GBE

Archaic Introgression Shapes Genetic Variation at Loci Associated With DNA Virus Load in Modern Humans

Rutvi Rajpara [1], Sofiia Stefaniia Polishchuk [1], Danat Yermakovich [1], Michael Dannemann [1,*]

[1]Center of Genomics, Evolution and Medicine, Institute of Genomics, University of Tartu, Tartu, Estonia

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

Accepted: April 13, 2026

Abstract

Archaic introgression has shaped the modern human immune system, particularly components involved in RNA virus responses. In contrast, its contribution to DNA virus defense remains poorly understood. Here, we investigate the contribution of Neandertal- and Denisovan-introgressed haplotypes to viral load of five common DNA viruses in UK Biobank samples, using genome-wide association summary statistics. We identified 18 genome-wide significant associations, predominantly involving Epstein–Barr virus (EBV) and loci within the Major Histocompatibility Complex, including a Denisovan-like haplotype tightly linked to HLA-A*11:01. Notably, the archaic alleles of these haplotypes showed a directional bias toward increased viral loads. Focusing on two chromosome 17 haplotypes associated with higher EBV load, we identified phenotypic associations with blood cell traits and disease markers, as well as functional effects on immune-relevant genes, including *GSDMB*, *ARRB2*, and *ALOX15*. Allele frequency analysis of one of the chromosome 17 haplotypes revealed signatures of shifting modes of selection, possibly reflecting changes in pathogen landscapes over time. Our results suggest that archaic DNA systematically contributes to DNA virus immunity in modern humans, with effects distinct from previously described RNA virus associations. These findings provide novel evolutionary and functional insights into host–virus interactions and the role of archaic admixture in antiviral defense.

Key words: human evolution, archaic admixture, immune response, host–pathogen interaction.

Significance

Introgressed archaic DNA from Neandertals and Denisovans has been shown to modulate immune function in present-day people. In particular, Neandertal DNA has been strongly linked to RNA virus defense, with several introgressed haplotypes showing signatures of selection and associations with RNA virus response. However, little is known about the impact of archaic DNA on DNA viruses, a class of pathogens that shows characteristic differences from RNA viruses in infection signatures as well as evolutionary and genomic dynamics. In this study, we address this gap by exploring the functional and evolutionary signatures of introgressed archaic haplotypes associated with DNA virus load. We show that these haplotypes map to key immune regions and are disproportionately linked to higher viral loads. We also identify signatures of recent negative selection on one prominent candidate region. Overall, our results suggest notable differences in how archaic introgression contributes to DNA versus RNA virus response.

Introduction

Modern humans have been exposed to a wide variety of pathogens throughout their existence. Among them, viruses posed a particularly challenging threat because of their ability to rapidly evolve and evade immunity. Consequently, genomic regions involved in immune responses have frequently been identified as targets of various modes of natural selection in modern humans (Barreiro and Quintana-Murci 2020). Among the targets of positive selection are genetic variants of archaic ancestry introduced through admixture with Neandertals and Denisovans (Zeberg et al. 2024). Variants inherited from Neandertals still comprise ~2% of the genomes of all present-day non-Africans, while an additional 2% to 4% of the genomes of Oceanians are of Denisovan ancestry. Relative to their prevalence in present-day human genomes, Neandertal variants have been overproportionally linked to responses to RNA viruses (Quach et al. 2016; Enard and Petrov 2018). However, much less is known about the impact of archaic admixture to the defense of DNA viruses. Due to their different makeup, RNA and DNA viruses show distinct infection signatures, including their replication strategies, latency, seasonality, and tissue tropism. For example, RNA viruses generally show higher mutation and evolutionary rates than DNA viruses due to the more error-prone RNA replication, which lacks a proofreading mechanism of DNA (Rampersad and Tennant 2018). Patterns in seasonality and tropism may differ because RNA viruses often cause acute infections with high transmission, while DNA viruses frequently persist in immune-privileged tissues (Jones et al. 2021; Shaw Stewart and Bach 2022). These mechanistic differences between RNA and DNA viruses suggest that host immune pressures and responses, and thus evolutionary dynamics, might differ between virus classes. Hence, the motivation of our study to assess the impact of archaic admixture on DNA viral immune response in present-day people. To explore this question, we analyzed genome-wide association study (GWAS) summary statistics for viral load measurement derived from whole-genome sequencing data from the UK Biobank (Kamitaki et al. 2026). Using these summary statistics, we investigate to what extent introgressed archaic haplotypes influence DNA virus load in modern humans. By combining GWAS summary statistics with functional annotation we aim to identify molecular and phenotypic links of virus-associated introgressed haplotypes. Furthermore, we will explore signatures of selection associated with candidate haplotypes, to understand their potential role in host-pathogen evolution. Overall, our results aim to uncover how archaic introgression has shaped human antiviral defense, and potentially reveal differential functional targets and evolutionary pressures for DNA versus RNA viruses.

Results

Mapping the Archaic Ancestry Among Significant Virus Load-associated Variants

In this study, we explored GWAS summary statistics from viral load measurements for associations with introgressed archaic haplotypes (Kamitaki et al. 2026). Viral load has been assessed using whole genome sequencing data derived from blood samples of UK Biobank donors (Bycroft et al. 2018). GWAS summary statistics were available for five DNA viruses with sufficient sample size and detectable load variance. These viruses were the Human herpesvirus 7 (HHV-7), Epstein–Barr virus (EBV), and three torque teno viruses (TTV1, TTV16, and TTV17). All five viruses commonly infect humans and often establish long-term, largely asymptomatic infections (Timmerman et al. 2024; Dotto-Maurel et al. 2025). Furthermore, their viral levels in body fluids like blood or saliva can reflect current aspects of immune function or physiological conditions (de Oliveira Neto et al. 2024; Sabbaghian et al. 2024). To identify significantly associated introgressed archaic haplotypes within the GWAS data, we used a reference list of those previously annotated in Eurasian populations from the 1000 Genomes Project (1000 Genomes Project Consortium et al. 2015; Yermakovich et al. 2024). We only tested archaic marker single nucleotide polymorphisms (aSNPs, Materials and Methods) that reside on archaic haplotypes containing at least five aSNPs and have an archaic allele frequency of at least 1% in a given GWAS. To account for multiple testing, we lowered the association *P*-value significance threshold to 10^{-9}. In total, we found 18 instances where at least one aSNP matched these criteria, distributed across 13 distinct introgression loci (Table S1). The majority of those associations were with EBV levels (12 associations, HHV7: 4 associations, TTV16 and TTV17: each one association). The number of associations for each virus is consistent with the overall genome-wide significant association proportions, which report 58% of associated loci for EBV load, 31% for HHV7 load, and 11% for TTV load (χ^2 test, $P = 0.75$). Archaic haplotypes at 10 of the 13 introgression loci showed the closest sequence similarity to 1 of the 3 high-coverage Neandertal genomes, with haplotypes at the remaining three loci showing a closer sequence affinity to the Denisovan genome (Materials and Methods, Table S1). Archaic haplotypes at 12 loci were found in 1000 Genomes Europeans, while only eight each were also present in either East or South Asians (Table S1). The predominant presence

of these haplotypes in Europeans, together with their closer sequence similarity to Neandertals, is consistent with the mainly European ancestry of the GWAS cohort and the fact that European populations carry Neandertal, but little to no Denisovan, DNA (Prüfer et al. 2017).

Genomic Distribution and Association Significance of Archaic Haplotypes

Ten of the 13 archaic loci were located within the Major Histocompatibility complex—a genomic region encoding a number of proteins that play a key role in immune recognition and antigen presentation (Neefjes et al. 2011). Two additional loci were located on chromosome 17 and one on chromosome 5 (Table S1). To assess the contribution to viral load association signals, we explored the landscape of linkage disequilibrium (LD) of the top-associated aSNPs on a given candidate haplotype with other genetic variants. To address this question, we leveraged LD information from the UK Biobank cohort and explored possible instances where aSNPs were in LD with substantially stronger associated non-archaic variants (Materials and Methods). This task is notoriously challenging to perform within the MHC because of its high functionality, diversity, and complexity that is characterized by high levels of LD, large levels of pleiotropy with substantial effect sizes, and regions that cannot be assessed for genomic features like LD (Trowsdale and Knight 2013). We found that all archaic haplotypes within the MHC, with one exception, were in moderate LD with one of several more strongly associated non-aSNPs ($r^2 > 0.2$), but not in high LD ($r^2 < 0.8$), suggesting that these associations were indirect and not driven by the archaic haplotypes themselves (Fig. S1). The exception was an EBV-associated aSNP (chr6:29,818,280_G/T) segregating on an archaic haplotype that was located within a region in the MHC for which we could not assess the level of LD to other variants due to the missing LD information in the region (Fig. S1). Similarly to other MHC-associated haplotypes, aSNPs at the archaic locus on chromosome 5 also showed an LD pattern ($0.2 < r^2 < 0.8$) to a number of stronger associated (EBV and HHV7) non-aSNPs within close proximity (Fig. S1), positioning the archaic haplotype as a weak candidate for carrying the potentially causal variant underlying the association signal. In contrast, for both EBV level association loci on chromosome 17, the significantly associated archaic variants were either the lead SNP (chr17:39,893,872_C/T, Fig. 1) or in very high LD with the lead single nucleotide polymorphism (SNP) (aSNP chr17:4,664,269_T/C with $r^2 = 0.92$ with lead SNP chr17:4,685,914_T/C, Fig. 1). In general, we found 43 variants in high LD ($r^2 > 0.9$) with chr17:39,893,872_C/T—21 of which were aSNPs themselves (Table S2). For 18 of the remaining 22 variants, genotype information was not available in the three Neandertals and the Denisovan genomes (Materials and Methods)—preventing them from being identified as aSNPs. Consistent with their potential archaic origin, 17 of the 22 non-aSNPs were absent in the African Yoruba population—a population frequently used as an un-admixed reference population because of their low archaic ancestry estimates (Prüfer et al. 2017). For chr17:4,664,269_T/C, we found that only 10 of the 86 variants in LD with the lead aSNP were annotated as aSNPs. Of the remaining 76 non-aSNPs, 28 had genotype data available in at least one archaic individual, and 21 of these had data for all 4 archaic genomes—limiting or even preventing their ability to be annotated as aSNPs (Table S3). Among the variants absent in archaic genomes were all seven (five SNPs and two indels) that showed a stronger association than the lead aSNP chr17:4,664,269_T/C. However, in contrast to the other lead aSNP on chromosome 17, for the 76 non-aSNPs in LD with chr17:4,664,269_T/C, we found a substantially larger fraction of variants present in the Yoruba population. A total of 40 non-aSNPs segregated in this African population, with frequencies between 4% and 63%, suggesting that they might not be of archaic origin. In order to understand the haplotype complexity in both candidate regions and the relationship between virus load-associated variants and the local haplotype structure, we generated haplotype networks for both regions using the 1000 Genomes cohort (Materials and Methods). For the region of chr17:39,893,872_C/T we detected ten core haplotypes (Fig. S2). One of those haplotypes (haplotype III) carried all 43 EBV-associated variants included in the network analysis. The haplotype also showed the closest sequence affinity to the three Neandertal genomes (80% to 90% sequence similarity) and was primarily found in non-African populations (Fig. S2). None of the other core haplotypes carried more than four virus load-associated variants and showed a higher sequence similarity than 35% to any of the Neandertals (Table S2). Our analysis confirms the presence of a Neandertal-like haplotype that is strongly linked to the EBV association signal we detected for chr17:39,893,872_C/T. For the region encompassing chr17:4,664,269_T/C, we detected ten core haplotypes as well. One of those core haplotypes (haplotype II, Fig. S2) carried all 63 variants included in the network analysis that were in high LD with the lead aSNP ($r^2 > 0.9$). The haplotype also carried the regional GWAS lead SNP and showed the closest sequence similarity to Neandertal genomes (92% to 94% sequence similarity) and was found almost exclusively outside of Africa (Fig. S2). However, we also detected

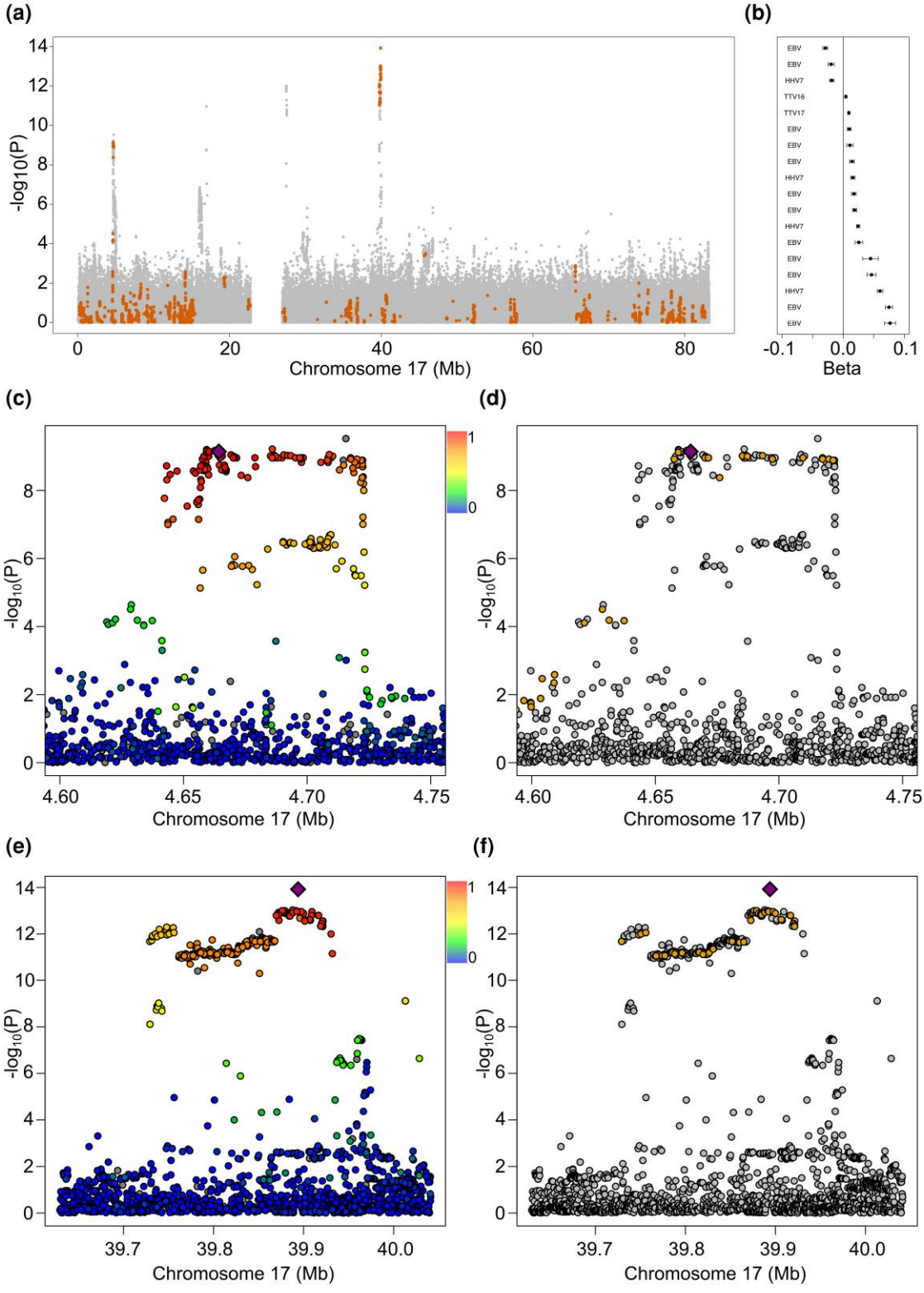

Fig. 1. Archaic haplotypes with DNA virus level associations. a) Manhattan plot for EBV level associations on chromosome 17. Neandertal marker variants (aSNPs) are highlighted in amber. b) Effect size estimates (beta ± 95% CI) from GWAS summary statistics of viral load measurements for aSNPs with $P < 10^{-9}$, shown relative to the archaic allele and sorted by the beta value. c to f) Manhattan plots for two candidate regions on chromosome 17 (c,d: lead aSNP chr17:4,664,269_T/C/e,f: lead aSNP chr17:39,893,872_C/T), displaying the LD levels (UK Biobank) relative to the lead aSNP (c, e; lead aSNP highlighted by purple diamond) and the general aSNP content (d,f; aSNPs highlighted in amber) in a given region.

eight core haplotypes—all primarily segregating in Africa—that carried some of the EBV-associated SNPs (between 4 and 20) as well (Table S3). One of those haplotypes (VI) carried 9 virus load-associated SNPs, including the top associated SNP among the 64 variants included in the haplotype network analysis. These results suggest that virus load-associated variants were found on both a Neandertal haplotype and several haplotypes present in Africa. Based on allele frequency differences between EBV-associated SNPs restricted to the Neandertal core haplotype and those also present on the other five core haplotypes, we estimate that the combined frequency of the latter is below 2.5% in the GWAS cohort and therefore a minor contributor to the association signal, which is predominantly driven by the archaic haplotype. Taking these observations together, we considered both loci on chromosome 17 to be promising candidates for associations driven by archaic haplotypes and examined both loci in more detail.

Functional Investigation of Archaic Association Loci

In total, four loci on chromosome 17 showed significant associations with EBV levels, two of which were not linked to archaic haplotypes and two corresponding to the aforementioned archaic haplotypes. The locus marked by the lead aSNP chr17:39,893,872_C/T represented the chromosome-wide significant association. Both archaic haplotypes were EBV-specific and not significantly associated with the levels of any of the other four tested viruses ($P > 0.05$). Furthermore, both haplotypes were associated with a higher EBV load (Fig. 1, Table S1). This observation was particularly noteworthy since 15 of the 18 significant archaic haplotype associations showed the same directional pattern (exact binomial test, $P = 7.5 \times 10^{-3}$, Fig. 1). We note that the viral load distribution—centered near zero with many non-carriers—is expected to produce a modest bias toward positive effect estimates for low-frequency variants. Consistent with this, the genome-wide associations in the viral load GWAS show a weak, nonsignificant correlation between allele frequency and effect direction (Spearman's rho = 0.25, $P = 0.09$). Thus, while this artifact may contribute to the observed directional bias, it is unlikely to fully account for it.

To evaluate whether the viral load associations of the two archaic haplotypes might have wider phenotypic relevance, we explored their association profiles across other traits in the Open Targets Platform (Materials and Methods). To address this task, we examined whether the two lead aSNPs were statistically associated with other lead GWAS signals ($P < 10^{-8}$, Table S4, Materials and Methods). Lead aSNP chr17:4,664,269_T/C was associated with 15 traits, several of which were repetitive, including 8 associated with lower eosinophil levels, and other traits including blood pressure and lower risk for nasal polyps. For lead aSNP chr17:39,893,872_C/T, we found one significant association with a lower risk of a thyroid gland disorder (Table S4). The associations and directional effects for both loci suggest a broader link of the candidate haplotypes to immune response processes, potentially including reduced immune-mediated pathology.

Our analysis suggests that archaic haplotypes in the MHC are generally not the driver of the observed viral load GWAS associations. Hence, any evaluation of co-associations with other traits would be difficult to interpret. However, the MHC harbors thousands of HLA alleles, many of which are known to influence immune function (Dendrou et al. 2018). Therefore, we explored whether any of the significantly viral load-associated archaic haplotypes in the MHC are linked to any of those alleles. To address this task, we leveraged HLA allele annotations from the 1000 Genomes cohort and computed LD with aSNPs on the archaic candidate haplotypes (Materials and Methods) (Abi-Rached et al. 2018). We found that 113 aSNPs on one archaic haplotype showed large levels of LD ($0.85 < r^2 < 0.96$) with HLA-A*11:01 (Fig. S3). This haplotype exhibited the largest sequence similarity with the Denisovan and had previously been reported to be introgressed (Abi-Rached et al. 2011). This haplotype lies in a region of the MHC lacking LD information in the UK Biobank, which previously prevented a robust assessment of its contribution to association signals. The archaic haplotype is composed of 131 aSNPs and overlaps seven protein genes, including HLA genes *HLA-A*, *HLA-F* and *HLA-G*, as well as *RNF39* and *MOG*—all of which have a link to immune regulation (Materials and Methods, Table S5). Additionally, no phenotype data could be extracted, as the aSNPs tagging this haplotype were also absent from the Open Targets Platform. This result is consistent with those of aSNPs for 11 other haplotypes in our study, which were also not included in the GWAS databases we queried (Materials and Methods). The two remaining aSNPs that we detected were listed in the Open Targets Platform, but did not show any significant trait associations (Table S4).

Next, we screened variants associated with the two archaic haplotypes on chromosome 17 for their predicted molecular and phenotypic effects. To address this question, we ran ENSEMBL's variant effect predictor (McLaren et al. 2016) for all variants showing $r^2 > 0.9$ with the corresponding two lead aSNPs. For the archaic haplotype carrying chr17:4,664,269_T/C, we tested 87 variants, including 11 aSNPs (Table S5). We identified links to three protein-coding genes, Arrestin Beta 2 (*ARRB2*), Arachidonate 15-Lipoxygenase (*ALOX15*),

and Proline, Glutamate, and Leucine Rich Protein 1 (*PELP1*). While no variant overlapped the genic parts of *ARRB2* or *ALOX15*, six variants were located in either exonic or UTR regions of *PELP1*. One of those six variants was a missense aSNP (chr17:4,671,456_T/A, $r^2 = 0.99$ to chr17:4,664,269_T/C), introducing a Serine instead of the Threonine (Table S5). Because of the similar characteristics of both amino acids, this substitution likely results in milder structural and functional effects on the protein. The archaic lead SNP chr17:4,664,269_T/C was also part of 23 credible eQTL sets, involving *ARRB2* (8 sets), *ALOX15* (12 sets), and *PELP1* (3 sets) (Table S6, Materials and Methods). While affected tissues for *ARRB2* and *PELP1* were predominantly in an immune/epithelial/neural context, eQTLs for *ALOX15* involved a more diverse range of tissues, suggesting a broader regulatory role of the archaic haplotype. Notably, both *ARRB2* and *PELP1* were included among the RNA virus-interacting proteins investigated in a previous study that identified signatures of adaptive introgression in this class of genes (Enard and Petrov 2018), indicating that this locus might have a broader link to host–virus interaction. Based on those results, variants associated with the archaic haplotype are predicted to have potential regulatory function for all three overlapping protein-coding genes and also protein-sequence-altering effects for *PELP1*.

Next, for the archaic haplotype carrying aSNP chr17:39,893,872_C/T, we tested 44 variants, including 22 aSNPs. Those variants were located in proximity of five protein-coding genes: Gasdermin B (*GSDMB*), Ikaros Family Zinc Finger 3 (*IKZF3*), Leucine Rich Repeat Containing 3C (*LRRC3C*), ORMDL sphingolipid biosynthesis regulator 3 (*ORMDL3*), and Zona pellucida binding protein 2 (*ZPBP2*). Similarly to variants associated with the first archaic haplotype, the majority of variants (38 of 44) were located in nongenic regions. Among the remaining six variants, four variants were located in 3′UTRs (*ZPBP2*: one variant, *ORMDL3*: three variants) and one situated in the 5′UTR of *GSDMB*. Furthermore, we found that the lead aSNP was associated with four eQTLs, three of which were with *GSDMB* (two brain tissues and esophagus muscularis mucosa, Table S6) and one in B cells for *IKZF3*. Additionally, the archaic allele of aSNP chr17:39,906,250_T/C ($r^2 = 0.95$ with lead aSNP) introduced a substitution of a Valine to an Isoleucine in *GSDMB*—a substitution predicted to be mild in its effects on protein structure and function due to the similar properties of both amino acids. Furthermore, seven tested variants of the first haplotype and one tested variant of the second haplotype were located in regulatory elements such as enhancers and open chromatin regions, suggesting further regulatory potential for variants associated with both archaic haplotypes (Table S5).

Assessment of Selection Signatures

Immune-associated genetic variation has frequently been a target of natural selection (Quintana-Murci 2019). We therefore explored whether any of our candidate loci exhibit selection signatures as well. First, we explored their frequency distribution in present-day populations. We found that a Neandertal haplotype carrying one of the lead aSNPs (chr17:4,664,269_T/C) showed a frequency of 22.1% in Europeans, which put it among the second percentile of archaic haplotypes in that population (Materials and Methods, Fig. 2). In East and South Asians, this haplotype was found to be in the 14th and 74th percentile, respectively, of the corresponding archaic haplotype frequency distributions. Haplotypes carrying the second lead aSNP on chromosome 17 (chr17:39,893,872_C/T) were detected only in Europeans and South Asians, and their frequencies fell well within the respective frequency ranges (Europe: 3.9%, 31st percentile; South Asia: 1.8%, 44th percentile). In addition, we also explored the frequency distributions of the HLA-A*11:01-associated haplotype and found that its frequencies in East (20.8%) and South (16.4%) Asians fall within the top 5% of detected archaic haplotypes in both populations (East Asia: 4.0 percentile, South Asia: 1.6 percentile). In Europeans, the haplotype was found at a lower frequency of 5.8% (20th percentile).

In addition, we explored all three haplotypes for additional signatures of selection. To evaluate this possibility, we investigated their reconstructed allele frequencies in the AGES database based on the genomes of ancient West Eurasian modern humans living over the last ~10,000 years (Materials and Methods). We would like to note that the composition of this database, with ancient individuals of predominantly European ancestry, is well matched to present-day European reference populations but not East and South Asian populations we included in our frequency analysis of present-day populations. For aSNPs associated with the HLA-A-linked haplotype, no information was available and no significant allele trajectory changes were detected for chr17:39,893,872_C/T ($P = 0.56$, model-based test for deviation from neutrality using allele age estimates, see Materials and Methods). However, for aSNP chr17:4,664,269_T/C, we identified a reconstructed allele frequency trajectory with significant changes over the last ~10,000 years. The trajectory showed a substantial decline of the archaic allele over the last 11,200 years from an estimated 42.6% to a present-day value of 22.3% ($P = 1.0 \times 10^{-5}$, Fig. 2, Materials and Methods). The local selection signal is tightly linked to our candidate aSNP and thus directly tied to the viral-load association (Fig. 2c/d). It is notable

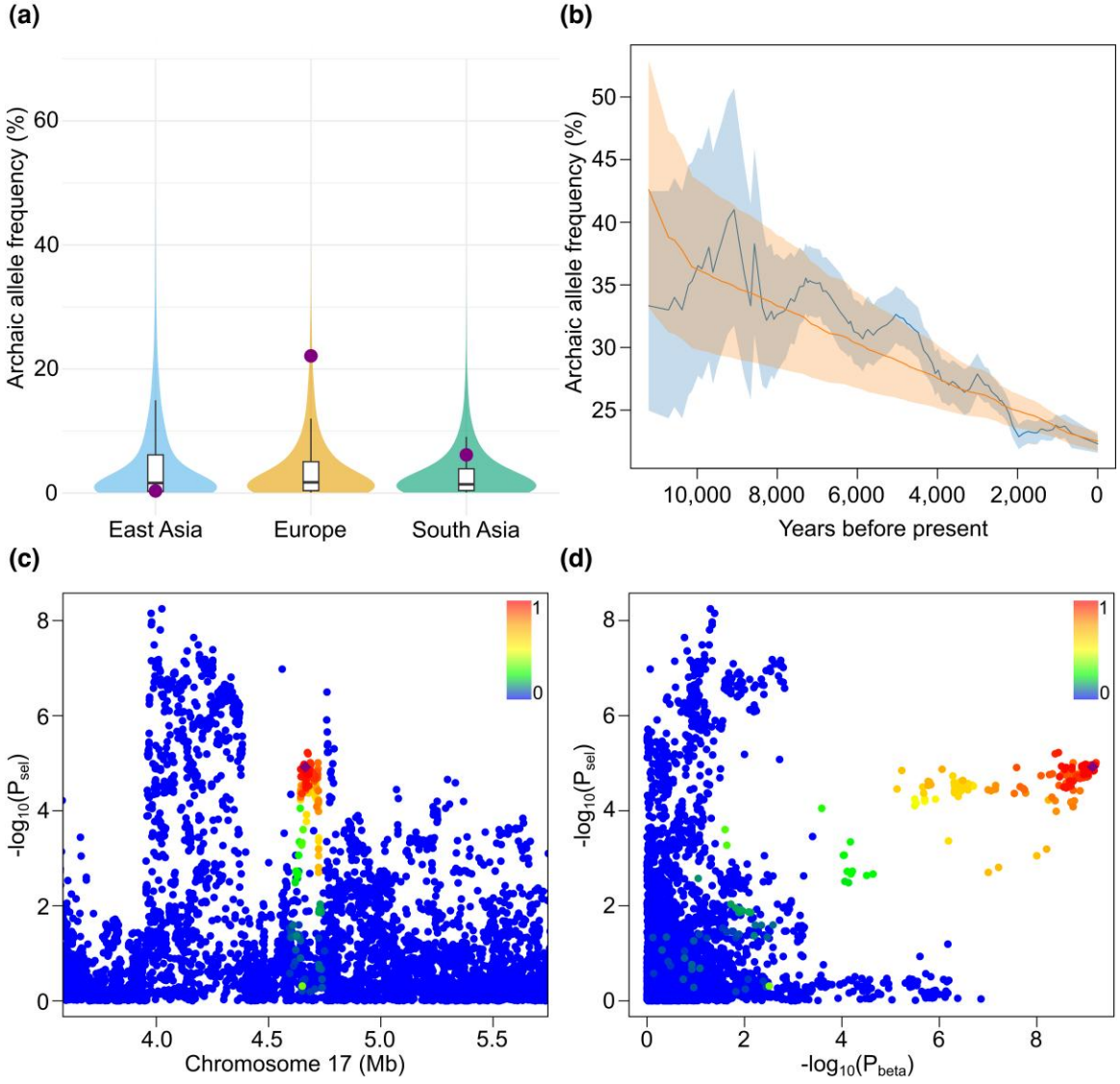

Fig. 2. Signatures of selection for archaic haplotype tagged by aSNP chr17:4,664,269_T/C. a) Frequency distribution of archaic haplotypes in three Eurasian populations. The frequency of the archaic haplotype associated with candidate aSNP chr17:4,664,269_T/C in each population is highlighted in purple. b) Empirical (blue) and statistically modeled (orange) archaic allele frequency trajectory for chr17:4,664,269_T/C in ancient human genomes from Europe and the Near East over the last ~11,500 years. The 95% confidence interval regions are shown in light blue and light orange. c) Manhattan plot displaying the selection *P*-value distribution of variants in a ± 1 megabase window around chr17:4,664,269_T/C. Variants are colored according to their LD values (derived from ancient individuals cohort) and the chr17:4,664,269_T/C is highlighted in purple (diamond shape). d) Figure showing both the selection *P*-value and EBV association *P*-value for variants in a ± 1 megabase window around chr17:4,664,269_T/C. The LD-based coloring schema is identical to (c).

that, despite this frequency decline, the haplotype remains among the highest 2% of archaic haplotypes in present-day Europeans. To put this result further in perspective, a frequency of 42.6% would place it within the top 0.1% of haplotypes in present-day Europeans. Based on this observation, it is conceivable that the locus may have been subject to different modes of selection over time, with signatures of positive selection early on and signatures of negative selection later. In this light, we also examined whether our second candidate locus on chromosome 17 might have experienced a recent frequency decline. However, we found no significant frequency changes for this locus over the last ~10,000 years ($P = 0.56$). Finally, we assessed whether either of the chromosome 17 candidate haplotypes had been identified by a recent study that explored archaic haplotypes with changing allele-frequency trajectories over time (Iasi et al. 2024). Neither locus was

detected in that study, an absence that could also be due to differences in data availability, temporal resolution, or statistical power to infer such trajectories.

Discussion

In this study, we explored associations of introgressed archaic haplotypes with DNA levels of five common DNA viruses in blood-derived WGS data. We detected 18 genome-wide significant associations, most of which involved archaic haplotypes within the MHC (15) and with viral levels of EBV (12).

We found that MHC-associated archaic haplotypes were often in LD with stronger associated non-archaic variants, weakening their case as key association drivers. One exception was an archaic haplotype, which was associated with higher EBV levels and linked to HLA-A*11:01. Its elevated allele frequency in East and South Asian populations may reflect possible signatures of positive selection. However, due to missing LD data, this region was excluded from functional and selection analyses, restricting additional phenotypic and evolutionary interpretations. Nevertheless, previous studies have implicated a strong immunogenic role of HLA-A*11 on peptide recognition and presentation impacting CD8+ T cell recognition and control of EBV (de Campos-Lima et al. 1993; Midgley et al. 2003; Tang et al. 2012). The results of the studies also indicate a long-term host-pathogen co-evolutionary dynamic between HLA-A11*, EBV and virus's epitope evolution.

We detected two additional archaic haplotypes that were significantly associated with higher EBV levels. Both haplotypes were located on chromosome 17 and had aSNPs tightly linked to the association signal. The first haplotype carrying aSNP chr17:4,664,269_T/C, was functionally linked to three genes: *PELP1*, *ARRB2*, and *ALOX15*. *PELP1* acts as a nuclear coactivator and is involved in several regulatory processes and is known as an oncogene in some cancers (Sareddy and Vadlamudi 2016). The gene has no established link to viral immune response, but viruses often exploit transcriptional coactivators. *ARRB2* regulates G protein-coupled receptor signaling and is involved in multiple signaling pathways, such as MAPK, NF-kB and apoptosis (Kim et al. 2023). *ARRB2* has been linked to innate immune responses, T cell activity, inflammation, and HIV (Kee et al. 2024). Viruses often hijack G protein-coupled receptors to enter cells or modulate immunity (Montaner et al. 2013). *ALOX15* is an enzyme involved in fatty acid saturation that also plays a role in inflammation resolution (Ma et al. 2022). Both *ALOX15* and *ARRB2* also act as eQTL genes for the archaic haplotype across a diverse range of tissues, including blood. In addition, an aSNP on the archaic haplotype modifies the amino acid sequence of *PELP1*. Nevertheless, the high regulatory link of both *ALOX15* and *ARRB2* to the archaic haplotype, in addition to their known functions, positions them as potentially stronger candidate genes in the context of EBV load and the archaic haplotype than *PELP1*. The lead aSNP on the haplotype also showed significant association with blood pressure, blood cell type traits and nasal polyps—all of which have been reported in the context of altered EBV load, however, without any established causal links between the virus and traits (Tao et al. 1996; Richard et al. 2010; Lin et al. 2023).

The second EBV-associated archaic haplotype carrying aSNP chr17:39,893,872_C/T, overlapped five genes: *GSDMB*, *IKZF3*, *LRRC3C*, *ORMDL3* and *ZPBP2*. *ZPBP2* is a gene involved in sperm-egg binding without any direct link to immunity (Ghanami Gashti et al. 2021). *GSDMB* mediates inflammatory programed cell death (pyroptosis), which is a key antiviral defense where infected cells are cleared (Zhu et al. 2024). *IKZF3* is a key B cell transcription factor regulated by EBV super-enhancers in lymphoblastoid cells (Zhou et al. 2015; Verhoeven et al. 2019). It has been associated with EBV-driven transcriptional processes that support viral response and cell proliferation. In general, *IKZF3* is a critical component for B cell differentiation and development with a direct link to pathogen and viral response (Lazarian et al. 2021). In addition, while its exact role in viral immune response is yet to be determined, the LRR domain that is also a key component of *LRRC3C* has often been linked to pathogen recognition (Li and Wu 2021). Finally, *ORMDL3* regulates sphingolipid metabolism and is involved in endoplasmic reticulum (ER) stress response, autophagy, and inflammation (Luthers et al. 2020). Both sphingolipids and ER stress are often exploited by viruses for cell entry and replication. Among these genes, *GSDMB* emerged as the strongest candidate interaction partner for the introgressed haplotype in the context of virus load, with aSNPs affecting regulatory function and protein sequence. However, given their immune-related function, it is also conceivable that *IKZF3*, *LRRC3C* or *ORMDL3* interact with the archaic haplotype in the context of EBV load. The lead aSNP of this haplotype is also significantly associated with a higher risk of an undefined thyroid gland disorder. Some thyroid gland autoimmune disorders have been described in the context of altered EBV levels (Moghoofei et al. 2019). However, just like for the haplotype discussed before, associations between EBV levels and thyroid disorders have been reported to be correlated without a direct causal mechanism.

The EBV associations for both chromosome 17 haplotypes and the HLA-A*11:01-linked haplotype, as well as 12 of the remaining 15 DNA virus associations, showed

increased viral loads in carriers of the respective archaic alleles. And although a direct link to the peak association signal remains uncertain for many of the archaic loci, the consistent directional bias suggests a nonrandom pattern that would be unlikely under the assumption of no functional contribution of archaic DNA to viral load variation. However, the phenotypic associations with a higher risk of thyroid disorders and nasal polyps are notable, as both conditions share similar patterns of inflammation, autoimmune tendencies, and possibly viral or environmental triggers that drive chronic immune activation. This overlap might suggest that DNA virus levels reflect overall health status rather than a direct effect on antiviral immune response. This is particularly relevant in the case of EBV, as the virus has been implicated in several diseases, including multiple sclerosis, lymphoma, stomach cancer, and nasopharyngeal carcinoma (Gandhi et al. 2004; Dogan et al. 2014; Pannone et al. 2014; Yau et al. 2014; Bjornevik et al. 2022), potentially providing a connection between virus load and broader disease risk.

However, the ongoing host-pathogen arms race—with pathogen strains continually emerging, replacing one another, and imposing shifting selective pressures on hosts—has also been linked to processes such as frequency-dependent selection (Wegner et al. 2019; Palmer et al. 2022). The directional effects on DNA virus load observed in our study may therefore not necessarily reflect those in Neandertals tens of thousands of years ago, when different viral strains were circulating. Consistent with this, a recent study proposed that the functional role of some archaic introgressed haplotypes may have shifted over time (Iasi et al. 2024), whereby previously causal and potentially advantageous archaic alleles lose their functional relevance, and linked nonarchaic alleles take over that role. Consistent with this hypothesis, among the 13 archaic haplotypes that showed moderate LD ($0.2 < r^2 < 0.8$) with more strongly associated nonarchaic variants, the strongest nonarchaic variant showed the same directional effect in all cases and a higher allele frequency in ten cases (Table S7). While this imbalance in allele frequencies is not statistically significant (binomial test, $P = 0.09$), it is tentatively consistent with the proposed model.

In general, the patterns of DNA virus-associated archaic haplotypes differ to those linked to RNA virus response, several of which show signatures of positive selection (Mendez et al. 2012; Quach et al. 2016; Sams et al. 2016; Enard and Petrov 2018). In general, DNA and RNA viruses have distinct characteristics. While RNA viruses often show a seasonal pattern, with an acute and localized infection signature, many DNA viruses are not seasonal, with latent and systemic disease characteristics (Griffin 2022). Our results might indicate that Neandertal immune systems were optimized for acute rather than chronic infections. Moreover, previous studies have shown that introgressed Neandertal DNA carries adaptive signatures related to UV exposure and seasonality (Reilly et al. 2022). This parallel between environmental adaptation and RNA virus seasonality may help explain the differing selective pressures and phenotypic patterns observed for RNA and DNA virus responses.

However, our results also indicate that their evolutionary interpretation might be challenging and complex. For example, we found that one of the haplotypes on chromosome 17 (lead aSNP chr17:4,664,269_T/C) shows an archaic allele frequency of 22% in present-day Europeans—a value that places it within the top 2% of archaic haplotypes in that population today. Despite its current high frequency, it has actually declined by nearly half over the past ~10,000 years, suggesting that it was rather negative selection acting on it during that time frame. Its high frequency ~10,000 years ago also suggests that the haplotype might have been under positive selection before. The past 10,000 years largely encompass the cultural transition to agriculture, which may have altered pathogen exposure. It is possible that changes in environmental conditions like this or selective tradeoffs shifted over time, painting a more complex picture of the fate of DNA virus-associated introgressed DNA in modern humans (Harrison et al. 2019). We also found that several top-associated SNPs at this locus segregate on nonarchaic haplotypes in African populations, indicating that archaic introgression is not the sole source of variation contributing to DNA virus-related phenotypes at this locus in present-day humans. Such patterns suggest that while archaic introgression may have contributed functionally relevant alleles, the locus itself may have been a repeated target of selection across human populations.

Because several key datasets in our analysis—most notably the viral-load GWAS cohort, the GWAS database used to identify co-associations with candidate haplotypes, and the ancient modern human samples used for allele-frequency reconstruction—are predominantly of European ancestry, our study was restricted to archaic variants segregating in Europeans. This limitation implies that we were unable to analyze a substantial fraction of Neandertal DNA not present in Europeans, as well as most Denisovan DNA that segregates in present-day populations. Hence, further computational and experimental work using datasets of a wider range of ancestries will be needed to refine our understanding of how DNA virus immune response signatures relate to introgressed archaic DNA in present-day humans. Nevertheless, our findings provide a new

perspective on the role of archaic DNA in shaping antiviral immunity, suggesting that immune response signatures may differ between virus types and that distinct immune components in archaic humans may have been better or worse equipped to handle specific viral challenges.

Materials and Methods

Detection of Archaic Haplotypes With Viral Level Association

We downloaded publicly available GWAS summary statistics for viral level measurements of five viruses from https://data.broadinstitute.org/lohlab/virome_summary_statistics/blood/. We focused on datasets derived from the UK Biobank, as LD matrices required for downstream analyses were available for this cohort at https://registry.opendata.aws/ukbb-ld (accessed on 2025 October 13). We conducted all analyses based on the human genome version hg38. In instances where datasets provided information in hg19 format, we used liftOver to convert those coordinates to hg38 (Hinrichs et al. 2006). All GWAS had been generated using the BOLT-LMM linear mixed model method (Loh et al. 2015). For our significance assessment, we applied a *P*-value threshold of 10^{-9} to account for multiple testing across the five GWAS we analyzed. Wherever available, we applied the threshold to the P_BOLT_LMM *P*-value, and in all other cases, to P_BOLT_LMM_INF. We explored significant viral level associations for archaic haplotypes that had previously been identified in 1000 Genomes Eurasian populations (Yermakovich et al. 2024). Haplotypes in this list were annotated as carrying aSNPs that show allele-sharing signatures consistent with archaic admixture. Furthermore, archaic haplotypes detected by the method used by Yermakovich et al. were highly overlapping those of other introgression maps (Dannemann et al. 2020). These signatures were defined by the absence of one of the two alleles in the African Yoruba population and the presence of that same allele in at least one of the three high-coverage Neandertal genomes (Prüfer et al. 2017, 2014; Mafessoni et al. 2020) or in the Denisovan genome (Meyer et al. 2012). In addition, aSNPs had been required to show high levels of LD with other aSNPs across haplotypes exceeding the length expected from incomplete lineage sorting (Yermakovich et al. 2024). In total, 378,433 aSNPs were used to screen for the presence of archaic haplotypes in the DNA load GWAS data.

Haplotype Network Analysis

We performed a haplotype network analysis for two candidate loci on chromosome 17 that were linked to aSNPs chr17:4,664,269_T/C and chr17:39,893,872_C/T. In both instances, we considered the genomic region encompassing the range of variants that were found to be in LD of $r^2 > 0.9$ in the UK Biobank cohort. Within those regions, we extracted all SNPs available in the 1000 Genomes cohort and generated haplotypes for all 1000 Genomes individuals (phase 3). Next, we clustered all haplotypes into core haplotypes using the vsearch software (Rognes et al. 2016), choosing the largest nucleotide distance that resulted in exactly ten core haplotypes. Based on the resulting core haplotypes, we generated haplotype networks using the R package pegas (Paradis 2010).

Molecular and Phenotypic Annotation of Archaic Candidate Haplotypes

We screened for trait and eQTL associations for candidate aSNPs in the OpenTargets database (Buniello et al. 2025). We annotated all trait and eQTL associations in which the candidate aSNP was part of a credible set, and the corresponding lead variant of the respective GWAS/eQTL had a *P*-value below 10^{-8}. Furthermore, we annotated predicted molecular effects for all variants in high LD ($r^2 > 0.9$) with the two candidate aSNPs on chromosome 17 and 131 aSNPs associated with an archaic haplotype in LD with HLA-A*11:01 using ENSEMBL's variant effect predictor (McLaren et al. 2016). In addition, we explored LD patterns between MHC archaic haplotypes associated with viral load and HLA alleles. We conducted this analysis using the HLA annotation of individuals from the 1000 Genomes cohort (1000 Genomes Project Consortium et al. 2015). As HLA alleles were not directly annotated to haploid sequences in this cohort, we calculated the r^2 as the squared Spearman's correlation coefficient based on genotypes converted into three categories: 0 (no alternative allele), 1 (heterozygous), and 2 (two alternative alleles), along with the counts of a given HLA allele (0, 1, 2) in each individual. We performed this calculation for all aSNPs associated with archaic haplotypes within the MHC that showed a significant association with viral load.

Frequency Analysis for Archaic Haplotypes

We explored the frequencies for three archaic candidate haplotypes within Eurasian super-populations from the 1000 Genomes cohort. Each archaic haplotype's frequency within a given super-population was calculated based on the number of identified archaic segments among individuals within the respective haplotype (Yermakovich et al. 2024). Frequency outlier statistics for those haplotypes were computed based on the distribution of frequencies for all detected archaic haplotypes within a given super-population. Furthermore,

for the candidate aSNP chr17:4,664,269_T/C, we explored the reconstructed allele frequency trajectory and selection information from the AGES database (Akbari et al. 2026). AGES is a public database that uses over 15,000 ancient and 6,000 present-day modern human genomes from Western Eurasia to trace genetic variants over the past 10,000 years. Allele frequency changes have been assessed for signatures of selection using a generalized linear mixed model and *P*-values derived from a test whether a modeled selection coefficient is zero.

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Acknowledgments

We are thankful to Kaur Alasoo for drawing our attention to the preprint by Kamitaki et al., that prompted this study. Data analyses were carried out in part in the High-Performance Computing Center of the University of Tartu.

Author Contributions

Rutvi Rajpara (Data curation [supporting], Formal analysis [supporting], Investigation [supporting], Writing—review & editing [supporting]), Sofiia Stefaniia Polishchuk (Data curation [supporting], Formal analysis [supporting], Investigation [supporting]), Danat Yermakovich (Data curation [supporting], Formal analysis [supporting], Writing—review & editing [supporting]), Michael Dannemann (Conceptualization [lead], Data curation [lead], Formal analysis [lead], Investigation [lead], Methodology [lead], Project administration [lead], Supervision [lead], Visualization [lead], Writing—original draft [lead], Writing—review & editing [lead]).

Funding

This work was supported by the Ministry of Education and Research Centres of Excellence grant TK214 "Centre of Excellence for Personalised Medicine" and the Estonian Research Council grant PRG3113.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability

The data underlying this article are available in the article and in its online supplementary material. Furthermore, R scripts related to the data analyses of this study can be found at https://github.com/SillySabertooth/DNA_virus_load_Neand_Intrgoression.

Literature Cited

1000 Genomes Project Consortium, et al. A global reference for human genetic variation. Nature. 2015:526:68–74. https://doi.org/10.1038/nature15393.

Abi-Rached L, et al. The shaping of modern human immune systems by multiregional admixture with archaic humans. Science. 2011:334:89–94. https://doi.org/10.1126/science.1209202.

Abi-Rached L, et al. Immune diversity sheds light on missing variation in worldwide genetic diversity panels. PLoS One. 2018:13:e0206512. https://doi.org/10.1371/journal.pone.0206512.

Akbari A, et al. Ancient DNA reveals pervasive directional selection across West Eurasia. Nature. 2026. https://doi.org/10.1038/s41586-026-10358-1.

Barreiro LB, Quintana-Murci L. Evolutionary and population (epi)-genetics of immunity to infection. Hum Genet. 2020:139:723–732. https://doi.org/10.1007/s00439-020-02167-x.

Bjornevik K, et al. Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. Science. 2022:375:296–301. https://doi.org/10.1126/science.abj8222.

Buniello A, et al. Open Targets Platform: facilitating therapeutic hypotheses building in drug discovery. Nucleic Acids Res. 2025:53:D1467–D1475. https://doi.org/10.1093/nar/gkae1128.

Bycroft C, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018:562:203–209. https://doi.org/10.1038/s41586-018-0579-z.

Dannemann M, et al. Human stem cell resources are an inroad to Neandertal DNA functions. Stem Cell Reports. 2020:15:214–225. https://doi.org/10.1016/j.stemcr.2020.05.018.

de Campos-Lima PO, et al. HLA-A11 epitope loss isolates of Epstein-Barr virus from a highly A11+ population. Science. 1993:260:98–100. https://doi.org/10.1126/science.7682013.

de Oliveira Neto NF, et al. The emergence of saliva as a diagnostic and prognostic tool for viral infections. Viruses. 2024:16:1759. https://doi.org/10.3390/v16111759.

Dendrou CA, Petersen J, Rossjohn J, Fugger L. HLA variation and disease. Nat Rev Immunol. 2018:18:325–339. https://doi.org/10.1038/nri.2017.143.

Dogan S, et al. Human papillomavirus and Epstein-Barr virus in nasopharyngeal carcinoma in a low-incidence population. Head Neck. 2014:36:511–516. https://doi.org/10.1002/hed.23318.

Dotto-Maurel A, Arzul I, Morga B, Chevignon G. Herpesviruses: overview of systematics, genomic complexity and life cycle. Virol J. 2025:22:155. https://doi.org/10.1186/s12985-025-02779-7.

Enard D, Petrov DA. Evidence that RNA viruses drove adaptive introgression between Neanderthals and modern humans. Cell. 2018:175:360–371.e13. https://doi.org/10.1016/j.cell.2018.08.034.

Gandhi MK, Tellam JT, Khanna R. Epstein-Barr virus-associated Hodgkin's lymphoma. Br J Haematol. 2004:125:267–281. https://doi.org/10.1111/j.1365-2141.2004.04902.x.

Ghanami Gashti N, et al. The germ cell-specific markers ZPBP2 and PGK2 in testicular biopsies can predict the presence as well as the quality of sperm in non-obstructive azoospermia patients.

Reprod Sci. 2021:28:1466–1475. https://doi.org/10.1007/s43032-020-00427-9.
Griffin DE. Why does viral RNA sometimes persist after recovery from acute infections? PLoS Biol. 2022:20:e3001687. https://doi.org/10.1371/journal.pbio.3001687.
Harrison GF, et al. Natural selection contributed to immunological differences between hunter-gatherers and agriculturalists. Nat Ecol Evol. 2019:3:1253–1264. https://doi.org/10.1038/s41559-019-0947-6.
Hinrichs AS, et al. The UCSC genome browser database: update 2006. Nucleic Acids Res. 2006:34:D590–D598. https://doi.org/10.1093/nar/gkj144.
Iasi LNM, et al. Neanderthal ancestry through time: insights from genomes of ancient and present-day humans. Science. 2024:386:eadq3010. https://doi.org/10.1126/science.adq3010.
Jones JE, Le Sage V, Lakdawala SS. Viral and host heterogeneity and their effects on the viral life cycle. Nat Rev Microbiol. 2021:19:272–282. https://doi.org/10.1038/s41579-020-00449-9.
Kamitaki N, Tang D, McCarroll SA, Loh P-R. The DNA virome varies with human genes and environments. Nature. 2026. https://doi.org/10.1038/s41586-026-10288-y.
Kee TR, et al. The multifaceted functions of β-arrestins and their therapeutic potential in neurodegenerative diseases. Exp Mol Med. 2024:56:129–141. https://doi.org/10.1038/s12276-023-01144-4.
Kim JY, et al. β-arrestin 2 negatively regulates lung cancer progression by inhibiting the TRAF6 signaling axis for NF-κB activation and autophagy induced by TLR3 and TLR4. Cell Death Dis. 2023:14:422. https://doi.org/10.1038/s41419-023-05945-3.
Lazarian G, et al. A hotspot mutation in transcription factor IKZF3 drives B cell neoplasia via transcriptional dysregulation. Cancer Cell. 2021:39:380–393.e8. https://doi.org/10.1016/j.ccell.2021.02.003.
Li D, Wu M. Pattern recognition receptors in health and diseases. Signal Transduct Target Ther. 2021:6:291. https://doi.org/10.1038/s41392-021-00687-0.
Lin K, et al. Association between hypertension and Epstein-Barr virus reactivation among the population in a high-risk area for nasopharyngeal carcinoma. Virus Res. 2023:331:199117. https://doi.org/10.1016/j.virusres.2023.199117.
Loh P-R, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nat Genet. 2015:47:284–290. https://doi.org/10.1038/ng.3190.
Luthers CR, Dunn TM, Snow AL. ORMDL3 and asthma: linking sphingolipid regulation to altered T cell function. Front Immunol. 2020:11:597945. https://doi.org/10.3389/fimmu.2020.597945.
Ma X-H, et al. ALOX15-launched PUFA-phospholipids peroxidation increases the susceptibility of ferroptosis in ischemia-induced myocardial damage. Signal Transduct Target Ther. 2022:7:288. https://doi.org/10.1038/s41392-022-01090-z.
Mafessoni F, et al. A high-coverage Neandertal genome from Chagyrskaya Cave. Proc Natl Acad Sci U S A. 2020:117:15132–-15136. https://doi.org/10.1073/pnas.2004944117.
McLaren W, et al. The Ensembl variant effect predictor. Genome Biol. 2016:17:122. https://doi.org/10.1186/s13059-016-0974-4.
Mendez FL, Watkins JC, Hammer MF. A haplotype at STAT2 Introgressed from neanderthals and serves as a candidate of positive selection in Papua New Guinea. Am J Hum Genet. 2012:91:265–274. https://doi.org/10.1016/j.ajhg.2012.06.015.
Meyer M, et al. A high-coverage genome sequence from an archaic Denisovan individual. Science. 2012:338:222–226. https://doi.org/10.1126/science.1224344.
Midgley RS, et al. HLA-A11-restricted epitope polymorphism among Epstein-Barr virus strains in the highly HLA-A11-positive Chinese population: incidence and immunogenicity of variant epitope sequences. J Virol. 2003:77:11507–11516. https://doi.org/10.1128/JVI.77.21.11507-11516.2003.
Moghoofei M, et al. Epstein-Barr virus and thyroid cancer: the role of viral expressed proteins. J Cell Physiol. 2019:234:3790–3799. https://doi.org/10.1002/jcp.27144.
Montaner S, Kufareva I, Abagyan R, Gutkind JS. Molecular mechanisms deployed by virally encoded G protein-coupled receptors in human diseases. Annu Rev Pharmacol Toxicol. 2013:53:331–354. https://doi.org/10.1146/annurev-pharmtox-010510-100608.
Neefjes J, Jongsma MLM, Paul P, Bakke O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. Nat Rev Immunol. 2011:11:823–836. https://doi.org/10.1038/nri3084.
Palmer WH, Telford M, Navarro A, Santpere G, Norman PJ. Human herpesvirus diversity is altered in HLA class I binding peptides. Proc Natl Acad Sci U S A. 2022:119:e2123248119. https://doi.org/10.1073/pnas.2123248119.
Pannone G, et al. The role of EBV in the pathogenesis of Burkitt's Lymphoma: an Italian hospital based survey. Infect Agent Cancer. 2014:9:34. https://doi.org/10.1186/1750-9378-9-34.
Paradis E. Pegas: an R package for population genetics with an integrated-modular approach. Bioinformatics. 2010:26:419–420. https://doi.org/10.1093/bioinformatics/btp696.
Prüfer K, et al. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature. 2014:505:43–49. https://doi.org/10.1038/nature12886.
Prüfer K, et al. A high-coverage Neandertal genome from Vindija Cave in Croatia. Science. 2017:358:655–658. https://doi.org/10.1126/science.aao1887.
Quach H, et al. Genetic adaptation and Neandertal admixture shaped the immune system of human populations. Cell. 2016:167:643–656.e17. https://doi.org/10.1016/j.cell.2016.09.024.
Quintana-Murci L. Human immunology through the lens of evolutionary genetics. Cell. 2019:177:184–199. https://doi.org/10.1016/j.cell.2019.02.033.
Rampersad S, Tennant P. Replication and expression strategies of viruses. In: Viruses. Elsevier; 2018. p. 55–82.
Reilly PF, Tjahjadi A, Miller SL, Akey JM, Tucci S. The contribution of Neanderthal introgression to modern human traits. Curr Biol. 2022:32:R970–R983. https://doi.org/10.1016/j.cub.2022.08.027.
Richard Y, et al. Changes in blood B cell phenotypes and Epstein-Barr virus load in chronically human immunodeficiency virus–infected patients before and after antiretroviral therapy. J Infect Dis. 2010:202:1424–1434. https://doi.org/10.1086/656479.
Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. PeerJ. 2016:4:e2584. https://doi.org/10.7717/peerj.2584.
Sabbaghian M, Gheitasi H, Shekarchi AA, Tavakoli A, Poortahmasebi V. The mysterious anelloviruses: investigating its role in human diseases. BMC Microbiol. 2024:24:40. https://doi.org/10.1186/s12866-024-03187-7.
Sams AJ, et al. Adaptively introgressed Neandertal haplotype at the OAS locus functionally impacts innate immune responses in humans. Genome Biol. 2016:17:246. https://doi.org/10.1186/s13059-016-1098-6.
Sareddy GR, Vadlamudi RK. PELP1: structure, biological function and clinical significance. Gene. 2016:585:128–134. https://doi.org/10.1016/j.gene.2016.03.017.

Shaw Stewart PD, Bach JL. Temperature dependent viral tropism: understanding viral seasonality and pathogenicity as applied to the avoidance and treatment of endemic viral respiratory illnesses. Rev Med Virol. 2022:32:e2241. https://doi.org/10.1002/rmv.2241.

Tang M, et al. The principal genetic determinants for nasopharyngeal carcinoma in China involve the HLA class I antigen recognition groove. PLoS Genet. 2012:8:e1003103. https://doi.org/10.1371/journal.pgen.1003103.

Tao Q, Srivastava G, Dickens P, Ho FC. Detection of Epstein-Barr virus-infected mucosal lymphocytes in nasal polyps. Am J Pathol. 1996:149:1111–1118.

Timmerman AL, Schönert ALM, van der Hoek L. Anelloviruses versus human immunity: how do we control these viruses? FEMS Microbiol Rev. 2024:48:fuae005. https://doi.org/10.1093/femsre/fuae005.

Trowsdale J, Knight JC. Major histocompatibility complex genomics and human disease. Annu Rev Genomics Hum Genet. 2013:14:301–323. https://doi.org/10.1146/annurev-genom-091212-153455.

Verhoeven RJA, et al. Epstein-Barr virus BART long non-coding RNAs function as epigenetic modulators in nasopharyngeal carcinoma. Front Oncol. 2019:9:1120. https://doi.org/10.3389/fonc.2019.01120.

Wegner F, Lassalle F, Depledge DP, Balloux F, Breuer J. Coevolution of sites under immune selection shapes Epstein-Barr virus population structure. Mol Biol Evol. 2019:36:2512–2521. https://doi.org/10.1093/molbev/msz152.

Yau TO, Tang C-M, Yu J. Epigenetic dysregulation in Epstein-Barr virus-associated gastric carcinoma: disease and treatments. World J Gastroenterol. 2014:20:6448–6456. https://doi.org/10.3748/wjg.v20.i21.6448.

Yermakovich D, et al. Denisovan admixture facilitated environmental adaptation in Papua New Guinean populations. Proc Natl Acad Sci U S A. 2024:121:e2405889121. https://doi.org/10.1073/pnas.2405889121.

Zeberg H, Jakobsson M, Pääbo S. The genetic changes that shaped Neandertals, Denisovans, and modern humans. Cell. 2024:187:1047–1058. https://doi.org/10.1016/j.cell.2023.12.029.

Zhou H, et al. Epstein-Barr virus oncoprotein super-enhancers control B cell growth. Cell Host Microbe. 2015:17:205–216. https://doi.org/10.1016/j.chom.2014.12.013.

Zhu C, et al. The gasdermin family: emerging therapeutic targets in diseases. Signal Transduct Target Ther. 2024:9:87. https://doi.org/10.1038/s41392-024-01801-8.

Associate editor: Emilia Huerta-Sanchez