Ancient DNA reveals that few GWAS loci have been strongly selected during recent human history

Corinne N. Simonti[1] and Joseph Lachance[1*]

[1]School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA

*Corresponding author: joseph.lachance@biology.gatech.edu

Keywords: ancient DNA, evolutionary medicine, genetics, GWAS, natural selection

Abstract

Genetic data from ancient humans has provided new evidence in the study of loci thought to be under historic selection, and thus is a powerful tool for identifying instances of selection that might be missed by methods that use present-day samples alone. Using a curated set of disease-associated variants from the NHGRI-EBI GWAS Catalog, we provide an analysis to identify disease-associated variants that bear signatures of selection over time. After accounting for the fact that not every ancient individual contributed equally to modern genomes, a Bayesian inference method was used to infer allele frequency trajectories over time and determine which disease-associated loci exhibit signatures of natural selection. Of the 2,709 variants analyzed in this study, 895 show at least a weak signature of selection ($|s| > 0.001$), including multiple variants that are introgressed from Neanderthals. However, only nine disease-associated variants show a signature of strong selection ($|s| > 0.01$). Additionally, we find that many risk-associated alleles have increased in frequency during the past 10,000 years. Overall, we find that disease-associated variants from GWAS are governed by nearly neutral evolution. Exceptions to this broad pattern include GWAS loci that protect against asthma and variants in MHC genes. Ancient samples allow us an unprecedented look at how our species has changed over time, and our results represent an important early step in using this new source of data to better understand the evolution of hereditary disease risks.

Introduction

Many studies have been conducted to understand the genetic basis of diseases and other traits. Indeed, thousands of disease loci have been implicated in genome-wide association studies (GWAS) across human populations (Welter et al., 2014). Due to a balance between mutation and selection, alleles that are associated with Mendelian diseases tend to be rare (Weghorn et al., 2019). By contrast, alleles that are associated with complex diseases often segregate at intermediate frequencies (Timpson et al., 2018). GWAS loci are also enriched for derived, as opposed to ancestral, alleles that increase the risks of complex polygenic diseases (Gorlova et al., 2012; Lachance, 2010). There has been a great deal of interest in understanding how allele frequencies—and thus, traits—have changed over time; however, until recently, only modern genomes have been available. Analysis of modern human genomes has revealed that genes that confer resistance to infectious diseases have been targets of recent selection (Karlsson et al., 2014), and that natural selection contributes to differences in hereditary disease risks between human populations (Lachance et al., 2018; Marigorta et al., 2011). Classic debates about the relative importance of natural selection and genetic drift have also been rekindled (Jensen et al., 2019; Kern and Hahn, 2018). It has also been proposed that modern medicine has led to the relaxation of selection against mildly deleterious mutations and an increase in genetic load over recent human history (Lynch, 2016).

During the past decade, an increasing number of ancient genomes have been sequenced, some to very high read depth (Fu et al., 2014; Mathieson et al., 2015; Meyer et al., 2012; Narasimhan et al., 2019; Olalde et al., 2018; Prüfer et al., 2014). High quality genome sequences of our archaic hominin cousins have revealed previously unknown introgression events, and follow-up studies

have shown how these introgressed regions have impacted modern human health and traits (Dannemann et al., 2016; Gunz et al., 2018; Simonti et al., 2016). Ancient DNA has also revealed that selection on the fatty acid desaturase genes *FADS1* and *FADS2*, which are relevant to lipid disorders, did not coincide with the Neolithic transition and the discovery of agriculture (Mathieson and Mathieson, 2018). Application of polygenic risk scores to ancient genomes has revealed that genetic disease risks have changed over the last 10,000 years (Berens et al., 2017). Furthermore, ancient human genomes have revealed that many variants associated with disease risk have increased in frequency during recent history (Aris-Brosou, 2019). Because ancient DNA enables times series data to be analyzed, it is now possible to determine the evolutionary histories of individual disease-associated loci. Large allele frequency changes during the last 10,000 years are indicative of disease loci that are under natural selection, while relatively flat allele frequency trajectories are indicative of disease loci that are evolving neutrally (Schraiber et al., 2016).

An open question in evolutionary medicine is whether the effects of genetic variation on disease risks are decoupled from fitness effects (Rodríguez et al., 2014). Here, we use 143 ancient human samples from 18 previously published studies to examine the role of selection on variants from the NHGRI-EBI GWAS Catalog. Most of the ancient samples analyzed here lived during the past 10,000 years. First, we correct for whether ancient genomes contribute to modern human populations. We then use time series data to infer allele frequency trajectories and selection coefficients for a curated set of 2,709 different disease-associated variants. We also examine whether protective alleles are more likely to be positively selected, and identify which diseases are enriched for signatures of weak selection. Collectively, these analyses reveal that most disease-associated loci are governed by nearly neutral evolution.

Results

Most GWAS variants do not show signatures of strong selection

For each disease-associated locus, we employed a Bayesian MCMC method (Schraiber et al., 2016) to infer allele frequency trajectories and selection coefficients that match time series data from ancient genomes. However, inference of natural selection requires an understanding of population continuity, or how much ancient individuals have contributed to modern genomes, if at all. As previous studies have found differences in contribution based on lifestyle (Haak et al., 2015), we grouped our ancient samples into one of three lifestyles: hunter-gatherers, agriculturalists, and pastoralists. We then calculated maximum possible genetic contribution of each of these groups to each variant using the method developed by Sjödin et al. (2014). Using this information, we could ensure we did not include ancient individuals who might skew allele frequencies in a way that did not reflect ancient demography (Figure S1).

Most GWAS loci have flat allele trajectories and selection coefficients that are close to zero. Of the 2,709 variants analyzed here, we find that 895 show at least a weak signature of selection ($|s| > 0.001$) and only 9 show a signature of strong selection ($|s| > 0.01$; Table 1). We also examined whether protective alleles were more likely to increase in frequency during recent history than risk-increasing alleles. A slight majority (52.6%) of putatively selected loci have protective alleles with positive selection coefficients. However, this trend was not statistically significant (471 of 895 variants, binomial test, $P = 6.21 \times 10^{-2}$). We also found that derived alleles are more likely to have positive selection coefficients than ancestral alleles at GWAS loci (528 of 895 variants; binomial test, $P < 10^{-7}$). Taken together, our findings reveal that GWAS loci are largely governed by nearly neutral evolution.

Despite most variants appearing to evolve neutrally, exceptions exist for some disease-associated loci (Supplemental Table S1). For example, six of the nine variants under strong selection ($s > 0.01$) are associated with autoimmune diseases. One of them (rs13277113) is associated with systemic lupus erythematosus (Hom et al., 2008), and has been found by the GTEx Project (v8) to significantly affect expression of *BLK*, a gene involved in B cell development (Gauld and Cambier, 2004). Another of these variants (rs2270450) is associated with developing Hashimoto thyroiditis versus Graves' disease (Oryoji et al., 2015). Curiously, this variant is significantly associated with expression of *TDRD6*, a gene involved in spermiogenesis as well as autoimmune polyendocrine syndrome (Akpınar et al., 2017; Bensing et al., 2007). The other four strongly selected variants associated with autoimmune diseases are found within the MHC region, which is known to be dense with rapidly evolving genes critical to immune function.

Table 1. Selection coefficients inferred from ancient genomes

Variant type	Total number of variants	$\|s\| > 0.001$ (weak selection)	Protective allele selected	$\|s\| > 0.01$ (strong selection)	Protective allele selected
All	2,709	895 (33.0%)	471 of 895 (52.6%)	9 (0.3%)	4 of 9 (44.4%)
MHC	148	76 (51.4%)	36 of 76 (47.4%)	4 (2.7%)	0 of 4 (0%)
Introgressed	24	8 (33.3%)	4 of 8 (50.0%)	0 (0.0%)	-

Archaic introgression introduced some selected alleles at GWAS loci

Disentangling the effects of selection and archaic introgression can be difficult if one is analyzing modern genomes (Racimo et al., 2016). However, time series data from ancient genomes can sidestep many of these difficulties. Of the 2,709 disease-associated loci analyzed here, 24 contain introgressed alleles from Neanderthals. Of these 24 loci, nine introgressed variants have signatures of weak selection in the recent past (Figure 1). Some introgressed alleles appear to have reached moderate allele frequencies prior to being negatively selected, while other introgressed alleles appear to have only been positively selected during recent human history. Overall, we find that disease-associated alleles that have a Neanderthal origin have similar selection coefficients to other disease-associated alleles from the GWAS Catalog (Table 1). This is not wholly surprising, given that purging of strongly deleterious introgressed alleles occurred shortly after admixture (Petr et al., 2019), and introgressed alleles that are able to persist to the present day will have managed to avoid being filtered out due to Dobzhansky-Muller incompatibilities or other hybrid fitness effects.

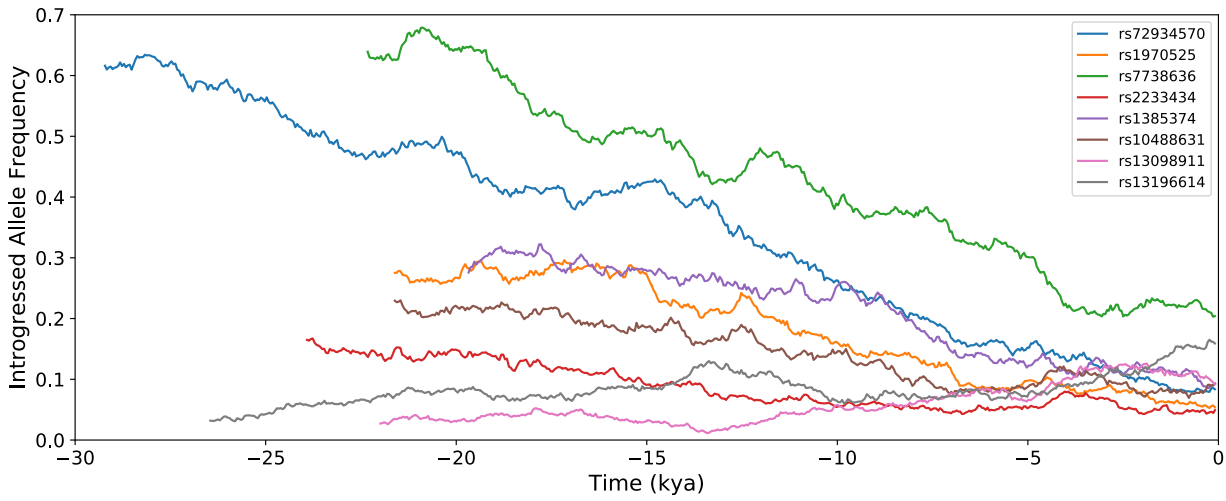

Figure 1. Some introgressed disease alleles are under weak selection. Inferred allele frequency trajectories of introgressed alleles with $|s| > 0.001$ are plotted here

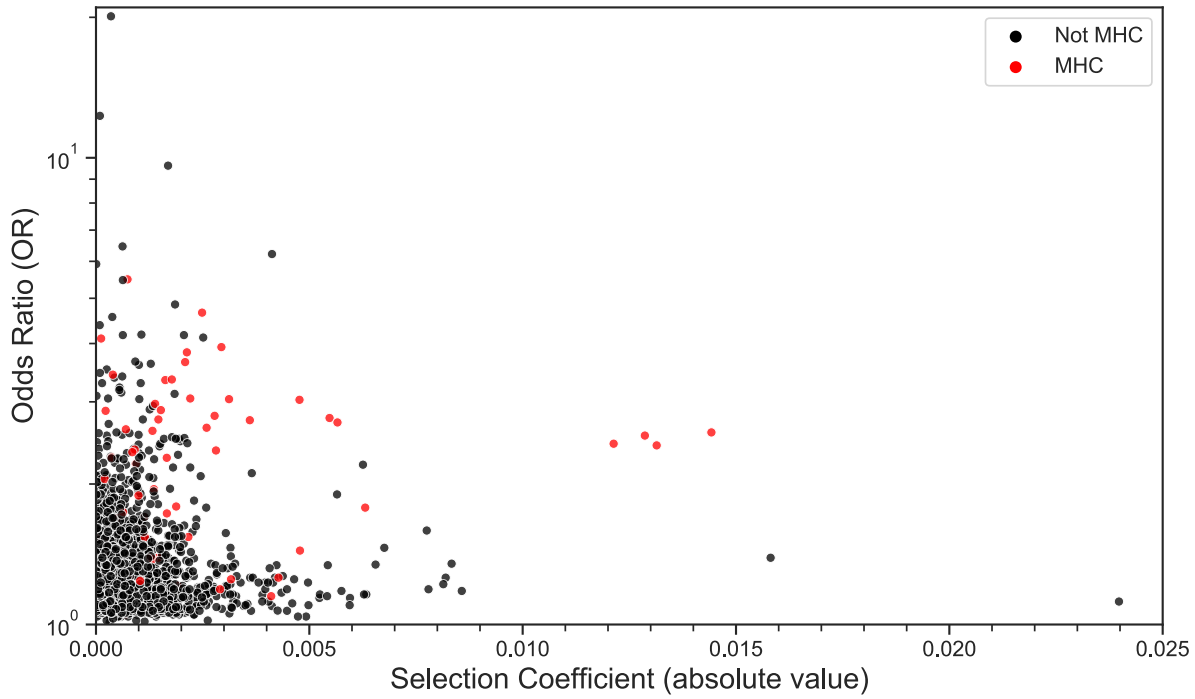

Figure 2. **GWAS effect sizes are generally unrelated to selection coefficients.** We plotted the odds ratios (OR) of GWAS loci verses inferred selection coefficients for 1,910 disease-associated variants that were identified in a GWAS that used samples of European ancestry (MHC variants: red, non-MHC variants: black).

Most selected alleles at disease loci have small effect sizes

We examined whether selection coefficients are correlated with disease-related effect sizes. Given that our time series dataset exclusively contains samples from Europe and effect sizes are contingent on a host of factors including study population and environmental factors, this analysis focused on the subset of 1,910 variants that were identified in a GWAS that used samples of European ancestry (Figure 2). Overall, we find that there is no relationship between odds ratio

(OR) and magnitude of s for non-MHC variants (r = -0.012). This pattern arises because many GWAS Catalog traits are largely irrelevant to fitness. For example, the variant with the highest OR analyzed here (rs3825942, OR=20.1) is associated with glaucoma, a disease which normally strikes long after reproductive age. Although the variants in Figure 2 are associated with a heterogenous set of diseases, our results suggest that the selection coefficients are largely decoupled from the effects of GWAS loci on disease risks.

Variants in the MHC region show stronger signatures of selection

Compared to other disease-associated alleles, MHC variants show a strikingly different pattern. Disease-associated variants in the MHC region on chromosome 6 have larger effects sizes than disease-associated variants that are not in the MHC region (mean MHC OR: 2.11, mean non-MHC OR: 1.36, t test $P = 4.12 \times 10^{-19}$). Disease-associated variants in MHC genes also show a modest positive correlation between s and OR (r = 0.167). Variants located in the MHC region were significantly more likely to show signatures of weak selection than GWAS variants found in other parts of the human genome (Table 1, hypergeometric $P < 10^{-6}$). Disease-associated variants in the MHC region were also significantly depleted of selection for protective alleles (hypergeometric $P < 10^{-6}$). Given the highly pleiotropic nature of the immune system, it is possible that alleles that increase risk for some immune diseases have been historically protective against others unrepresented in the GWAS Catalog (i.e., infectious diseases).

Disease-specific tests for enrichment of selection signatures

Although we caution against overinterpreting disease-specific results (due to genetic hitchhiking and the pleiotropic nature of variants analyzed in this paper), multiple diseases were enriched for

signatures of weak selection. Idiopathic membranous nephropathy, an autoimmune disease that severely impacts kidney function, showed the strongest signature of selection, with 17 of 19 variants under at least weak selection (Figure 3A; hypergeometric $P < 1.00 \times 10^{-6}$). Unsurprising given the strong immune component to this disease, all but one of the variants associated with this disease are found in the MHC region, which likely contributes to enrichment for selected alleles that are associated with idiopathic membranous nephropathy. Curiously, alleles that protect against this disease have tended to decrease in frequency during recent human history (Figure 3B; hypergeometric $P = 4.76 \times 10^{-4}$). As stated above, this result could represent either genetic hitchhiking or pleiotropic effects. The next strongest enrichment signal was from asthma, with 26 of the 49 variants associating with it being under at least weak selection (Figure 3A; hypergeometric $P = 2.12 \times 10^{-3}$). We found that alleles that protect against asthma were strongly enriched for being under positive selection (Figure 3B; 20 of 26 selected alleles were protective; hypergeometric $P = 8.94 \times 10^{-3}$). This result was robust to the exclusion of MHC variants. A third disease that showed enrichment for signatures of weak selection was acne, with 7 of its 10 associated variants under selection (hypergeometric $P = 1.79 \times 10^{-2}$; Supplemental Table S2).

In addition, we found three diseases that were significantly depleted for signatures of weak selection (Supplemental Table S2). The strongest signal was for breast cancer (Figure 3B; 21 of 96 variants; hypergeometric $P = 1.20 \times 10^{-2}$), followed by major depressive disorder (10 of 55 associated variants; hypergeometric $P = 1.27 \times 10^{-2}$), and addiction (3 of 23 variants; hypergeometric $P = 2.96 \times 10^{-2}$). Taken together, these results reveal that not all diseases are equally affected by selection, even when selection does not appear to be broadly acting.

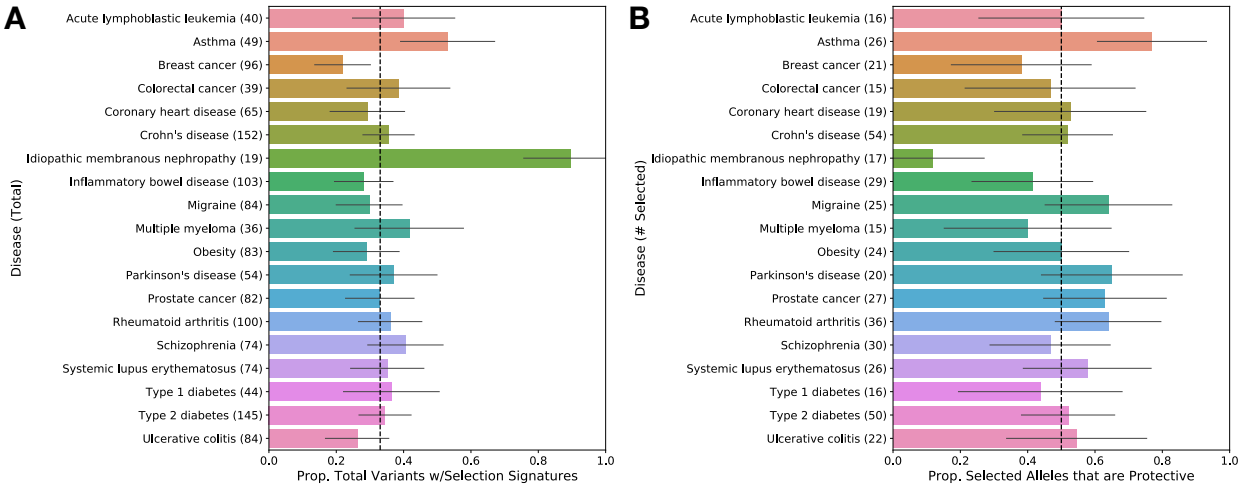

Figure 3. **Few diseases are enriched for signatures of selection.** Dashed vertical lines indicate mean values for all disease-associated alleles analyzed in this study. **(A)** For diseases with at least 15 variants under weak selection, the proportion of variants with $|s| > 0.001$ is plotted. Details about other diseases can be found in Supplemental Table S2. **(B)** The proportion of selected variants where the protective allele has a positive selection coefficient.

Given that severity and when a disease manifests determine its fitness impact, we also examined whether diseases with an early age of onset have different characteristics than diseases with a late age of onset. Overall, 55.9% of alleles that are associated with increased disease risks in our dataset were derived, as opposed to ancestral (binomial $P < 1.00 \times 10^{-6}$). This trend was more pronounced in diseases whose onset was before the age of 18 (hypergeometric $P = 3.50 \times 10^{-5}$; Figure 4A). Despite this effect, age of onset does not seem to impact the likelihood of a variant being under selection (Figure 4B). However, there is a slight bias in favor of the protective allele for both the 0-18 age of onset (binomial $P = 0.047$) and the 19-40 age of onset disease variants (binomial $P = 0.042$; Figure 4C).

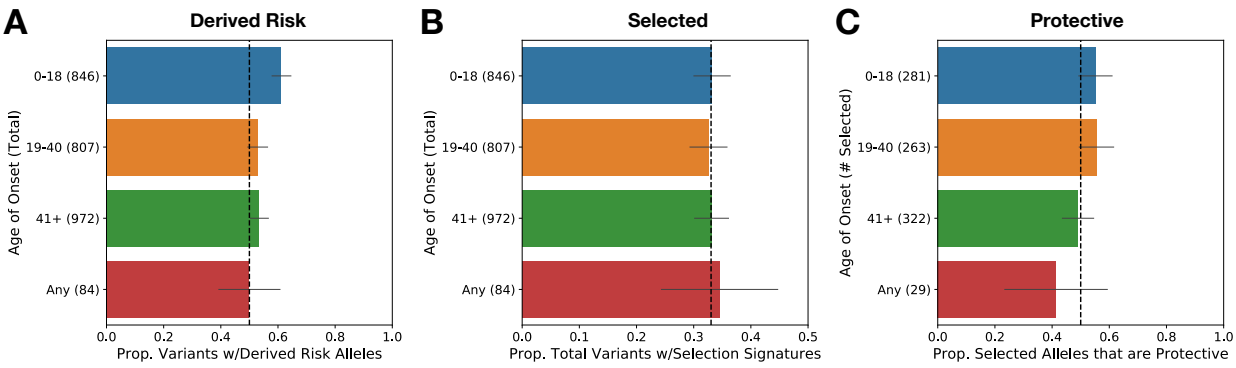

Figure 4. Comparisons between GWAS loci that are associated with diseases that have different ages of onset. (A) Proportion of variants where the risk-increasing allele is the derived allele. The dashed line represents the null expectation (50%). **(B)** Proportion of variants under at least weak selection ($|s| > 0.001$). The dashed line represents the overall variants under weak selection (33.0%). **(C)** Proportion of selected variants where protective alleles have increased in frequency over recent human history. The dashed line represents the null expectation (50%).

Discussion

We find that the majority of GWAS variants do not appear to be under strong selection, and that selection coefficients (s) and odds ratios (OR) are uncorrelated. Allele frequency changes at many disease-associated loci may be due to the indirect effects of genetic hitchhiking, i.e., selection acting at closely linked loci. Because of this, our results suggest an "upper bound" for the amount of selection experienced by GWAS variants. In general, we find that our results reinforce the notion that diseases do not necessarily impact fitness. Our study also demonstrates the utility of time series data to infer the evolutionary histories of individual disease-associated loci.

Disease phenotypes are obvious targets for natural selection, so the general lack of variants under selection in this set of GWAS loci at first seems unexpected. However, a previous study estimating selection coefficients from C-scores showed a remarkably similar proportion of GWAS variants under weak selection (Racimo and Schraiber, 2014). Indeed, there are many reasons why we would not expect many of these variants to be under strong directional selection in the timescale examined. First, many of these diseases are unlikely to impact fitness, a scenario supported by our result that OR and the magnitude of s are uncorrelated for most variants. Of the 178 diseases considered here, 20 are late-onset (age of onset of 60 or later) and while they shorten lifespans, they do so well after reproductive age. Barring strong effects from scenarios such as the grandmother hypothesis (Williams, 1957) and presuming these variants substantially affect no other fitness-impacting traits, this makes this group of disease variants essentially invisible to selection. The fact that these diseases show no marked depletion in the proportion of variants under weak selection when compared to diseases with earlier ages of onset could imply that genetic hitchhiking or pleiotropy may be obscuring expected patterns of selection.

Another possibility is that selection has acted on disease traits, but the polygenic nature of the disease means individual loci affecting risk might be close to neutrality. Indeed, this is expected as many GWAS variants are segregating at allele frequencies that no strongly deleterious variants could ever reach. Previous studies have demonstrated negative correlations between allele frequency and effect size, even when study power was taken into account (Park et al., 2011). This suggests that purifying selection constrains evolutionarily deleterious alleles, supporting the idea that variants in the GWAS Catalog have not historically had large fitness effects. Another possibility is that due to environmental changes, some of these diseases have become more

prevalent in recent history (so-called diseases of modernity). This would cause selection to only be able to act on them recently, outside the timescale of our analyses. These scenarios do not represent an exhaustive list of possibilities, nor are they mutually exclusive.

One exception to this overall pattern is that many putatively selected variants are found in the highly polymorphic MHC region of the genome. Genes in this region impact resistance to infectious disease, mate choice, kin recognition, reproductive success, among others (Bernatchez and Landry, 2003; Piertney and Oliver, 2006). As such, they have been frequent targets of recurring balancing and positive selection. When considered, these theories make it clear why GWAS variants overall would be depleted of signatures of selection, apart from MHC variants. It is worth noting the highly repetitive and highly polymorphic nature of this region makes genotyping in the MHC region notoriously difficult, so it is possible that this has introduced bias in our results (Kiyotani et al., 2017). However, there is no a priori reason why difficulty genotyping would consistently result in an excess of signal for selection.

Interestingly, diseases whose onset is during pre-reproductive or reproductive age are slightly more likely to have protective alleles that are positively selected. A prime example of this involves variants affecting asthma, which are strongly enriched for selection overall, and positive selection on protective alleles in particular. Asthma is a chronic inflammatory disorder of the airways primarily affecting children whose incidence has increased substantially over recent decades (Akinbami et al., 2015), which has been partially attributed to both the hygiene hypothesis as well as increasing air pollution through urbanization (Bowatte et al., 2015; WHO, 2019). These factors together suggest that the environmental risk for asthma has been on the rise over recent human

history, with commensurate increases in selective pressure resulting in the strong enrichment for positive selection on protective alleles that we observed.

It is worth mentioning that the approach taken here is subject to the limitations of the GWAS data to which it has been applied. While incredibly useful, GWAS are not easily conducted on certain diseases that likely carry great importance to fitness, such as infectious diseases or diseases and health events that result in patient death before they can be enrolled in a study (i.e., myocardial infarction, stroke). Our approach here is well suited to examining the common variants typically found in GWAS, but it is ill-suited to rare or family-specific variation. Additionally, the timescales we can examine using this approach are defined by the density and timeline of ancient samples. As many of our samples are from the last 10,000 years and the youngest is from ~1,208 years ago, very recent allele frequency shifts are difficult to detect. It is possible that some variants may have been under episodic selection, but our method is unable to detect selection if allele frequency trajectories are U-shaped.

Conclusion

Our study shows the utility of ancient genomic data and time series analyses in understanding the effects of selection versus neutral processes on complex disease risk. These results contribute to the growing body of data on how selection has shaped the human genome, as well as the importance of disambiguating fitness effects from disease status. It is also worth noting that incorporating information from ancient samples can help give context in situations where traditional markers of selection (linkage disequilibrium, singleton density, etc.) are obscured due to the history of the region. This is a particular source of trouble in regions such as haplotypes

introgressed from ancient hominins, whose history violates many assumptions of methods using modern data alone (Vernot and Akey, 2014). As more high-quality ancient genomes are collected from around the world, our understanding of evolutionary medicine will continue to grow.

Materials and Methods

GWAS variant and ancient DNA curation

Data was downloaded and curated as described in (Berens et al., 2017). Briefly, variants were downloaded from the GWAS catalog (accessed 28 September 2016) and associations with benign phenotypes removed. Specific variant associations were removed from consideration if two variants associated the same disease and were within 100 kb of each other, with a preference for maintaining the variant-phenotype association with the lower p-value. The risk allele for each variant was classified as derived or ancestral based on data from the 1000 Genomes Project (Auton et al., 2015). We used genomic data from 143 ancient individuals from 18 studies (see (Berens et al., 2017) for further details on the curation of this data). As time series data requires a continuous population, non-European ancient human samples (e.g., Mota and Anzick) and archaic hominin Neanderthal and Denisovan samples were not included from the larger set described in Berens et al (2017).

Age of onset determination and phenotype collapsing

When available, we used the disease information collated by Google from multiple data sources (for more information, see: https://support.google.com/websearch/answer/2364942). Onset was broken down into seven age ranges: 0-2, 3-5, 6-13, 14-18, 19-40, 41-60, and 60+. For each disease,

we used the age range determined by the highest frequency of ages affected. When multiple age ranges were equally likely, we used the youngest age range. For the diseases that Google did not have this information, we used either age of onset data or the age of study participants from the studies contributing to the GWAS Catalog (Welter et al., 2014). For phenotypes that lacked an age of onset (allergic reactions, injuries, etc.), we used the category "Any." In addition, we collapsed similar phenotypes into an umbrella phenotype (Supplemental Table S2). For example, all symptoms associated with type II diabetes were collapsed into the type II diabetes umbrella category.

Population continuity and contributions to modern genomes

We used a method developed by Sjödin et al (2014). Ancient individuals were assigned into three groups based on lifestyle: agriculturalists, hunter-gatherers, and pastoralists. For each variant, we used this program to determine the maximum possible genetic contribution of individuals from each lifestyle to modern EUR individuals. Maximum contribution was calculated by decile using a uniform prior and scaled admixture time of 0.01 (5000 years ago, given a generation time of 25 years and N_e of 10,000). This method was shown to be robust to a difference in the time of admixture compared to the age of the sample as well as variation in sample age. Consistent with previous findings (Sjödin et al., 2014), similar results were observed when using the Griffiths distribution, which incorporates ancestral and derived state information.

Allele frequency trajectory and selection inference

We used a method developed by Schraiber et al (2016). For variants where the maximum contribution of individuals of a particular lifestyle was less than 100%, we down-sampled the

individuals from that lifestyle by enforcing a chance of inclusion equal to the maximum contribution of that lifestyle. Individuals were then binned by age (>9500, 9500-8000, 8000-6500, 6500-5000, 5000-3500, and <3500 years ago). Each time point was determined by calculating the median age of all individuals included in the bin. Time was then scaled by generation time (25 years) and N_0 (10,000). The program was then run twice for each variant, once using a demographic model of constant population size, and the other including the out of Africa bottleneck and subsequent exponential growth - parameters used from Gravel et al (2011). Growth was modelled as a step-wise exponential increase in population size every 2500 years until a population size of 20,000 was reached, after which population size was held constant through to present day. Variants were only included for analysis if at least 20 ancient individuals total contributed over at least two time points (determined after individuals were down-sampled). In instances of individual down-sampling, the same individuals were used for each variant for both constant and exponential runs. The final selection coefficient (s) was determined by averaging the s_1 (selection on heterozygote) and ½ s_2 (selection on homozygote) values generated by this method. Due to overestimation of s when the derived allele frequency started very high in ancient populations, we repolarized the test to treat an ancestral allele as if it was derived when the derived allele frequency at the first time point was greater than or equal to 75%.

Introgressed and MHC variant identification

We determined the introgression status of variants using the haplotypes and variants identified in a study of Neanderthal and Denisovan introgression across modern Eurasian and Melanesian groups (Vernot et al., 2016). Coordinates defined by the Genome Reference Consortium were used

to specify the MHC region (hg38 chr6:28,510,020-33,480,577). When necessary, we used liftOver to convert data from hg38 to hg19 or vice versa (Hinrichs et al., 2006).

Binomial and hypergeometric enrichment

We used the numpy module in python to generate binomial and hypergeometric distributions. In all cases, we simulated 1,000,000 draws. Unless otherwise stated in the text, hypergeometric distributions were generated based on the results from the overall set of 2,709 GWAS variants with selection information.

Acknowledgements

We thank Josh Schraiber and members of the Lachance Lab for their helpful feedback on this manuscript. This work was supported by NIH grant R35GM133727.

References

Akinbami LJ, Simon AE, Rossen LM. 2015. Changing Trends in Asthma Prevalence Among Children. *Pediatrics*. doi:10.1542/peds.2015-2354

Akpınar M, Lesche M, Fanourgakis G, Fu J, Anasstasiadis K, Dahl A, Jessberger R. 2017. TDRD6 mediates early steps of spliceosome maturation in primary spermatocytes. *PLOS Genet* **13**:e1006660. doi:10.1371/journal.pgen.1006660

Aris-Brosou S. 2019. Direct Evidence of an Increasing Mutational Load in Humans. *Mol Biol Evol* **36**:2823–2829. doi:10.1093/molbev/msz192

Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A, Clark AG, Donnelly P, Eichler EE, Flicek P, Gabriel SB, Gibbs RA, Green ED, Hurles ME, Knoppers BM, Korbel JO, Lander ES, Lee C, Lehrach H, Mardis ER, Marth GT, McVean GA, Nickerson DA, Schmidt JP, Sherry ST, Wang J, Wilson RK, Boerwinkle E, Doddapaneni H, Han Y, Korchina V, Kovar C, Lee S, Muzny D, Reid JG, Zhu Y, Chang Y, Feng Q, Fang X, Guo X, Jian M, Jiang H, Jin X, Lan T, Li G, Li J, Li Yingrui, Liu S, Liu Xiao, Lu Y, Ma X, Tang M, Wang B, Wang G, Wu H, Wu R, Xu X, Yin Y, Zhang D, Zhang W, Zhao J, Zhao M, Zheng X, Gupta N, Gharani N, Toji LH, Gerry NP, Resch AM, Barker J, Clarke L, Gil L, Hunt SE, Kelman G, Kulesha E, Leinonen R, McLaren WM, Radhakrishnan R, Roa A, Smirnov D, Smith RE, Streeter I, Thormann A, Toneva I, Vaughan B, Zheng-Bradley X, Grocock R, Humphray S, James T, Kingsbury Z, Sudbrak R, Albrecht MW, Amstislavskiy VS, Borodina TA, Lienhard M, Mertes F, Sultan M, Timmermann B, Yaspo M-L, Fulton L, Fulton R, Ananiev V, Belaia Z, Beloslyudtsev D, Bouk N, Chen C, Church D, Cohen R, Cook C, Garner J, Hefferon T, Kimelman M, Liu C, Lopez J, Meric P, O'Sullivan C, Ostapchuk Y, Phan L, Ponomarov S, Schneider V, Shekhtman E, Sirotkin K, Slotta D, Zhang H, Balasubramaniam S, Burton J, Danecek P, Keane TM, Kolb-Kokocinski A, McCarthy S, Stalker J, Quail M, Davies CJ, Gollub J, Webster T, Wong B, Zhan Y, Campbell CL, Kong Y, Marcketta A, Yu F, Antunes L, Bainbridge M, Sabo A, Huang Z, Coin LJM, Fang L, Li Q, Li Z, Lin H, Liu B, Luo R, Shao H, Xie Y, Ye C, Yu C, Zhang F, Zheng H, Zhu H, Alkan C, Dal E, Kahveci F, Garrison EP, Kural D, Lee W-P, Fung Leong W, Stromberg M, Ward AN, Wu J, Zhang M, Daly MJ, DePristo MA, Handsaker RE, Banks E, Bhatia G, del Angel G, Genovese G, Li H, Kashin S, McCarroll SA, Nemesh JC, Poplin RE, Yoon SC, Lihm J, Makarov V, Gottipati S, Keinan A, Rodriguez-Flores JL,

Rausch T, Fritz MH, Stütz AM, Beal K, Datta A, Herrero J, Ritchie GRS, Zerbino D, Sabeti PC, Shlyakhter I, Schaffner SF, Vitti J, Cooper DN, Ball E V., Stenson PD, Barnes B, Bauer M, Keira Cheetham R, Cox A, Eberle M, Kahn S, Murray L, Peden J, Shaw R, Kenny EE, Batzer MA, Konkel MK, Walker JA, MacArthur DG, Lek M, Herwig R, Ding L, Koboldt DC, Larson D, Ye Kai, Gravel S, Swaroop A, Chew E, Lappalainen T, Erlich Y, Gymrek M, Frederick Willems T, Simpson JT, Shriver MD, Rosenfeld JA, Bustamante CD, Montgomery SB, De La Vega FM, Byrnes JK, Carroll AW, DeGorter MK, Lacroute P, Maples BK, Martin AR, Moreno-Estrada A, Shringarpure SS, Zakharia F, Halperin E, Baran Y, Cerveira E, Hwang J, Malhotra A, Plewczynski D, Radew K, Romanovitch M, Zhang C, Hyland FCL, Craig DW, Christoforides A, Homer N, Izatt T, Kurdoglu AA, Sinari SA, Squire K, Xiao C, Sebat J, Antaki D, Gujral M, Noor A, Ye Kenny, Burchard EG, Hernandez RD, Gignoux CR, Haussler D, Katzman SJ, James Kent W, Howie B, Ruiz-Linares A, Dermitzakis ET, Devine SE, Min Kang H, Kidd JM, Blackwell T, Caron S, Chen W, Emery S, Fritsche L, Fuchsberger C, Jun G, Li B, Lyons R, Scheller C, Sidore C, Song S, Sliwerska E, Taliun D, Tan A, Welch R, Kate Wing M, Zhan X, Awadalla P, Hodgkinson A, Li Yun, Shi X, Quitadamo A, Lunter G, Marchini JL, Myers S, Churchhouse C, Delaneau O, Gupta-Hinch A, Kretzschmar W, Iqbal Z, Mathieson I, Menelaou A, Rimmer A, Xifara DK, Oleksyk TK, Fu Yunxin, Liu Xiaoming, Xiong M, Jorde L, Witherspoon D, Xing J, Browning BL, Browning SR, Hormozdiari F, Sudmant PH, Khurana E, Tyler-Smith C, Albers CA, Ayub Q, Chen Y, Colonna V, Jostins L, Walter K, Xue Y, Gerstein MB, Abyzov A, Balasubramanian S, Chen J, Clarke D, Fu Yao, Harmanci AO, Jin M, Lee D, Liu J, Jasmine Mu X, Zhang J, Zhang Yan, Hartl C, Shakir K, Degenhardt J, Meiers S, Raeder B, Paolo Casale F, Stegle O, Lameijer E-W, Hall I, Bafna

V, Michaelson J, Gardner EJ, Mills RE, Dayama G, Chen K, Fan X, Chong Z, Chen T, Chaisson MJ, Huddleston J, Malig M, Nelson BJ, Parrish NF, Blackburne B, Lindsay SJ, Ning Z, Zhang Yujun, Lam H, Sisu C, Challis D, Evani US, Lu J, Nagaswamy U, Yu J, Li W, Habegger L, Yu H, Cunningham F, Dunham I, Lage K, Berg Jespersen J, Horn H, Kim D, Desalle R, Narechania A, Wilson Sayres MA, Mendez FL, David Poznik G, Underhill PA, Coin L, Mittelman D, Banerjee R, Cerezo M, Fitzgerald TW, Louzada S, Massaia A, Ritchie GR, Yang F, Kalra D, Hale W, Dan X, Barnes KC, Beiswanger C, Cai H, Cao H, Henn B, Jones D, Kaye JS, Kent A, Kerasidou A, Mathias R, Ossorio PN, Parker M, Rotimi CN, Royal CD, Sandoval K, Su Y, Tian Z, Tishkoff S, Via M, Wang Y, Yang H, Yang L, Zhu J, Bodmer W, Bedoya G, Cai Z, Gao Y, Chu J, Peltonen L, Garcia-Montero A, Orfao A, Dutil J, Martinez-Cruzado JC, Mathias RA, Hennis A, Watson H, McKenzie C, Qadri F, LaRocque R, Deng X, Asogun D, Folarin O, Happi C, Omoniwa O, Stremlau M, Tariyal R, Jallow M, Sisay Joof F, Corrah T, Rockett K, Kwiatkowski D, Kooner J, Tịnh Hiê`n T, Dunstan SJ, Thuy Hang N, Fonnie R, Garry R, Kanneh L, Moses L, Schieffelin J, Grant DS, Gallo C, Poletti G, Saleheen D, Rasheed A, Brooks LD, Felsenfeld AL, McEwen JE, Vaydylevich Y, Duncanson A, Dunn M, Schloss JA. 2015. A global reference for human genetic variation. *Nature* **526**:68–74. doi:10.1038/nature15393

Bensing S, Fetissov SO, Mulder J, Perheentupa J, Gustafsson J, Husebye ES, Oscarson M, Ekwall O, Crock PA, Hökfelt T, Hulting AL, Kämpe O. 2007. Pituitary autoantibodies in autoimmune polyendocrine syndrome type 1. *Proc Natl Acad Sci U S A* **104**:949–954. doi:10.1073/pnas.0610070104

Berens AJ, Cooper TL, Lachance J. 2017. The Genomic Health of Ancient Hominins. *Hum Biol*. doi:10.1101/145193

Bernatchez L, Landry C. 2003. MHC studies in nonmodel vertebrates: What have we learned about natural selection in 15 years? *J Evol Biol*. doi:10.1046/j.1420-9101.2003.00531.x

Bowatte G, Lodge C, Lowe AJ, Erbas B, Perret J, Abramson MJ, Matheson M, Dharmage SC. 2015. The influence of childhood traffic-related air pollution exposure on asthma, allergy and sensitization: a systematic review and a meta-analysis of birth cohort studies. *Allergy* **70**:245–256. doi:10.1111/all.12561

Dannemann M, Andrés AM, Kelso J. 2016. Introgression of Neandertal- and Denisovan-like Haplotypes Contributes to Adaptive Variation in Human Toll-like Receptors. *Am J Hum Genet* **98**:22–33. doi:10.1016/j.ajhg.2015.11.015

Fu Q, Li H, Moorjani P, Jay F, Slepchenko SM, Bondarev AA, Johnson PLF, Aximu-Petri A, Prüfer K, de Filippo C, Meyer M, Zwyns N, Salazar-García DC, Kuzmin Y V, Keates SG, Kosintsev PA, Razhev DI, Richards MP, Peristov N V, Lachmann M, Douka K, Higham TFG, Slatkin M, Hublin J-J, Reich D, Kelso J, Viola TB, Pääbo S. 2014. Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**:445–9. doi:10.1038/nature13810

Gauld SB, Cambier JC. 2004. Src-family kinases in B-cell development and signaling. *Oncogene*. doi:10.1038/sj.onc.1208075

Gorlova OY, Ying J, Amos CI, Spitz MR, Peng B, Gorlov IP. 2012. Derived SNP Alleles Are Used More Frequently Than Ancestral Alleles As Risk-Associated Variants In Common Human Diseases. *J Bioinform Comput Biol*. doi:10.1142/S0219720012410089

Gravel S, Henn BM, Gutenkunst RN, Indap AR, Marth GT, Clark AG, Yu F, Gibbs RA, Bustamante CD, Altshuler DL, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, Collins FS, De La Vega FM, Donnelly P, Egholm M, Flicek P, Gabriel SB,

Gibbs RA, Knoppers BM, Lander ES, Lehrach H, Mardis ER, McVean GA, Nickerson DA, Peltonen L, Schafer AJ, Sherry ST, Wang J, Wilson RK, Gibbs RA, Deiros D, Metzker M, Muzny D, Reid J, Wheeler D, Wang J, Li J, Jian M, Li G, Li R, Liang H, Tian G, Wang B, Wang J, Wang W, Yang H, Zhang X, Zheng H, Lander ES, Altshuler DL, Ambrogio L, Bloom T, Cibulskis K, Fennell TJ, Gabriel SB, Jaffe DB, Shefler E, Sougnez CL, Bentley DR, Gormley N, Humphray S, Kingsbury Z, Koko-Gonzales P, Stone J, McKernan KJ, Costa GL, Ichikawa JK, Lee CC, Sudbrak R, Lehrach H, Borodina TA, Dahl A, Davydov AN, Marquardt P, Mertes F, Nietfeld W, Rosenstiel P, Schreiber S, Soldatov A V., Timmermann B, Tolzmann M, Egholm M, Affourtit J, Ashworth D, Attiya S, Bachorski M, Buglione E, Burke A, Caprio A, Celone C, Clark S, Conners D, Desany B, Gu L, Guccione L, Kao K, Kebbel A, Knowlton J, Labrecque M, McDade L, Mealmaker C, Minderman M, Nawrocki A, Niazi F, Pareja K, Ramenani R, Riches D, Song W, Turcotte C, Wang S, Mardis ER, Wilson RK, Dooling D, Fulton L, Fulton R, Weinstock G, Durbin RM, Burton J, Carter DM, Churcher C, Coffey A, Cox A, Palotie A, Quail M, Skelly T, Stalker J, Swerdlow HP, Turner D, De Witte A, Giles S, Gibbs RA, Wheeler D, Bainbridge M, Challis D, Sabo A, Yu F, Yu J, Wang J, Fang X, Guo X, Li R, Li Y, Luo R, Tai S, Wu H, Zheng H, Zheng X, Zhou Y, Li G, Wang J, Yang H, Marth GT, Garrison EP, Huang W, Indap A, Kural D, Lee W-P, Leong WF, Quinlan AR, Stewart C, Stromberg MP, Ward AN, Wu J, Lee C, Mills RE, Shi X, Daly MJ, DePristo MA, Altshuler DL, Ball AD, Banks E, Bloom T, Browning BL, Cibulskis K, Fennell TJ, Garimella K V., Grossman SR, Handsaker RE, Hanna M, Hartl C, Jaffe DB, Kernytsky AM, Korn JM, Li H, Maguire JR, McCarroll SA, McKenna A, Nemesh JC, Philippakis AA, Poplin RE, Price A, Rivas MA, Sabeti PC, Schaffner SF, Shefler E, Shlyakhter IA, Cooper DN, Ball E V., Mort M, Phillips

AD, Stenson PD, Sebat J, Makarov V, Ye K, Yoon SC, Bustamante CD, Clark AG, Boyko A, Degenhardt J, Gravel S, Gutenkunst RN, Kaganovich M, Keinan A, Lacroute P, Ma X, Reynolds A, Clarke L, Flicek P, Cunningham F, Herrero J, Keenen S, Kulesha E, Leinonen R, McLaren WM, Radhakrishnan R, Smith RE, Zalunin V, Zheng-Bradley X, Korbel JO, Stutz AM, Humphray S, Bauer M, Cheetham RK, Cox T, Eberle M, James T, Kahn S, Murray L, Chakravarti A, Ye K, De La Vega FM, Fu Y, Hyland FCL, Manning JM, McLaughlin SF, Peckham HE, Sakarya O, Sun YA, Tsung EF, Batzer MA, Konkel MK, Walker JA, Sudbrak R, Albrecht MW, Amstislavskiy VS, Herwig R, Parkhomchuk D V., Sherry ST, Agarwala R, Khouri HM, Morgulis AO, Paschall JE, Phan LD, Rotmistrovsky KE, Sanders RD, Shumway MF, Xiao C, McVean GA, Auton A, Iqbal Z, Lunter G, Marchini JL, Moutsianas L, Myers S, Tumian A, Desany B, Knight J, Winer R, Craig DW, Beckstrom-Sternberg SM, Christoforides A, Kurdoglu AA, Pearson J V., Sinari SA, Tembe WD, Haussler D, Hinrichs AS, Katzman SJ, Kern A, Kuhn RM, Przeworski M, Hernandez RD, Howie B, Kelley JL, Melton SC, Abecasis GR, Li Y, Anderson P, Blackwell T, Chen W, Cookson WO, Ding J, Kang HM, Lathrop M, Liang L, Moffatt MF, Scheet P, Sidore C, Snyder M, Zhan X, Zollner S, Awadalla P, Casals F, Idaghdour Y, Keebler J, Stone EA, Zilversmit M, Jorde L, Xing J, Eichler EE, Aksay G, Alkan C, Hajirasouliha I, Hormozdiari F, Kidd JM, Sahinalp SC, Sudmant PH, Mardis ER, Chen K, Chinwalla A, Ding L, Koboldt DC, McLellan MD, Dooling D, Weinstock G, Wallis JW, Wendl MC, Zhang Q, Durbin RM, Albers CA, Ayub Q, Balasubramaniam S, Barrett JC, Carter DM, Chen Y, Conrad DF, Danecek P, Dermitzakis ET, Hu M, Huang N, Hurles ME, Jin H, Jostins L, Keane TM, Le SQ, Lindsay S, Long Q, MacArthur DG, Montgomery SB, Parts L, Stalker J, Tyler-Smith C, Walter K, Zhang Y, Gerstein MB, Snyder M, Abyzov A, Balasubramanian S, Bjornson

R, Du J, Grubert F, Habegger L, Haraksingh R, Jee J, Khurana E, Lam HYK, Leng J, Mu XJ, Urban AE, Zhang Z, Li Y, Luo R, Marth GT, Garrison EP, Kural D, Quinlan AR, Stewart C, Stromberg MP, Ward AN, Wu J, Lee C, Mills RE, Shi X, McCarroll SA, Banks E, DePristo MA, Handsaker RE, Hartl C, Korn JM, Li H, Nemesh JC, Sebat J, Makarov V, Ye K, Yoon SC, Degenhardt J, Kaganovich M, Clarke L, Smith RE, Zheng-Bradley X, Korbel JO, Humphray S, Cheetham RK, Eberle M, Kahn S, Murray L, Ye K, De La Vega FM, Fu Y, Peckham HE, Sun YA, Batzer MA, Konkel MK, Walker JA, Xiao C, Iqbal Z, Desany B, Blackwell T, Snyder M, Xing J, Eichler EE, Aksay G, Alkan C, Hajirasouliha I, Hormozdiari F, Kidd JM, Chen K, Chinwalla A, Ding L, McLellan MD, Wallis JW, Hurles ME, Conrad DF, Walter K, Zhang Y, Gerstein MB, Snyder M, Abyzov A, Du J, Grubert F, Haraksingh R, Jee J, Khurana E, Lam HYK, Leng J, Mu XJ, Urban AE, Zhang Z, Gibbs RA, Bainbridge M, Challis D, Coafra C, Dinh H, Kovar C, Lee S, Muzny D, Nazareth L, Reid J, Sabo A, Yu F, Yu J, Marth GT, Garrison EP, Indap A, Leong WF, Quinlan AR, Stewart C, Ward AN, Wu J, Cibulskis K, Fennell TJ, Gabriel SB, Garimella K V., Hartl C, Shefler E, Sougnez CL, Wilkinson J, Clark AG, Gravel S, Grubert F, Clarke L, Flicek P, Smith RE, Zheng-Bradley X, Sherry ST, Khouri HM, Paschall JE, Shumway MF, Xiao C, McVean GA, Katzman SJ, Abecasis GR, Blackwell T, Mardis ER, Dooling D, Fulton L, Fulton R, Koboldt DC, Durbin RM, Balasubramaniam S, Coffey A, Keane TM, MacArthur DG, Palotie A, Scott C, Stalker J, Tyler-Smith C, Gerstein MB, Balasubramanian S, Chakravarti A, Knoppers BM, Abecasis GR, Bustamante CD, Gharani N, Gibbs RA, Jorde L, Kaye JS, Kent A, Li T, McGuire AL, McVean GA, Ossorio PN, Rotimi CN, Su Y, Toji LH, TylerSmith C, Brooks LD, Felsenfeld AL, McEwen JE, Abdallah A, Juenger CR, Clemm NC, Collins FS, Duncanson A, Green ED, Guyer MS, Peterson JL, Schafer AJ,

Abecasis GR, Altshuler DL, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA. 2011. Demographic history and rare allele sharing among human populations. *Proc Natl Acad Sci*. doi:10.1073/pnas.1019276108

Gunz P, Tilot AK, Wittfeld K, Teumer A, Shapland CY, van Erp TGM, Dannemann M, Vernot B, Neubauer S, Guadalupe T, Fernández G, Brunner HG, Enard W, Fallon J, Hosten N, Völker U, Profico A, Di Vincenzo F, Manzi G, Kelso J, St Pourcain B, Hublin J-J, Franke B, Pääbo S, Macciardi F, Grabe HJ, Fisher SE. 2018. Neandertal Introgression Sheds Light on Modern Human Endocranial Globularity. *Curr Biol* **29**:120-127.e5. doi:10.1016/j.cub.2018.10.065

Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, Brandt G, Nordenfelt S, Harney E, Stewardson K, Fu Q, Mittnik A, Bánffy E, Economou C, Francken M, Friederich S, Pena RG, Hallgren F, Khartanovich V, Khokhlov A, Kunst M, Kuznetsov P, Meller H, Mochalov O, Moiseyev V, Nicklisch N, Pichler SL, Risch R, Rojo Guerra MA, Roth C, Szécsényi-Nagy A, Wahl J, Meyer M, Krause J, Brown D, Anthony D, Cooper A, Alt KW, Reich D. 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*. doi:10.1038/nature14317

Hinrichs A, Karolchik D, Baertsch R, Barber G, Bejerano G, Clawson H. 2006. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res* **34**:D590–D598. doi:10.1093/nar/gkj144

Hom G, Graham RR, Modrek B, Taylor KE, Ortmann W, Garnier S, Lee AT, Chung SA, Ferreira RC, Pant PVK, Ballinger DG, Kosoy R, Demirci FY, Kamboh MI, Kao AH, Tian C, Gunnarsson I, Bengtsson AA, Rantapää-Dahlqvist S, Petri M, Manzi S, Seldin MF, Rönnblom L, Syvänen A-C, Criswell LA, Gregersen PK, Behrens TW. 2008. Association of

systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. *N Engl J Med*. doi:10.1056/NEJMoa0707865

Jensen JD, Payseur BA, Stephan W, Aquadro CF, Lynch M, Charlesworth D, Charlesworth B. 2019. The importance of the Neutral Theory in 1968 and 50 years on: A response to Kern and Hahn 2018. *Evolution (N Y)* **73**:111–114. doi:10.1111/evo.13650

Karlsson EK, Kwiatkowski DP, Sabeti PC. 2014. Natural selection and infectious disease in human populations. *Nat Rev Genet*. doi:10.1038/nrg3734

Kern AD, Hahn MW. 2018. The Neutral Theory in Light of Natural Selection. *Mol Biol Evol* **35**:1366–1371. doi:10.1093/molbev/msy092

Kiyotani K, Mai TH, Nakamura Y. 2017. Comparison of exome-based HLA class i genotyping tools: Identification of platform-specific genotyping errors. *J Hum Genet*. doi:10.1038/jhg.2016.141

Lachance J. 2010. Disease-associated alleles in genome-wide association studies are enriched for derived low frequency alleles relative to HapMap and neutral expectations. *BMC Med Genomics*. doi:10.1186/1755-8794-3-57

Lachance J, Berens AJ, Hansen MEB, Teng AK, Tishkoff SA, Rebbeck TR. 2018. Genetic hitchhiking and population bottlenecks contribute to prostate cancer disparities in men of african descent. *Cancer Res* **78**:2432–2443. doi:10.1158/0008-5472.CAN-17-1550

Lynch M. 2016. Mutation and Human Exceptionalism: Our Future Genetic Load. *Genetics* **202**:869–875. doi:10.1534/genetics.115.180471

Marigorta UM, Lao O, Casals F, Calafell F, Morcillo-Suárez C, Faria R, Bosch E, Serra F, Bertranpetit J, Dopazo H, Navarro A. 2011. Recent human evolution has shaped geographical differences in susceptibility to disease. *BMC Genomics* **12**:55.

doi:10.1186/1471-2164-12-55

Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, Harney E, Stewardson K, Fernandes D, Novak M, Sirak K, Gamba C, Jones ER, Llamas B, Dryomov S, Pickrell J, Arsuaga JL, De Castro JMB, Carbonell E, Gerritsen F, Khokhlov A, Kuznetsov P, Lozano M, Meller H, Mochalov O, Moiseyev V, Guerra MAR, Roodenberg J, Vergès JM, Krause J, Cooper A, Alt KW, Brown D, Anthony D, Lalueza-Fox C, Haak W, Pinhasi R, Reich D. 2015. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**:499–503. doi:10.1038/nature16152

Mathieson S, Mathieson I. 2018. *FADS1* and the Timing of Human Adaptation to Agriculture. *Mol Biol Evol* **35**:2957–2970. doi:10.1093/molbev/msy180

Meyer M, Kircher M, Gansauge M, Li H, Racimo F, Siebauer M, Green RE, Bryc K, Briggs AW, Stenzel U, Dabney J, Shendure J, Kitzman J, Slatkin M, Reich D. 2012. A High-Coverage Genome Sequence from an Archaic Denisovan Individual A High-Coverage Genome Sequence from an Archaic Denisovan Individual A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Sci (New York, NY)* **222**:1–14. doi:10.1126/science.1224344

Narasimhan VM, Patterson N, Moorjani P, Rohland N, Bernardos R, Mallick S, Lazaridis I, Nakatsuka N, Olalde I, Lipson M, Kim AM, Olivieri LM, Coppa A, Vidale M, Mallory J, Moiseyev V, Kitov E, Monge J, Adamski N, Alex N, Broomandkhoshbacht N, Candilio F, Callan K, Cheronet O, Culleton BJ, Ferry M, Fernandes D, Freilich S, Gamarra B, Gaudio D, Hajdinjak M, Harney É, Harper TK, Keating D, Lawson AM, Mah M, Mandl K, Michel M, Novak M, Oppenheimer J, Rai N, Sirak K, Slon V, Stewardson K, Zalzala F, Zhang Z, Akhatov G, Bagashev AN, Bagnera A, Baitanayev B, Bendezu-Sarmiento J, Bissembaev

AA, Bonora GL, Chargynov TT, Chikisheva T, Dashkovskiy PK, Derevianko A, Dobeš M, Douka K, Dubova N, Duisengali MN, Enshin D, Epimakhov A, Fribus A V., Fuller D, Goryachev A, Gromov A, Grushin SP, Hanks B, Judd M, Kazizov E, Khokhlov A, Krygin AP, Kupriyanova E, Kuznetsov P, Luiselli D, Maksudov F, Mamedov AM, Mamirov TB, Meiklejohn C, Merrett DC, Micheli R, Mochalov O, Mustafokulov S, Nayak A, Pettener D, Potts R, Razhev D, Rykun M, Sarno S, Savenkova TM, Sikhymbaeva K, Slepchenko SM, Soltobaev OA, Stepanova N, Svyatko S, Tabaldiev K, Teschler-Nicola M, Tishkin AA, Tkachev V V., Vasilyev S, Velemínský P, Voyakin D, Yermolayeva A, Zahir M, Zubkov VS, Zubova A, Shinde VS, Lalueza-Fox C, Meyer M, Anthony D, Boivin N, Thangaraj K, Kennett DJ, Frachetti M, Pinhasi R, Reich D. 2019. The formation of human populations in South and Central Asia. *Science (80-)* **365**. doi:10.1126/science.aat7487

Olalde I, Brace S, Allentoft ME, Armit I, Kristiansen K, Booth T, Rohland N, Mallick S, Szécsényi-Nagy A, Mittnik A, Altena E, Lipson M, Lazaridis I, Harper TK, Patterson N, Broomandkhoshbacht N, Diekmann Y, Faltyskova Z, Fernandes D, Ferry M, Harney E, De Knijff P, Michel M, Oppenheimer J, Stewardson K, Barclay A, Alt KW, Liesau C, Rios P, Blasco C, Miguel JV, Garcia RM, Fernandez AA, Banffy E, Bernabo-Brea M, Billoin D, Bonsall C, Bonsall L, Allen T, Buster L, Carver S, Navarro LC, Craig OE, Cook GT, Cunliffe B, Denaire A, Dinwiddy KE, Dodwell N, Ernee M, Evans C, Kucharik M, Farre JF, Fowler C, Gazenbeek M, Pena RG, Haber-Uriarte M, Haduch E, Hey G, Jowett N, Knowles T, Massy K, Pfrengle S, Lefranc P, Lemercier O, Lefebvre A, Martinez CH, Olmo VG, Ramirez AB, Maurandi JL, Majo T, McKinley JI, McSweeney K, Mende BG, Mod A, Kulcsar G, Kiss V, Czene A, Patay R, Endrodi A, Kohler K, Hajdu T, Szeniczey T, Dani J, Bernert Z, Hoole M, Cheronet O, Keating D, Veleminsky P, Dobe M, Candilio F, Brown F,

Fernandez RF, Herrero-Corral AM, Tusa S, Carnieri E, Lentini L, Valenti A, Zanini A, Waddington C, Delibes G, Guerra-Doce E, Neil B, Brittain M, Luke M, Mortimer R, Desideri J, Besse M, Brucken G, Furmanek M, Hauszko A, Mackiewicz M, Rapinski A, Leach S, Soriano I, Lillios KT, Cardoso JL, Pearson MP, Wodarczak P, Price TD, Prieto P, Rey PJ, Risch R, Guerra MAR, Schmitt A, Serralongue J, Silva AM, Smrcka V, Vergnaud L, Zilhao J, Caramelli D, Higham T, Thomas MG, Kennett DJ, Fokkens H, Heyd V, Sheridan A, Sjogren KG, Stockhammer PW, Krause J, Pinhasi R, Haak W, Barnes I, Lalueza-Fox C, Reich D. 2018. The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* **555**:190–196. doi:10.1038/nature25738

Oryoji D, Ueda S, Yamamoto K, Noh JY, Okamura K, Noda M, Watanabe N, Yoshihara A, Ito K, Sasazuki T. 2015. Identification of a hashimoto thyroiditis susceptibility locus via a genome-wide comparison with graves' disease. *J Clin Endocrinol Metab*. doi:10.1210/jc.2014-3431

Park J-H, Gail MH, Weinberg CR, Carroll RJ, Chung CC, Wang Z, Chanock SJ, Fraumeni JF, Chatterjee N. 2011. Distribution of allele frequencies and effect sizes and their interrelationships for common genetic susceptibility variants. *Proc Natl Acad Sci*. doi:10.1073/pnas.1114759108

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

Piertney SB, Oliver MK. 2006. The evolutionary ecology of the major histocompatibility complex. *Heredity (Edinb)*. doi:10.1038/sj.hdy.6800724

Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G, Sudmant PH, de Filippo C, Li H, Mallick S, Dannemann M, Fu Q, Kircher M, Kuhlwilm M,

Lachmann M, Meyer M, Ongyerth M, Siebauer M, Theunert C, Tandon A, Moorjani P, Pickrell J, Mullikin JC, Vohr SH, Green RE, Hellmann I, Johnson PLF, Blanche H, Cann H, Kitzman JO, Shendure J, Eichler EE, Lein ES, Bakken TE, Golovanova L V, Doronichev VB, Shunkov M V, Derevianko AP, Viola B, Slatkin M, Reich D, Kelso J, Pääbo S. 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**:43–9. doi:10.1038/nature12886

Racimo F, Marnetto D, Huerta-Sánchez E. 2016. Signatures of Archaic Adaptive Introgression in Present-Day Human Populations. *Mol Biol Evol* **34**:msw216. doi:10.1093/molbev/msw216

Racimo F, Schraiber JG. 2014. Approximation to the Distribution of Fitness Effects across Functional Categories in Human Segregating Polymorphisms. *PLoS Genet*. doi:10.1371/journal.pgen.1004697

Rodríguez JA, Marigorta UM, Navarro A. 2014. Integrating genomics into evolutionary medicine. *Curr Opin Genet Dev*. doi:10.1016/j.gde.2014.08.009

Schraiber JG, Evans SN, Slatkin M. 2016. Bayesian inference of natural selection from allele frequency time series. *Genetics*. doi:10.1534/genetics.116.187278

Simonti CN, Vernot B, Bastarache L, Bottinger E, Carrell DS, Chisholm RL, Crosslin DR, Hebbring SJ, Jarvik GP, Kullo IJ, Li R, Pathak J, Ritchie MD, Roden DM, Verma SS, Tromp G, Prato JD, Bush WS, Akey JM, Denny JC, Capra JA. 2016. The phenotypic legacy of admixture between modern humans and Neandertals. *Science (80-)* **351**:737–741. doi:10.1126/science.aad2149

Sjödin P, Skoglund P, Jakobsson M. 2014. Assessing the maximum contribution from ancient populations. *Mol Biol Evol*. doi:10.1093/molbev/msu059

Timpson NJ, Greenwood CMT, Soranzo N, Lawson DJ, Richards JB. 2018. Genetic architecture:

The shape of the genetic contribution to human traits and disease. *Nat Rev Genet*. doi:10.1038/nrg.2017.101

Vernot B, Akey J. 2014. Resurrecting surviving Neanderthal lineages from modern human genomes. *Science (80-)* **20**:96–103. doi:10.1002/evan.20301

Vernot B, Tucci S, Kelso J, Schraiber JG, Wolf AB, Gittelman RM, Dannemann M, Grote S, McCoy RC, Norton H, Scheinfeldt LB, Merriwether DA, Koki G, Friedlaender JS, Wakefield J, Pääbo S, Akey JM. 2016. Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. *Science* **352**:235–9. doi:10.1126/science.aad9416

Weghorn D, Balick DJ, Cassa C, Kosmicki JA, Daly MJ, Beier DR, Sunyaev SR. 2019. Applicability of the Mutation – Selection Balance Model to Population Genetics of Heterozygous Protein-Truncating Variants in Humans. *Mol Biol Evol* **36**:1701–1710. doi:10.1093/molbev/msz092

Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorff L, Parkinson H. 2014. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* **42**. doi:10.1093/nar/gkt1229

WHO | Ambient (outdoor) air quality and health. 2019. WHO | Ambient (outdoor) air quality and health. *Who*.

Williams GC. 1957. PLEIOTROPY, NATURAL SELECTION, AND THE EVOLUTION OF SENESCENCE. *Evolution (N Y)*. doi:10.2307/2406060