000). Use of population isolates for mapping complex traits. Nat. Rev. Genet. 1, 182–190.

Peng, Y., Yang, Z., Zhang, H., Cui, C., Qi, X., Luo, X., Tao, X., Wu, T., Ouzhuluobu, Basang, et al.. (2011). Genetic variations in Tibetan populations and high-altitude adaptation at the Himalayas. Mol. Biol. Evol. 28, 1075–1081.

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.

Pierce, B.L., Tong, L., Argos, M., Gao, J., Farzana, J., Roy, S., Paul-Brutus, R., Rahaman, R., Rakibuz-Zaman, M., Parvez, F., et al. (2013). Arsenic metabolism efficiency has a causal role in arsenic toxicity: Mendelian randomization and gene-environment interaction. Int. J. Epidemiol. 42, 1862–1871.

Prado-Martinez, J., Sudmant, P.H., Kidd, J.M., Li, H., Kelley, J.L., Lorente-Galdos, B., Veeramah, K.R., Woerner, A.E., O'Connor, T.D., Santpere, G., et al. (2013). Great ape genetic diversity and population history. Nature 499, 471–475.

Pritchard, J.K., Pickrell, J.K., and Coop, G. (2010). The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. Curr. Biol. 20, R208–R215.

Prüfer, K., de Filippo, C., Grote, S., Mafessoni, F., Korlević, P., Hajdinjak, M., Vernot, B., Skov, L., Hsieh, P., Peyrégne, S., et al. (2017). A high-coverage Neandertal genome from Vindija Cave in Croatia. Science 358, 655–658.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Getet. 81, 559–575.

Quach, H., Rotival, M., Pothlichet, J., Loh, Y.E., Dannemann, M., Zidane, N., Laval, G., Patin, E., Harmant, C., Lopez, M., et al. (2016). Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. Cell 167, 643–656.e17.

Racimo, F., Marnetto, D., and Huerta-Sánchez, E. (2017). Signatures of Archaic Adaptive Introgression in Present-Day Human Populations. Mol. Biol. Evol. 34, 296–317.

Racimo, F., Berg, J.J., and Pickrell, J.K. (2018). Detecting Polygenic Adaptation in Admixture Graphs. Genetics 208, 1565–1584.

Ramachandran, S., Deshpande, O., Roseman, C.C., Rosenberg, N.A., Feldman, M.W., and Cavalli-Sforza, L.L. (2005). Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. Proc. Natl. Acad. Sci. USA *102*, 15942–15947.

Rasmussen, S., Allentoft, M.E., Nielsen, K., Orlando, L., Sikora, M., Sjögren, K.-G., Pedersen, A.G., Schubert, M., Van Dam, A., Kapel, C.M.O., et al. (2015). Early divergent strains of Yersinia pestis in Eurasia 5,000 years ago. Cell *163*, 571–582.

Refoyo-Martínez, A., da Fonseca, R.R., Halldórsdóttir, K., Árnason, E., Mailund, T., and Racimo, F. (2018). Identifying loci under positive selection in complex population histories. bioRxiv. https://doi.org/10.1101/453092.

Reginsson, G.W., Ingason, A., Euesden, J., Bjornsdottir, G., Olafsson, S., Sigurdsson, E., Oskarsson, H., Tyrfingsson, T., Runarsdottir, V., Hansdottir, I., et al. (2018). Polygenic risk scores for schizophrenia and bipolar disorder associate with addiction. Addict. Biol. *23*, 485–492.

Rintamäki, H. (2007). Human responses to cold. Alaska Med. *49* (2, Suppl), 29–31.

Rivas, M.A., Avila, B.E., Koskela, J., Huang, H., Stevens, C., Pirinen, M., Haritunians, T., Neale, B.M., Kurki, M., Ganna, A., et al.; International IBD Genetics Consortium; NIDDK IBD Genetics Consortium; T2D-GENES Consortium (2018). Insights into the genetic epidemiology of Crohn's and rare diseases in the Ashkenazi Jewish population. PLoS Genet. *14*, e1007329.

Robinson, M.R., Hemani, G., Medina-Gomez, C., Mezzavilla, M., Esko, T., Shakhbazov, K., Powell, J.E., Vinkhuyzen, A., Berndt, S.I., Gustafsson, S., et al. (2015). Population genetic differentiation of height and body mass index across Europe. Nat. Genet. *47*, 1357–1362.

Sabeti, P.C., Walsh, E., Schaffner, S.F., Varilly, P., Fry, B., Hutcheson, H.B., Cullen, M., Mikkelsen, T.S., Roy, J., Patterson, N., et al. (2005). The case for selection at CCR5-Δ32. PLoS Biol. *3*, e378.

Samson, M., Libert, F., Doranz, B.J., Rucker, J., Liesnard, C., Farber, C.M., Saragosti, S., Lapoumeroulie, C., Cognaux, J., Forceille, C., et al. (1996). Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. Nature *382*, 722–725.

Sanjak, J.S., Long, A.D., and Thornton, K.R. (2017). A Model of Compound Heterozygous, Loss-of-Function Alleles Is Broadly Consistent with Observations from Complex-Disease GWAS Datasets. PLoS Genet. *13*, e1006573.

Sanjak, J.S., Sidorenko, J., Robinson, M.R., Thornton, K.R., and Visscher, P.M. (2018). Evidence of directional and stabilizing selection in contemporary humans. Proc. Natl. Acad. Sci. USA *115*, 151–156.

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.

Scheinfeldt, L.B., Soi, S., Thompson, S., Ranciaro, A., Woldemeskel, D., Beggs, W., Lambert, C., Jarvis, J.P., Abate, D., Belay, G., and Tishkoff, S.A. (2012). Genetic adaptation to high altitude in the Ethiopian highlands. Genome Biol. *13*, R1.

Schuenemann, V.J., Singh, P., Mendum, T.A., Krause-Kyora, B., Jäger, G., Bos, K.I., Herbig, A., Economou, C., Benjak, A., Busso, P., et al. (2013). Genome-wide comparison of medieval and modern Mycobacterium leprae. Science *341*, 179–183.

Scriver, C.R. (2001). Human genetics: lessons from Quebec populations. Annu. Rev. Genomics Hum. Genet. *2*, 69–101.

Ségurel, L., and Bon, C. (2017). On the Evolution of Lactase Persistence in Humans. Annu. Rev. Genomics Hum. Genet. *18*, 297–319.

Sikora, M., Carpenter, M.L., Moreno-Estrada, A., Henn, B.M., Underhill, P.A., Sánchez-Quinto, F., Zara, I., Pitzalis, M., Sidore, C., Busonero, F., et al. (2014). Population genomic analysis of ancient and modern genomes yields new insights into the genetic ancestry of the Tyrolean Iceman and the genetic structure of Europe. PLoS Genet. *10*, e1004353.

Sikora, M., Seguin-Orlando, A., Sousa, V.C., Albrechtsen, A., Korneliussen, T., Ko, A., Rasmussen, S., Dupanloup, I., Nigst, P.R., Bosch, M.D., et al. (2017). Ancient genomes show social and reproductive behavior of early Upper Paleolithic foragers. Science *358*, 659–662.

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.

Skoglund, P., and Mathieson, I. (2018). Ancient Genomics of Modern Humans: The First Decade. Annu. Rev. Genomics Hum. Genet. *19*, 381–404.

Skoglund, P., Malmström, H., Raghavan, M., Storå, J., Hall, P., Willerslev, E., Gilbert, M.T.P., Götherström, A., and Jakobsson, M. (2012). Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. Science *336*, 466–469.

Skoglund, P., Malmström, H., Omrak, A., Raghavan, M., Valdiosera, C., Günther, T., Hall, P., Tambets, K., Parik, J., Sjögren, K.-G., et al. (2014). Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers. Science *344*, 747–750.

Skotte, L., Koch, A., Yakimov, V., Zhou, S., Søborg, B., Andersson, M., Michelsen, S.W., Navne, J.E., Mistry, J.M., Dion, P.A., et al. (2017). *CPT1A* Missense Mutation Associated With Fatty Acid Metabolism and Reduced Height in Greenlanders. Circ Cardiovasc Genet *10*, e001618.

Smith, J.M., and Haigh, J. (1974). The hitch-hiking effect of a favourable gene. Genet. Res. *23*, 23–35.

Sohail, M., Maier, R.M., Ganna, A., Bloemendal, A., Martin, A.R., Turchin, M.C., Chiang, C.W.K., Hirschhorn, J.N., Daly, M.J., Patterson, N., et al. (2018). Signals of polygenic adaptation on height have been overestimated due to uncorrected population structure in genome-wide association studies. bioRxiv. https://doi.org/10.1101/355057.

Southam, L., Gilly, A., Süveges, D., Farmaki, A.-E., Schwartzentruber, J., Tachmazidou, I., Matchan, A., Rayner, N.W., Tsafantakis, E., Karaleftheri, M., et al. (2017). Whole genome sequencing and imputation in isolated populations identify genetic associations with medically-relevant complex traits. Nat. Commun. *8*, 15606.

Stearns, S.C. (1989). Trade-Offs in Life-History Evolution. Funct. Ecol. *3*, 259–268.

Strachan, D.P. (1989). Hay fever, hygiene, and household size. BMJ *299*, 1259–1260.

Tishkoff, S.A., Reed, F.A., Ranciaro, A., Voight, B.F., Babbitt, C.C., Silverman, J.S., Powell, K., Mortensen, H.M., Hirbo, J.B., Osman, M., et al. (2007). Convergent adaptation of human lactase persistence in Africa and Europe. Nat. Genet. *39*, 31–40.

Torkamani, A., Wineinger, N.E., and Topol, E.J. (2018). The personal and clinical utility of polygenic risk scores. Nat. Rev. Genet. *19*, 581–590.

Trinkaus, E. (2018). An abundance of developmental anomalies and abnormalities in Pleistocene people. Proc. Natl. Acad. Sci. USA *115*, 11941–11946.

Uricchio, L.H., Zaitlen, N.A., Ye, C.J., Witte, J.S., and Hernandez, R.D. (2016). Selection and explosive growth alter genetic architecture and hamper the detection of causal rare variants. Genome Res. *26*, 863–873.

Vernot, B., Tucci, S., Kelso, J., Schraiber, J.G., Wolf, A.B., Gittelman, R.M., Dannemann, M., Grote, S., McCoy, R.C., Norton, H., et al. (2016). Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. Science *352*, 235–239.

Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., and Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. Am. J. Hum. Genet. *101*, 5–22.

West, J.B. (2006). Human responses to extreme altitudes. Integr. Comp. Biol. *46*, 25–34.

Winkler, C.A., Nelson, G.W., and Smith, M.W. (2010). Admixture mapping comes of age. Annu. Rev. Genomics Hum. Genet. *11*, 65–89.

Wojcik, G., Graff, M., Nishimura, K.K., Tao, R., Haessler, J., Gignoux, C.R., Highland, H.M., Patel, Y.M., Sorokin, E.P., Avery, C.L., et al. (2018). The PAGE Study: How Genetic Diversity Improves Our Understanding of the Architecture of Complex Traits. bioRxiv. https://doi.org/10.1101/188094.

Woolhouse, M.E.J., Webster, J.P., Domingo, E., Charlesworth, B., and Levin, B.R. (2002). Biological and biomedical implications of the co-evolution of pathogens and their hosts. Nat. Genet. *32*, 569–577.

World Health Organization (2017). World Malaria Report 2016 (Geneva: World Health Organization).

World Health Organization (2018). Global status report on alcohol and health 2018 (Geneva: World Health Organization).

Xue, Y., Mezzavilla, M., Haber, M., McCarthy, S., Chen, Y., Narasimhan, V., Gilly, A., Ayub, Q., Colonna, V., Southam, L., et al. (2017). Enrichment of low-frequency functional variants revealed by whole-genome sequencing of multiple isolated European populations. Nat. Commun. *8*, 15927.

Yang, J., Bakshi, A., Zhu, Z., Hemani, G., Vinkhuyzen, A.A.E., Lee, S.H., Robinson, M.R., Perry, J.R.B., Nolte, I.M., van Vliet-Ostaptchouk, J.V., et al.; LifeLines Cohort Study (2015). Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat. Genet. *47*, 1114–1120.

Ye, K., Gao, F., Wang, D., Bar-Yosef, O., and Keinan, A. (2017). Dietary adaptation of FADS 1121 genes in Europe varied across time and geography. Nat Ecol Evol *1*, 167.

Yeaman, S., and Whitlock, M.C. (2011). The genetic architecture of adaptation under migration–selection balance. Evolution *65*, 1897–1911.

Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z.X.P., Pool, J.E., Xu, X., Jiang, H., Vinckenbosch, N., Korneliussen, T.S., et al. (2010). Sequencing of 50 human exomes reveals adaptation to high altitude. Science *329*, 75–78.

Zeng, J., de Vlaming, R., Wu, Y., Robinson, M.R., Lloyd-Jones, L.R., Yengo, L., Yap, C.X., Xue, A., Sidorenko, J., McRae, A.F., et al. (2018). Signatures of negative selection in the genetic architecture of human complex traits. Nat. Genet. *50*, 746–753.