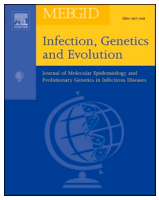
Infection, Genetics and Evolution 116 (2023) 105528

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid



Disentangling archaic introgression and genomic signatures of selection at human immunity genes

Alina Urnikyte *, Abigaile Masiulyte, Laura Pranckeniene, Vaidutis Kučinskas

Faculty of Medicine, Department of Human and Medical Genetics, Institute of Biomedical Sciences, Vilnius University, Santariskiu Street 2, Vilnius LT-08661, Lithuania

ARTICLE INFO

Keywords:
WGS
Lithuanian population
Positive selection
Immunity genes
Neanderthal
Introgression

ABSTRACT

Pathogens and infectious diseases have imposed exceptionally strong selective pressure on ancient and modern human genomes and contributed to the current variation in many genes. There is evidence that modern humans acquired immune variants through interbreeding with ancient hominins, but the impact of such variants on human traits is not fully understood. The main objectives of this research were to infer the genetic signatures of positive selection that may be involved in adaptation to infectious diseases and to investigate the function of Neanderthal alleles identified within a set of 50 Lithuanian genomes. Introgressed regions were identified using the machine learning tool ArchIE. Recent positive selection signatures were analysed using iHS. We detected high-scoring signals of positive selection at innate immunity genes (*EMB, PARP8, HLA—C,* and *CDSN*) and evaluated their interactions with the structural proteins of pathogens. Interactions with human immunodeficiency virus (HIV) 1 and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) were identified. Overall, genomic regions introgressed from Neanderthals were shown to be enriched in genes related to immunity, keratinocyte differentiation, and sensory perception.

1. Introduction

The evolutionary history of anatomically modern humans is complex, involving migration events, archaic introgression, and genetic adaptations (Dolgova and Lao, 2018). Despite progress in the healthcare system, infectious diseases caused by pathogens impose powerful global selective pressure that remains high today (e.g., the COVID-19 pandemic), with infectious diseases still representing one of the major causes of mortality (Cagliani and Sironi, 2013; Quintana-Murci, 2019). During the past 100,000 years, humans due to the out-of-Africa migration have been exposed to new environments, different climatic conditions, novel zoonotic pathogens, and local diets, to which they have been forced to adapt through the action of natural selection (Barreiro et al., 2008). Pathogens and infectious diseases have imposed exceptionally strong selective pressure on ancient and modern human genomes and contributed to the current variation in many genes (Barreiro and Quintana-Murci, 2010; Collen et al., 2022; Grossman et al., 2013; Karlsson et al., 2014; Quintana-Murci, 2019). Episodes of natural selection, as occur during epidemics of introduced infectious diseases, are expected to leave signatures in the genome (Barreiro and Quintana-Murci, 2010; Karlsson et al., 2014; Kerner et al., 2023; Quintana-Murci and Clark, 2013). The need to identify loci that have been targeted by natural selection in local populations and characterize how they influence adaptation to infectious diseases is increasing (Corona et al., 2018; Kerner et al., 2023; Owers et al., 2017). Loci under natural selection are more likely to harbour functional variants and can be prioritized in screening for associations with susceptibility or resistance to diseases and infections (Corona et al., 2018; Karlsson et al., 2014; Vasseur and Quintana-Murci, 2013). The identification of most phenotypes associated with variants showing signatures of selection is difficult, particularly for genetic variants conferring resistance to infectious diseases. In such cases, the best approach is to identify signatures of past natural selection in the genome, as opposed to the traditional approach of linking genetic variants with phenotypes. Furthermore, many immunity-related genes show patterns of genetic variation that are strongly associated with pathogen diversity (Fumagalli et al., 2011). Signals of adaptive evolution in immune-related genes tend to be recent and population specific, emphasizing the role of pathogens in local adaptation (Deschamps et al., 2016). Functional introgressed variants are another important source of adaptive variation (Yan et al., 2021). Modern Eurasians inherited 1% to 6% of their genomes from Neanderthal or Denisovan ancestors (Green et al., 2010; Meyer et al., 2012; Reich

* Corresponding author.
E-mail addresses: alina.urnikyte@mf.vu.lt (A. Urnikyte), laura.pranckeniene@mf.vu.lt (L. Pranckeniene), vaidutis.kucinskas@mf.vu.lt (V. Kučinskas).

https://doi.org/10.1016/j.meegid.2023.105528
Received 16 May 2023; Received in revised form 4 November 2023; Accepted 14 November 2023
Available online 15 November 2023

et al., 2010). Various genomic regions of adaptive introgression representing up to 64% of Neanderthal ancestry related to immunity, skin and hair pigmentation, and metabolism have been identified (Dannemann and Kelso, 2017; Sankararaman et al., 2014). For example, the *BNC2* gene associated with skin pigmentation (Jacobs et al., 2013) is a strong candidate for adaptive introgression (Vernot and Akey, 2014) whose archaic haplotype is present at a 70% frequency in Europeans (Racimo et al., 2015). There is evidence to suggest that modern humans have acquired advantageous variation through admixture with ancient hominins (Abi-Rached et al., 2011).

In this research, we sought to identify genetic signatures of positive selection that may be involved in resistance to infectious diseases in the Lithuanian population using whole-genome sequencing data. Genetic variation in innate immunity acquired through admixture with Neandertals was also evaluated. The idea of the study is not new, but it is absolutely relevant because the identification of positively selected genes and relationships with function is important to understand the adaptive evolution of modern humans. To date, many similar population genetics studies have been performed on large reference populations. The analysis of geographically specific regions and the characterization of loci targeted by selection can certainly contribute to progress in scientific knowledge and help us to understand host–pathogen coadaptation and impacts.

2. Materials and methods

2.1. Data

We made use of samples from 50 unrelated adult participants of Lithuanian nationality that were previously obtained for another project (Urnikyte et al., 2022). Written informed consent was obtained from all participants. DNA was extracted from whole blood (10 mL) samples using a QIAGEN GENTRA® Puregene® Blood Kit (Qiagen GmbH) according to the manufacturer's protocol. DNA concentration and quality were assessed using a NanoDropR ND-1000 spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE, USA)[9]. This work was approved by the Vilnius Regional Research Ethics Committee (No. 2020/6–1243-724, 22-06-2022).

Whole-genome sequencing (WGS) was performed on 100 ng of genomic DNA at the CeGaT company (Tubingen, Germany). Paired-end libraries were constructed using the TruSeq® Nano DNA Library Prep Kit (Illumina Inc., San Diego, CA, USA) and sequenced using 2 × 150 bp paired-end reads on an Illumina NovaSeq™ 6000 Sequencing System at coverage of 26.88–61.38× (an average of 36.27×). The sequenced filtered reads were aligned to the reference human genome hg19. The analysis of sequencing data was performed using the Illumina DRAGEN platform (v3.6.4).

2.2. Detection of positive selection

Whole-genome scans for footprints of selection were performed using the integrated haplotype homozygosity score (iHS) with the R package *rehh* (Gautier and Vitalis, 2012). The iHS statistic compares integrated extended haplotype homozygosity (EHH) (Sabeti et al., 2002) between derived and ancestral alleles within a population (Voight et al., 2006). Phased biallelic genetic markers are needed as the input. The phases of the haplotypes were estimated with the software shapeit4 (Delaneau et al., 2019). The ancestral and derived alleles were inferred by downloading ancestral annotation information for *Homo sapiens* from the 1000 Genomes Project based on EPO (Enredo, Pecan, Ortheus) (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase1/analysis_results/supporting/ancestral_alignments/human_ancestor_GRCh37_e59.tar.bz2). SNPs were excluded according to an MAF < 5%. The final dataset included 1,461,557 SNPs. For candidate region identification, we used overlapping sliding windows with a 100 kb width with an offset of 10 kb, and overlapping candidate windows were merged. Only windows with >5 extreme SNPs were retained.

2.3. Detection of introgressed Archaic variants

The identification of Neanderthal DNA fragments in 50 modern Lithuanian genomes was performed with ARCHaic Introgression Explorer (ArchIE), which is based on a logistic regression model and uses coalescent simulations to train a model to distinguish archaic fragments from nonarchaic fragments (Durvasula and Sankararaman, 2019). The *ArchIE* statistical method can use reference genomes from distant representatives (Prufer et al., 2014). Here, we used 108 Yoruban individuals from the 1000 Genomes Project Phase3 dataset (Auton et al., 2015). ArchIE takes phased genotype data as inputs, along with a file specifying the outgroup samples. Multiallelic SNPs were removed, and the phases of the haplotypes were estimated with the software shapeit4 (Delaneau et al., 2019). We applied ArchIE in 50 kb windows because this was the average introgressed archaic haplotype length after 2000 generations based on recombination frequency in simulations. For every DNA fragment, the probability of Neanderthal origin was calculated. Only fragments with more than a 99% probability were used for further analysis. Shared alleles between modern Lithuanians and the archaic genome (matches) were calculated as the match ratio between matching Neanderthal SNPs and the sum of matching and mismatching Neanderthal SNPs using step 4 of the *Sprime* protocol (Zhou and Browning, 2021). The Vindija Neanderthal genome was obtained from Prufer et al. (2017). The analysis workflow is summarised in Fig. S1.

2.4. Annotation

Functional gene and variant annotation in candidate selected and introgressed regions was performed with ANNOVAR v.20210123 (Wang et al., 2010) using GRCh37 (hg19), RefSeqGene, gnomAD, avsnp150 and CADD version 1.3 (Kircher et al., 2014). Human-pathogen interaction data were obtained from the BioGRID (Chatr-Aryamontri et al., 2013) database and used to identify human genes whose products interact with pathogens. The enrichment of biological processes in selected genes was tested using DAVID (Database for Annotation, Visualization, and Integrated Discovery) (Huang et al., 2009). We also performed two-sided Gene Ontology (GO) enrichment analyses (enrichment/depletion) and pathway annotation network tests with the Cytoscape (Shannon et al., 2003) plug-in ClueGo (Bindea et al., 2009) in the genes in introgressed regions. *P* values were corrected for multiple testing by the Benjamini–Hochberg (BH) procedure.

3. Results

3.1. Archaic introgression in the Lithuanian population

To identify segments of archaic ancestry in 50 modern unrelated Lithuanian genomes, we used a machine learning tool, ARCHaic Introgression Explorer (ArchIE), which makes use of training datasets to calculate a set of features that are potentially informative about introgression (Durvasula and Sankararaman, 2019). Genome data analysis with ArchIE identified a total of 1600 fragments among all autosomes that had lengths of 50 kb and covered approximately 80 Mb (2.5%) of the genome and showed a probability of being archaic greater than or equal to 99% (Fig. 1).

Among these fragments, 301 fragments showed matches ≥0.85 with Vindija Neanderthal genome, and 51 fragment showed matches ≥0.85 with the Altai Denisovan genome (Table S1). We plotted the two-way densities of match rates with the Vindija Neanderthal and Altai Denisovan genomes, as described in step 4 of the *Sprime* protocol (Zhou and Browning, 2021) (Fig. 2). We observed high matching with the Vindija Neanderthal genome (approximately 0.6) versus low matching with the Altai Denisovan genome (approximately 0.2). (See Fig. 3.)

For further analysis, we considered only genomic fragments

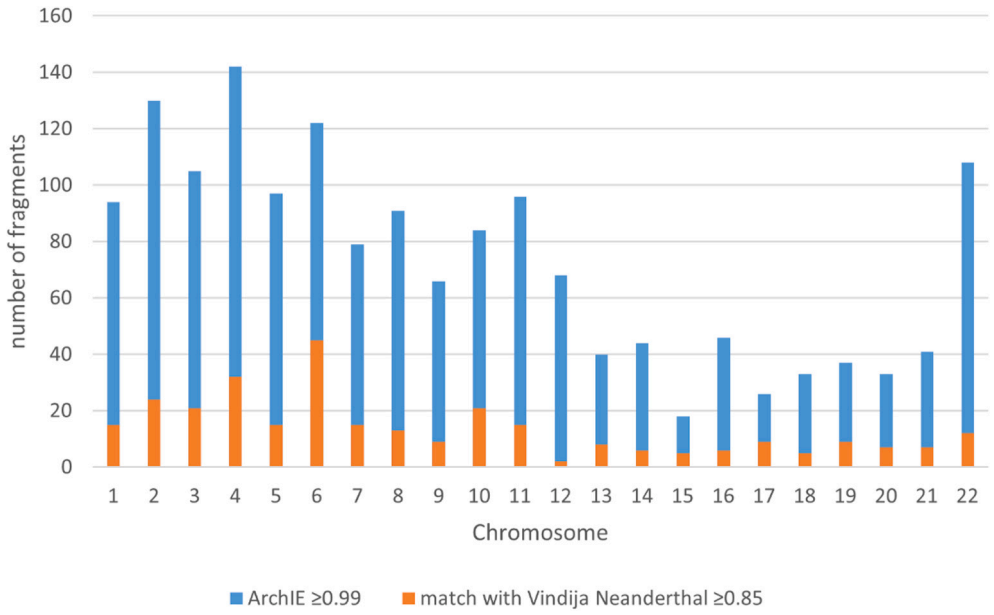


Fig. 1. Number of archaic fragments identified in Lithuanian WGS data. The columns in blue represent fragments that ArchIE inferred to be archaic with a probability greater than or equal to 99%, and orange segments matched the Vindija Neanderthal genome. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

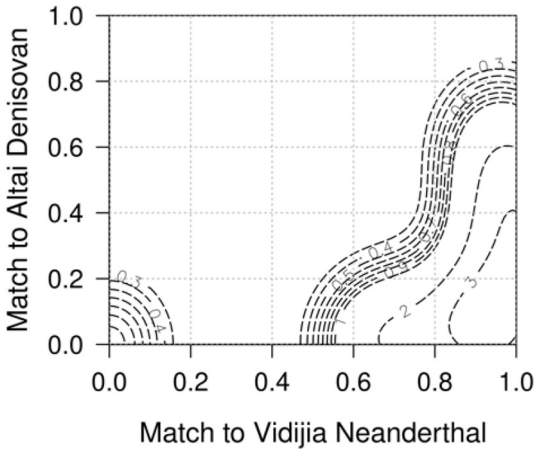


Fig. 2. Plot of the calculated match proportions of introgressed segments in the Vindija Neanderthal and Altai Denisovan genomes. Numbers inside the plots indicate the density value corresponding to each contour line.

identified as showing a Vindija Neanderthal origin. Introgressed sequences with matches ≥0.85 with the Vindija Neanderthal genome were annotated, and an enrichment pathway analysis was performed on the genes identified in the region (Table S2, Figs. S2 and 3).

Most of the pathways are involved in immune functions, in accord with previous studies showing that innate immunity genes are enriched in Neanderthal ancestry (Deschamps et al., 2016). Other significant pathways are related to the regulation of keratinocyte differentiation, antioxidant activity, positive regulation of interleukin-23 production, negative regulation of Rho protein signal transduction, and sensory perception (Table S4, S5). Next, we identified a list of immunity genes (IGs) by using the publicly available Reactome Pathway database (Fabregat et al., 2018) and InnateDB (Breuer et al., 2013). In total, 21 genes (overlapping putatively introgressed segments) of the innate and/or adaptive immune system were identified and annotated with DAVID (Huang et al., 2009) (Tables 1 and S5).

The largest number of Neanderthal introgressed fragments identified with the ArchIE program in Lithuanian WGS data were identified on chromosome 6, which contained *HLA* genes important in acquired immunity (*HLA-A, HLA—B, HLA—C, HLA-DRB1, HLA-DQA1*). The identified innate immunity-related genes were *REG3G, MAPK10, PHACTR2, HLA-A, HLA—B, HLA—C, HLA-DQB1, IL17A, IL17F, DLC1, MMP20, CLEC7A, CDH1,* and *TOM1*. Nonsynonymous variants were found in the immunity genes *HLA-A* (intronic variants: rs78306866, rs1136695, rs3173420), *HLA-B* (rs9266206), and *HLA-C* (rs1050147, rs1130838). According to Reactome Pathway Browser v.3.7, *HLA-A* participates in HIV infection pathway, host interaction of HIV factors (FDR: 8.52E-1); *CDH1* is related to Listeria monocytogene entry into the host cell (FDR: 4.22E-1); *DLC1* is related to HCMV infection (FDR: 8.63E-1); and *HLA—B, HLA—C, HLA-A, IL17A, IL17F, HLA—F, HLA-G,* and *HLA-H* participate in SARS-CoV-2 infection host interactions. Based on the pathogen–host interaction database PHILM2Web (Le et al., 2022), we identified a few human immune-related gene and pathogen associations. *HLA*-G and *HLA-DQB1* genes are related to human papillomavirus, *HLA-B* to human immunodeficiency virus, and *HLA-DRB1* to unidentified influenza virus.

3.2. Signals of positive selection mapped by iHS

We sought to identify regions that have evolved under recent positive selection in Lithuanians based on iHS (Voight et al., 2006) statistics using the R package *rehh* (Gautier and Vitalis, 2012). Fig. 4 shows positive selection signals identified in a Manhattan plot.

The positive and negative iHS values were considered to capture ancient and recent signals of selection, respectively. In total, 17 candidate genomic regions under selection that produced clusters of markers with outlier values were identified (Tables 2 and S5). Clusters of significant SNPs were observed on chromosomes 1–6, 9–12, 15, and 21. (Fig. 4). Thirteen of these regions have been identified in the selection catalogue of human adaptation (Murga-Moreno et al., 2019), and the remaining regions were novel and were involved in functions that could not be linked to specific phenotypes (Table 2). The region with the strongest signal on chromosome 21 contained the genes *TPTE*, *BAGE* and *BAGE2*. The *BAGE* (B melanoma antigen) gene family has been shown to be under purifying selection to eliminate deleterious amino acid changes. *BAGE* genes were generated via complex chromosome

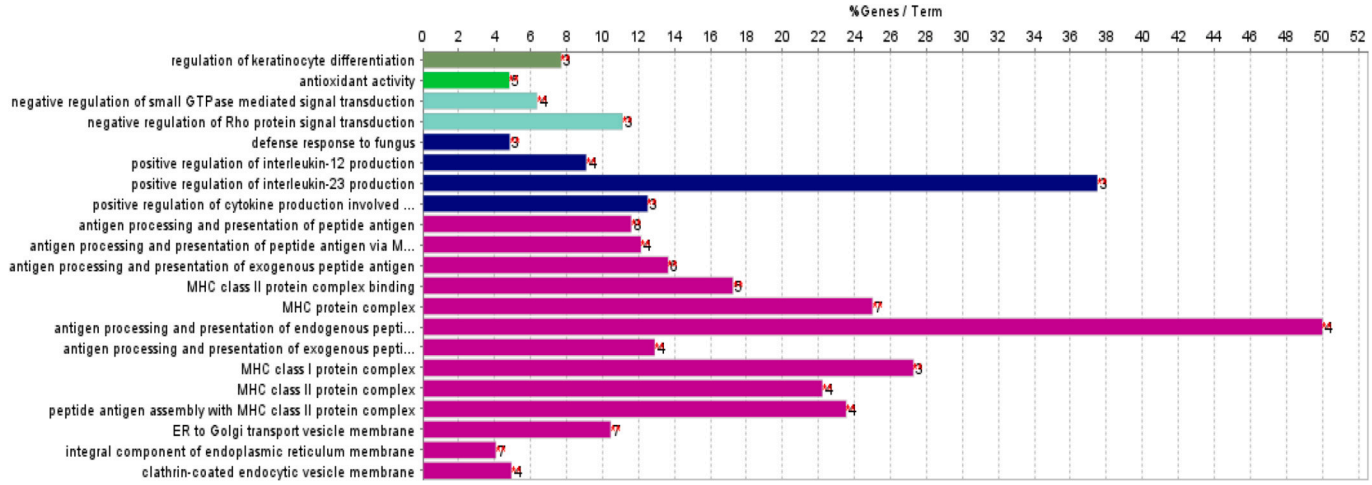


Fig. 3. Pathway enrichment analysis based on introgressed genes. Bar length is proportional to the percentage of genes identified as associated with the term, and the number of genes identified as associated with each term is shown. Pathways are coloured depending on their main biological functions. All terms are statistically significant (p value < 0.05, BH—FDR).

rearrangements that occurred in juxtacentromeric regions during hominoid evolution, which have functioned as birthplaces of novel genes (Ruault et al., 2003). *TPTE* and *BAGE2* are associated with Robertsonian Down syndrome (Shaw et al., 2008). It has been hypothesized, that genes involved in common diseases could often be targets of selection (Hancock et al., 2008). Another region with a strong signal on chromosome 1 (141600000–150,200,000) comprised the *NBPF* gene family. *NBPF* genes contain Olduvai domains, which have undergone the most dramatic copy-number increase of any protein-coding region in the human lineage. Their repeat numbers are strongly associated with the evolutionary expansion of brain volume, neuron counts and cognitive ability as well as with disorders of the autistic spectrum (Mangan et al., 2022).

Nonsynonymous variants were found within the region (129200000–131,100,000) on chromosome 3, which comprises the *COL6A4P2*, *COL6A5* and *COL6A6* genes encoding collagen VI chains. Two nonsynonymous variants in *COL6A5* (rs12488457: c.3838 A > C, Tr1280Pro, with a CADD (deleteriousness prediction score of single nucleotide variant, score of 10 or greater indicates that the SNP is predicted to be among the top 10% most deleterious variants of the human genome, while a score 20 or greater implies top 1% most deleterious) value of 23.2, and rs16827497: c.4765 T > C, Ser1589Pro, with a CADD value of 0.278) were identified. The SNP rs12488457 has been reported as an eQTL for *COL6A5* in testis ($P = 7.3 \times 10^{-21}$) and for *COL6A6* in the heart ($P = 2.5 \times 10^{-10}$), with the derived C allele implying lower expression in both cases. Lithuanians present a high frequency (0.82) for the derived C allele, which is found at 0.71 in Europeans and 0.15 in Africans according to 1000 Genomes Project (Auton et al., 2015). The SNP rs16827497 has been reported as an eQTL for *COL6A5* in testis ($P = 1.3 \times 10^{-15}$), for *COL6A6* in heart ($P = 3.9 \times 10^{-10}$), and for *COL6A4P2* in thyroid ($P = 0.000015$), artery-tibial ($P = 0.00013$), and oesophagus-mucosa ($P = 0.00018$), with the derived C allele implying lower expression according to the Genotype-Tissue Expression (GTEx) portal (https://gtexportal.org/home/; accessed on 14/03/23). Again, the highest frequency of the derived C allele was determined for Lithuanians (0.87), Europeans - 0.76, and Africans - 0.40 (Auton et al., 2015). The derived allele C of both variants rs12488457 and rs16827497 is present in Neanderthal genomes at frequencies 0.77 and 0.87, respectively (Sherry et al., 1999). A large region comprising olfactory receptor family genes was detected on chromosome 11. The nonsynonymous variant rs628524, causing a Ser171Asn substitution, with a CADD value of 5.6 in the *OR5M11* gene associated with proinsulin (Huyghe et al., 2013), was identified. A nonsynonymous variant detected in *HLA-C* (rs1130838: c.1087 A > C) has been associated with psoriasis and Behcet's disease (Lee et al., 2012). The C allele frequencies were 0.80, 0.68, and 0.75 in Lithuanian, European and African populations, respectively.

We found a region on chromosome 9 including the *BNC2* gene matching the Neanderthal genome, which is involved in skin pigmentation (Jacobs et al., 2013). This region has been shown to harbour introgressed Neanderthal haplotypes at an elevated frequency (Durvasula and Sankararaman, 2019; Sankararaman et al., 2014).

We detected signals of positive selection at three genes participating in immunity pathways, *HLA—C*, *CAPZA3,* and *TCF19*. A nonsynonymous variant was found in the *TCF19* gene (rs2073721:c.A631G: p. Met211Val), which encodes a protein with a PHD-type zinc finger domain involved in transcriptional regulation. The derived allele G is present at a high frequency of 0.81 in Lithuanians, 0.76 in Europeans, and 0.84 in Africans. Few mechanisms related to the role of *TCF19* in immune regulation have been reported (Yang et al., 2021). The *TCF19* gene is related to the immune response and encodes a protein involved in the proliferation and apoptosis of pancreatic beta cells (Hystad et al., 2007). There is evidence that the destruction of islet β-cells in type 1 diabetes (T1D) is a result of an immunoregulation disorder. *TCF19* may be one of the key players in maintaining such immunological balance (Cheung et al., 2011). The *TCF19* rs2073721 variant is associated with type 1 and type 2 diabetes (Brenner et al., 2020; Cheung et al., 2011; Spracklen et al., 2018). rs2073721 is located in the MHC region and has been reported to be a genuine variant associated with common variable immunodeficiency (CVID), which manifests as an insufficient quantity and quality of immunoglobulin, leading to susceptibility to bacterial infections (Maggadottir et al., 2015; Orange et al., 2011). The most significant eQTLs of rs2073721 are related to the oesophagus mucosa ($P = 7.0E-14$) and the tibial nerve ($P = 1.1E-80$) and with higher expression of the derived G allele.

To assess human-pathogen protein–protein interactions, we used the BioGRID (Chatr-Aryamontri et al., 2013) and PHILM2Web (Le et al., 2022) databases. We found an interaction of the human *HLA-C* gene with immunodeficiency virus (HIV) 1 and interactions of the *PARP8* and *EMB* genes with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).

4. Discussion

Although most alleles that introgressed from Neanderthals to modern humans have been eliminated by purifying selection (Juric et al., 2016), the genome still maintains adaptive archaic introgression signals

Table 1
Immune system-related introgressed genes.

POSITION	GENE	UP_KW_BIOLOGICAL_PROCESS
Chr2:71032316–71,079,245	*CD207*	GO:0044419 ~ interspecies interaction between organisms, GO:0051607 ~ defence response to virus. Adaptive immune system.
Chr2:79234333–79,280,710	*REG3G*	KW-0011 ~ Acute phase, KW-0395 ~ Inflammatory response, Innate immune system pathway.
Chr4:87217607–87,266,456	*MAPK10*	KW-0090 ~ Biological rhythms; Innate immune system pathway.
Chr6: 29798610–29,897,944	*HLA-G; HLA-H*	KW-0391 ~ Immunity.
Chr6:143867336–143,897,977	*PHACTR2*	Innate immune system pathway.
Chr6:29698426–29,746,527	*HLA-F*	KW-0391 ~ Immunity.
Chr6:29898105–29,947,740	*HLA-A*	KW-0391 ~ Immunity, KW-0399 ~ Innate immunity, KW-0945 ~ Host–virus interaction, KW-1064 ~ Adaptive immunity.
Chr6:31198205–31,348,022	*HLA-B; HLA-C*	KW-0391 ~ Immunity, KW-0399 ~ Innate immunity, KW-0945 ~ Host–virus interaction, KW-1064 ~ Adaptive immunity.
Chr6:32598042–32,644,388	*HLA-DQA1; HLA-DQB1*	KW-0391 ~ Immunity, KW-1064 ~ Adaptive immunity. Innate immune system pathway.
Chr6:32698044–32,748,039	*HLA-DQA2; HLA-DQB2*	KW-0391 ~ Immunity, KW-1064 ~ Adaptive immunity.
Chr6:32748045–32,797,488	*HLA-DOB*	KW-0391 ~ Immunity, KW-1064 ~ Adaptive immunity.
Chr6:32748045–32,797,488	*TAP2*	KW-0391 ~ Immunity, KW-0571 ~ Peptide transport, KW-0653 ~ Protein transport, KW-0813 ~ Transport, KW-0945 ~ Host–virus interaction, KW-1064 ~ Adaptive immunity.
Chr6:52050353–52,098,036	*IL17A; IL17F*	KW-0391 ~ Immunity, KW-0395 ~ Inflammatory response, KW-0399 ~ Innate immunity, KW-1064 ~ Adaptive immunity.
Chr8:13155109–13,201,971	*DLC1*	Involved in innate immune responses against DNA virus infection. Adaptive and innate immune system.
Chr11:102454640–102,503,213	*MMP20*	Innate immune system pathway.
Chr12:10229931–10,275,336	*CLEC7A*	KW-0391 ~ Immunity, KW-0395 ~ Inflammatory response, KW-0399 ~ Innate immunity.
Chr16:68760463–68,805,641	*CDH1*	KW-0130 ~ Cell adhesion; regulation of immune response.
Chr22:35654133–35,703,551	*TOM1*	KW-0653 ~ Protein transport, KW-0813 ~ Transport. Innate immune system pathway.

through positive selection. There is evidence that adaptive segments of Neanderthal ancestry in modern humans are enriched for proteins that interact with viruses (Enard and Petrov, 2018). It is hypothesized that due to positive selection against the exchanged viruses between Neanderthals and modern humans, these introgressed archaic alleles can be employed to assess the level of proportional protective evolution against the pathogen. (Enard and Petrov, 2018). Therefore, adaptive introgressed alleles should include immune genes that clearly function against pathogens, including viruses (Abi-Rached et al., 2011; Deschamps et al., 2016; Nedelec et al., 2016). Here, using whole-genome sequence data, we detected introgressed Neanderthal segments in Lithuanians and positive selection candidate genetic regions, in addition to immune-related genes and their interactions with viruses.

Genomic regions introgressed from Neanderthals are enriched in genes related to immunity, keratinocyte differentiation, and sensory perception, consistent with previously published findings (Deschamps et al., 2016; Gittelman et al., 2016; Racimo et al., 2015; Vernot and Akey, 2014). These genes include previously identified candidates in Europeans as *HLA* regions at the same locus 6p21.33 (Abi-Rached et al., 2011), and *IL17F* (Sankararaman et al., 2014; Deschamps et al., 2016). The largest number of introgressed Neanderthal fragments identified in Lithuanian WGS data were located on chromosome 6, where *HLA* genes (*HLA-A, HLA—B, HLA—C, HLA-DRB1,* and *HLA-DQA1*) involved in innate and adaptive immunity are located (Abi-Rached et al., 2011). *HLA* genes are associated with susceptibility and resistance to infectious diseases (Cagliani and Sironi, 2013). HLA class I archaic alleles comprise at least 50–85% of the HLA variants in Eurasian populations because of pathogen-driven selection and past interbreeding with archaic hominins (Abi-Rached et al., 2011). Archaic HLA variants probably provided modern humans with preadapted HLA variants that helped them survive in a new environment. Population-specific HLA-associated selective pressure might also shape the evolution not only of HIV-1 associated with HLA genes but also other pathogens that emerged recently, such as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Kist et al., 2020). Using iHS statistics, we identified a candidate positively selected region encompassing the *HLA-C* gene. Nonsynonymous *HLA-C* variation detected in rs1130838 has been associated with psoriasis and Behcet's disease (Lee et al., 2012). Psoriasis is a common multifactorial chronic skin disease with a genetic background that was prompted investigations in the early 1970s into associations with the HLA complex on chromosome 6p (Russell et al., 1972). The global psoriasis prevalence rate is approximately 2–3% among the world's population (Sewerin et al., 2019), reaching 8–11% in some Northern European countries (Egeberg et al., 2019) and ~ 4% in Lithuania. Strong associations with *HLA-Cw6* and *HLA-DR7* were identified in the Finnish population and in Germany. However, only ~10% of HLA-Cw6-positive individuals have developed psoriasis, suggesting a major role for additional genes and/or environmental triggers or certain protective mechanism (Bhalerao and Bowcock, 1998). According to the TreeWAS database, the identified *HLA-C* variant rs1130838 has a protective effect on the psoriasis phenotype (Cortes et al., 2020). Further research is needed to investigate the protective effect of this variant on Lithuanians. Psoriasis is a damaging autoimmune disorder, but in a pathogen-rich past, a highly active immune system could potentially have served a protective function, even if it increased the chance of an autoimmune response. Genes evolving through positive or balancing selection, such as *DARC*, *HLA* locus genes, and *ABO* blood group genes, are usually more permissive of functional variation, which may exert a protective effect against infections (Key et al., 2014).

Consequently, our subsequent inquiry delved into the potential interactions between these immune-related genes that have undergone positive selection and various viruses. Two interactions were detected, between the *HLA-C* gene and HIV1 and *PARP8* and the *EMB* and SARS-CoV-2. In the Chinese population, *HLA-C* rs1130838-T allele was found to be associated with a significantly elevated risk of hepatitis C virus infection, and was suggested as potential biomarker (Shen et al., 2021). The frequency in Lithuanians of rs1130838-T is 0.20, and in Chinese rs1130838-T 0.19. The *HLA-C* can also influence the risk of Human immunodeficiency virus disease (Cortes et al., 2020). Genome-wide association studies have also revealed the special importance of the *HLA-C* gene in early or establishing stages of HIV-1 infection. It was shown that a SNP located 35 kb upstream of the coding region of the *HLA-C* gene (-35C/T, rs9264942) is a major determinant accounting for the viral load set point in Europeans. The protective allele —35C may be important for the viral control in natural history of infection, involving both lower viremias and lower cellular HIV-1 reservoirs (Herráiz-Nicuesa et al., 2017). The C allele frequency of the SNP rs9264942 in Lithuanians is 0.65, in Europeans 0.37, and in Africans 0.30. According to UNAIDS estimations, the HIV incidence rate in Lithuania decreased from 5.4 to 4.3 cases per 100,000 individuals from 2019 to 2021 (WHO Regional Office for Europe, 2022). In Europe, it was 12 cases per 100,000 individuals.

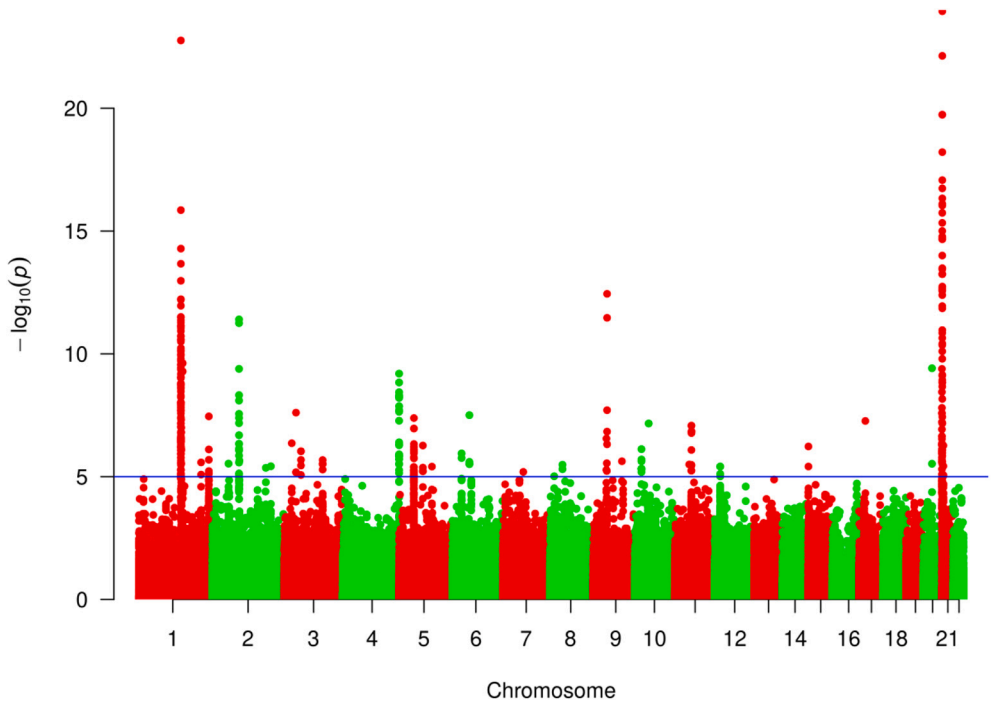


Fig. 4. Genome-wide distribution of selection signatures. -Log10(p value) for the integrated haplotype score (iHS) plotted against chromosome position, with the significance threshold highlighted with a dashed line ($P < 0.001$).

The next identified positively selected immune-related genes and virus combination was *PARP8* and the *EMB* with SARS-CoV-2. Poly-adenosine diphosphate-ribose polymerases (PARPs) promote ADP-ribosylation, a highly conserved, fundamental posttranslational modification (PTM) (Gupte et al., 2017). The EMB (embigin) gene encodes a transmembrane glycoprotein that is a member of the immunoglobulin superfamily. Embigin, like besigin, may be a second high-affinity receptor for the spike protein of SARS-CoV-2 (Fuentes-Prior, 2021). It is known that the major PARylating enzyme's role is to target viral macrodomains and have got the capacity to restrict viral replication following ADP-mediated ribosylation modification of the viral macro domains (Grunewald et al., 2019). A high-fat diet was able to activate PARP and PAR formation in mice as the major fatty acid components was able to elevate the expression of PARP in hypothalamic neurons (Scheibye-Knudsen et al., 2014). Based on Urnikyte et al., 2019 data, we know that the genome of Lithuanians is characterized by genes *PPARD* and *PNLIP* under positive selection that improve fat metabolism. Hence, this research enhances our current comprehension of immune-related signals within the *PARP8* and *EMB* genes, emphasizing their positive selection and proposing a specific protective mechanism against SARS-CoV-2 infection among Lithuanians.

New genome studies point to the importance of balancing selection in human evolution (Aqil et al., 2023; Bitarello et al., 2023). Balancing selection is a mode of adaptation that maintains advantageous genetic diversity in populations. Genes of immune function are prime candidates for long term balancing selection (Meyer and Thomson, 2001; Spurgin and Richardson, 2010; Bitarello et al., 2018). Aqil et al. (2023), have shown that a certain proportion of human ancient polymorphisms that are older than the anatomically modern human - archaic split (~700,000 years), have been maintained by balancing selection. Many of these ancient genes regulate the immune response and may protect against infectious diseases outbreaks and starvation, which have occurred periodically throughout human history (Aqil et al., 2023). We considered identified genes under balancing selection reported in Andrés et al. (2009), Leffler et al. (2013), DeGiorgio et al. (2014), B. Bitarello et al. (2018), and Aqil et al. (2023). We identify positively selected genes, *C6orf15, CDSN, PSORS1C1, CCHCR1, PSORS1C3, HCG27;HLA-C, TCF19, POU5F1*, detected by DeGiorgio et al. (2014), and *CDSN* detected by Andrés et al. (2009). Also a *HLA-C* gene identified by Bitarello et al. (2018). Between introgressed genes *HLA-B* was detected by Andrés et al. (2009), and *HLA-A, HLA—B; HLA—C, HLA—F, HLA-G; HLA—H, HLA-DQA1; HLA-DQB1, HLA-DQA2; HLA-DQB2, HLA-DOB*, and *TAP2* by DeGiorgio et al. (2014), and Bitarello et al. (2018), and *HLA-DQ* genes by Aqil et al. (2023). We identified genes with previous evidence of balancing selection, however, further investigations are needed to explain how much genetic variation is maintained by balancing selection and how it shapes the evolution.

The identified positively selected nonsynonymous variant rs12488457 in *COL6A5,* previously described by Urnikyte et al. (2019), has an important impact on protein structure that can increase or decrease protein stability and may represent a risk factor or molecular markers for human disorders such as lung, colorectal, prostate, thyroid or breast cancer (Kalmari et al., 2022). Another study showed that rs12488457 (A/C, *COL6A5*) was significantly associated with psoriasis and psoriatic arthritis in Italian patients (Caputo et al., 2020). This variant has also been associated with eczema (Naumann et al., 2011). Both nonsynonymous variants (rs12488457 and rs16827497) linked to selection signals identified in *COL6A5* were found to be associated with bone mineral density (Wang et al., 2016). These results once again demonstrate the importance of the *COL6A5* gene in disease etiopathogenesis in the Lithuanian population.

Neverless, the functional roles of the introgressed regulatory variants require further investigation, but our results clearly establish that archaic admixture, whether adaptive or not, has increased the immune diversity of modern Europeans.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2023.105528.

Ethics approval and consent to participate

This work was approved by the Vilnius Regional Research Ethics Committee (No. 2020/6–1243-724, 22-06-2022). All the study participants provided written informed consent.

Table 2
Detected candidate genetic regions under selection with iHS statistics.

POSITION	N_MRK*	N_EXTR_MRK*	GENES IN THE REGION
			NBPF8;NBPF9; FAM72C; NBPF20; PDE4DIPP2; NBPF19;NBPF20;
chr1:141600000–150,200,000	1698	75	*NBPF9*
chr1:236000000–237,900,000	1538	27	*HEATR1;ACTN2 EML6;RTN4; MTIF2; PRORSD1P; CCDC85A;VRK2;*
chr2:54100000–58,000,000	2192	19	*CFAP36*
chr2:89500000–91,400,000	107	45	*COL6A4P2; COL6A5;*
chr3:129200000–131,100,000	792	9	*COL6A6*
chr4:189700000–191,600,000	1091	28	*LOC105379514; FRG1-DT EMB;PARP8; LINC02106;*
chr5:48900000–51,300,000	922	51	*LOC642366 HCG22;C6orf15; CDSN; PSORS1C1; CCHCR1;TCF19; POU5F1; PSORS1C3; HCG27;*
chr6:30200000–32,100,000	776	16	*HLA-C*
chr6:62900000–64,800,000	1075	24	*KHDRBS2;LGSN*
chr9:16600000–16,900,000	91	8	*BNC2; CNTLN LOC100499489;*
chr10:21900000–23,800,000	407	11	*PIP4K2A LINC02644; ZNF438;ZEB1;*
chr10:31000000–32,900,000	911	7	*ARHGAP12 OR5W2;OR5I1; OR10AG1; OR7E5P;OR8H1; OR8K3;OR8J1; OR8U1;OR5AL1; OR5M8; OR5M11; OR5M10; OR5M1;OR5AR1; OR9G1; OR5AK4P; LRRC55;LRRC55;*
chr11:55200000–57,300,000	652	24	*APLNR CAPZA3;*
chr12:18200000–20,100,000	1052	23	*PLEKHA5 CHEK2P2;*
chr15:20400000–22,300,000	637	7	*HERC2P3 MIR10396A; TEKT4P2;TPTE;*
chr21:9000000–12,100,000	703	111	*BAGE,BAGE2 ANKRD30BP2; MIR3156–3;*
chr21:13800000–15,700,000	354	29	*LINC01674; LIPI*

* N_MRK: Number of markers; N_EXTR_MRK: Number of extreme markers. Novel regions are in bold.

CRediT authorship contribution statement

Alina Urnikyte: Conceptualization, Funding acquisition, Formal analysis, Writing - original draft, Writing - review & editing. **Abigaile Masiulyte:** Formal analysis, Visualization, Writing - original draft. **Laura Pranckeniene:** Formal analysis, Writing - review & editing. **Vaidutis Kučinskas:** Conceptualization, Writing - review & editing.

Funding

The research has received funding from European Social Fund (project No 09.3.3-LMT-K-712-23-0104) under grant agreement with the Research Council of Lithuania (LMTLT).

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used Springer Nature author service in order to edit English language. After using this service, the authors reviewed and edited the content as needed and takes full responsibility for the content of the publication.

Declaration of Competing Interest

None.

Data availability

The datasets analysed during the current study are available in the Mendeley repository, DOI:10.17632/d2xt5hdm5j.2.

References

Abi-Rached, L., Jobin, M.J., Kulkarni, S., McWhinnie, A., Dalva, K., Gragert, L., Babrzadeh, F., Gharizadeh, B., Luo, M., Plummer, F.A., Kimani, J., Carrington, M., Middleton, D., Rajalingam, R., Beksac, M., Marsh, S.G., Maiers, M., Guethlein, L.A., Tavoularis, S., Little, A.M., Green, R.E., Norman, P.J., Parham, P., 2011. The shaping of modern human immune systems by multiregional admixture with archaic humans. Science 334, 89–94. https://doi.org/10.1126/science.1209202.

Andrés, A.M., Hubisz, M.J., Indap, A., Torgerson, D.G., Degenhardt, J.D., Boyko, A.R., Gutenkunst, R.N., White, T.J., Green, E.D., Bustamante, C.D., Clark, A.G., Nielsen, R., 2009. Targets of balancing selection in the human genome. Mol. Biol. Evol. 12, 2755–2764. https://doi.org/10.1093/molbev/msp190.

Aqil, A., Speidel, L., Pavlidis, P., Gokcumen, O., 2023. Balancing selection on genomic deletion polymorphisms in humans. eLife 12, e79111. https://doi.org/10.7554/eLife.79111.

Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., Abecasis, G.R., 2015. A global reference for human genetic variation. Nature 526, 68–74. https://doi.org/10.1038/nature15393.

Barreiro, L.B., Quintana-Murci, L., 2010. From evolutionary genetics to human immunology: how selection shapes host defence genes. Nat. Rev. Genet. 11, 17–30. https://doi.org/10.1038/nrg2698.

Barreiro, L.B., Laval, G., Quach, H., Patin, E., Quintana-Murci, L., 2008. Natural selection has driven population differentiation in modern humans. Nat. Genet. 40, 340–345. https://doi.org/10.1038/ng.78.

Bhalerao, J., Bowcock, A.M., 1998. The Genetics of Psoriasis: a complex Disorder of the Skin and Immune System. Hum. Mol. Genet. 7 (10), 1537–1545. https://doi.org/10.1093/hmg/7.10.1537.

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.H., Pages, F., Trajanoski, Z., Galon, J., 2009. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics 25, 1091–1093. https://doi.org/10.1093/bioinformatics/btp101.

Bitarello, B.D., de Filippo, C., Teixeira, J.C., Schmidt, J.M., Kleinert, P., Meyer, D., Andrés, A.M., 2018. Signatures of Long-Term Balancing selection in Human Genomes. Genome Biol. Evol. 10 (3), 939–955. https://doi.org/10.1093/gbe/evy054.

Bitarello, B.D., Brandt, D.Y.C., Meyer, D., Andrés, A.M., 2023. Inferring Balancing selection from Genome-Scale Data. Genome Biol. Evol. 15 (3) https://doi.org/10.1093/gbe/evad032.

Brenner, L.N., Mercader, J.M., Robertson, C.C., Cole, J., Chen, L., Jacobs, S.B.R., Rich, S. S., Florez, J.C., 2020. Analysis of glucocorticoid-related genes reveal CCHCR1 as a new candidate gene for type 2 diabetes. J. Endocr. Soc. 4, bvaa121. https://doi.org/10.1210/jendso/bvaa121.

Breuer, K., Foroushani, A.K., Laird, M.R., Chen, C., Sribnaia, A., Lo, R., Winsor, G.L., Hancock, R.E., Brinkman, F.S., Lynn, D.J., 2013. InnateDB: systems biology of innate immunity and beyond–recent updates and continuing curation. Nucleic Acids Res. 41, D1228–D1233. https://doi.org/10.1093/nar/gks1147.

Cagliani, R., Sironi, M., 2013. Pathogen-driven selection in the human genome. Int. J. Evol. Biol. 2013, 204240 https://doi.org/10.1155/2013/204240.

Caputo, V., Strafella, C., Termine, A., Campione, E., Bianchi, L., Novelli, G., Giardina, E., Cascella, R., 2020. RNAseq-based prioritization revealed COL6A5, COL8A1, COL10A1 and MIR146A as common and differential susceptibility biomarkers for psoriasis and psoriatic arthritis: confirmation from genotyping analysis of 1417 Italian ubjects. Int. J. Mol. Sci. 21, 2740. https://doi.org/10.3390/ijms21082740.

Chatr-Aryamontri, A., Breitkreutz, B.J., Heinicke, S., Boucher, L., Winter, A., Stark, C., Nixon, J., Ramage, L., Kolas, N., O'Donnell, L., Reguly, T., Breitkreutz, A., Sellam, A., Chen, D., Chang, Rust C., Livstone, J., Oughtred, M., Dolinski, R., Tyers, K.M., 2013. The BioGRID interaction database: 2013 update. Nucleic Acids Res. 41, D816–D823. https://doi.org/10.1093/nar/gks1158.

Cheung, Y.H., Watkinson, J., Anastassiou, D., 2011. Conditional meta-analysis stratifying on detailed HLA genotypes identifies a novel type 1 diabetes locus around TCF19 in the MHC. Hum. Genet. 129, 161–176. https://doi.org/10.1007/s00439-010-0908-2.

Collen, E.J., Johar, A.S., Teixeira, J.C., Llamas, B., 2022. The immunogenetic impact of European colonization in the Americas. Front. Genet. 13, 918227 https://doi.org/10.3389/fgene.2022.918227.

Corona, E., Wang, L., Ko, D., Patel, C.J., 2018. Systematic detection of positive selection in the human-pathogen interactome and lasting effects on infectious disease susceptibility. PLoS One 13, e0196676. https://doi.org/10.1371/journal.pone.0196676.

Cortes, A., Albers, P.K., Dendrou, C.A., Fugger, L., McVean, G., 2020. Identifying cross-disease components of genetic risk across hospital data in the UK Biobank. Nat. Genet. 52, 126–134. https://doi.org/10.1038/s41588-019-0550-4.

Dannemann, M., Kelso, J., 2017. The contribution of Neanderthals to phenotypic variation in modern humans. Am. J. Hum. Genet. 101, 578–589. https://doi.org/10.1016/j.ajhg.2017.09.010.

DeGiorgio, M., Lohmueller, K.E., Nielsen, R., 2014. A model-based approach for identifying signatures of ancient balancing selection in genetic data. PLoS Genet. 10 (8), e1004561 https://doi.org/10.1371/journal.pgen.1004561.

Delaneau, O., Zagury, J.F., Robinson, M.R., Marchini, J.L., Dermitzakis, E.T., 2019. Accurate, scalable and integrative haplotype estimation. Nat. Commun. 10, 5436. https://doi.org/10.1038/s41467-019-13225-y.

Deschamps, M., Laval, G., Fagny, M., Itan, Y., Abel, L., Casanova, J.L., Patin, E., Quintana-Murci, L., 2016. Genomic signatures of selective pressures and introgression from archaic hominins at human innate immunity genes. Am. J. Hum. Genet. 98, 5–21. https://doi.org/10.1016/j.ajhg.2015.11.014.

Dolgova, O., Lao, O., 2018. Evolutionary and medical consequences of archaic introgression into modern human genomes. Genes (Basel) 9, 358. https://doi.org/10.3390/genes9070358.

Durvasula, A., Sankararaman, S., 2019. A statistical model for reference-free inference of archaic local ancestry. PLoS Genet. 15, e1008175 https://doi.org/10.1371/journal.pgen.1008175.

Egeberg, A., Andersen, Y.M.F., Thyssen, J.P., 2019. Prevalence and characteristics of psoriasis in Denmark: findings from the Danish skin cohort. BMJ Open 9, e028116. https://doi.org/10.1136/bmjopen-2018-028116.

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

Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., Milacic, M., Roca, C.D., Rothfels, K., Sevilla, C., Shamovsky, V., Shorser, S., Varusai, T., Viteri, G., Weiser, J., Wu, G., Stein, L., Hermjakob, H., D'Eustachio, P., 2018. The reactome pathway knowledgebase. Nucleic Acids Res. 46, D649–D655. https://doi.org/10.1093/nar/gkx1132.

Fuentes-Prior, P., 2021. Priming of SARS-CoV-2 S protein by several membrane-bound serine proteinases could explain enhanced viral infectivity and systemic COVID-19 infection. J. Biol. Chem. 296, 100135 https://doi.org/10.1074/jbc.REV120.015980.

Fumagalli, M., Sironi, M., Pozzoli, U., Ferrer-Admetlla, A., Pattini, L., Nielsen, R., 2011. Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. PLoS Genet. 7, e1002355 https://doi.org/10.1371/journal.pgen.1002355.

Gautier, M., Vitalis, R., 2012. Rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. Bioinformatics 28, 1176–1177. https://doi.org/10.1093/bioinformatics/bts115.

Gittelman, R.M., Schraiber, J.G., Vernot, B., Mikacenic, C., Wurfel, M.M., Akey, J.M., 2016. Archaic hominin admixture facilitated adaptation to out-of-africa environments. Curr. Biol. 26, 3375–3382. https://doi.org/10.1016/j.cub.2016.10.041.

Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H., Hansen, N.F., Durand, E.Y., Malaspinas, A.S., Jensen, J. D., Marques-Bonet, T., Alkan, C., Prufer, K., Meyer, M., Burbano, H.A., Good, J.M., Schultz, R., Aximu-Petri, A., Butthof, A., Hober, B., Hoffner, B., Siegemund, M., Weihmann, A., Nusbaum, C., Lander, E.S., Russ, C., Novod, N., Affourtit, J., Egholm, M., Verna, C., Rudan, P., Brajkovic, D., Kucan, Z., Gusic, I., Doronichev, V. B., Golovanova, L.V., Lalueza-Fox, C., de la Rasilla, M., Fortea, J., Rosas, A., Schmitz, R.W., Johnson, P.L.F., Eichler, E.E., Falush, D., Birney, E., Mullikin, J.C., Slatkin, M., Nielsen, R., Kelso, J., Lachmann, M., Reich, D., Paabo, S., 2010. A draft sequence of the Neandertal genome. Science 328, 710–722. https://doi.org/10.1126/science.1188021.

Grossman, S.R., Andersen, K.G., Shlyakhter, I., Tabrizi, S., Winnicki, S., Yen, A., Park, D. J., Griesemer, D., Karlsson, E.K., Wong, S.H., Cabili, M., Adegbola, R.A., Bamezai, R. N., Hill, A.V., Vannberg, F.O., Rinn, J.L., Genomes, P., Lander, E.S., Schaffner, S.F., Sabeti, P.C., 2013. Identifying recent adaptations in large-scale genomic data. Cell 152, 703–713. https://doi.org/10.1016/j.cell.2013.01.035.

Grunewald, M.E., Chen, Y., Kuny, C., Maejima, T., Lease, R., Ferraris, D., Aikawa, M., Sullivan, C.S., Perlman, S., Fehr, A.R., 2019. The coronavirus macrodomain is required to prevent PARP-mediated inhibition of virus replication and enhancement of IFN expression. PLoS Pathog. 15, e1007756. https://doi.org/10.1371/journal.ppat.1007756.

Gupte, R., Liu, Z., Kraus, W.L., 2017. PARPs and ADP-ribosylation: recent advances linking molecular functions to biological outcomes. Genes Dev. 31, 101–126. https://doi.org/10.1101/gad.291518.116.

Hancock, A.M., Witonsky, D.B., Gordon, A.S., Eshel, G., Pritchard, J.K., Coop, G., Di Rienzo, A., 2008. Adaptations to climate in candidate genes for common metabolic disorders. PLoS Genet. 4 (2), e32 https://doi.org/10.1371/journal.pgen.0040032.

Herráiz-Nicuesa, L., Hernández-Flórez, D.C., Valor, L., García-Consuegra, S., Navarro-Valdivieso, J.P., Fernández-Cruz, E., Rodríguez-Sainz, C., 2017. Impact of the Polymorphism rs9264942 near the HLA-C Gene on HIV-1 DNA Reservoirs in Asymptomatic Chronically Infected patients Initiating Antiviral Therapy. J Immunol Res 2017, 8689313. https://doi.org/10.1155/2017/8689313.

Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4, 44–57. https://doi.org/10.1038/nprot.2008.211.

Huyghe, J.R., Jackson, A.U., Fogarty, M.P., Buchkovich, M.L., Stancakova, A., Stringham, H.M., Sim, X., Yang, L., Fuchsberger, C., Cederberg, H., Chines, P.S., Teslovich, T.M., Romm, J.M., Ling, H., McMullen, I., Ingersoll, R., Pugh, E.W., Doheny, K.F., Neale, B.M., Daly, M.J., Kuusisto, J., Scott, L.J., Kang, H.M., Collins, F. S., Abecasis, G.R., Watanabe, R.M., Boehnke, M., Laakso, M., Mohlke, K.L., 2013. Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. Nat. Genet. 45, 197–201. https://doi.org/10.1038/ng.2507.

Hystad, M.E., Myklebust, J.H., Bo, T.H., Sivertsen, E.A., Rian, E., Forfang, L., Munthe, E., Rosenwald, A., Chiorazzi, M., Jonassen, I., Staudt, L.M., Smeland, E.B., 2007. Characterization of early stages of human B cell development by gene expression profiling. J. Immunol. 179, 3662–3671. https://doi.org/10.4049/jimmunol.179.6.3662.

Jacobs, L.C., Wollstein, A., Lao, O., Hofman, A., Klaver, C.C., Uitterlinden, A.G., Nijsten, T., Kayser, M., Liu, F., 2013. Comprehensive candidate gene study highlights UGT1A and BNC2 as new genes determining continuous skin color variation in Europeans. Hum. Genet. 132, 147–158. https://doi.org/10.1007/s00439-012-1232-9.

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

Kalmari, A., Heydari, M., Colagar, A.H., Arash, V., 2022. In Silico analysis of collagens missense SNPs and human abnormalities. Biochem. Genet. 60, 1630–1656. https://doi.org/10.1007/s10528-021-10172-6.

Karlsson, E.K., Kwiatkowski, D.P., Sabeti, P.C., 2014. Natural selection and infectious disease in human populations. Nat. Rev. Genet. 15, 379–393. https://doi.org/10.1038/nrg3734.

Kerner, G., Neehus, A.L., Philippot, Q., Bohlen, J., Rinchai, D., Kerrouche, N., Puel, A., Zhang, S.Y., Boisson-Dupuis, S., Abel, L., Casanova, J.L., Patin, E., Laval, G., Quintana-Murci, L., 2023. Genetic adaptation to pathogens and increased risk of inflammatory disorders in post-Neolithic Europe. Cell Genom. 3, 100248 https://doi.org/10.1016/j.xgen.2022.100248.

Key, F.M., Teixeira, J.C., de Filippo, C., Andrés, A.M., 2014. Advantageous diversity maintained by balancing selection in humans. Curr. Opin. Genet. Dev. 29, 45–51. https://doi.org/10.1016/j.gde.2014.08.001.

Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., Shendure, J., 2014. A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. 46, 310–315. https://doi.org/10.1038/ng.2892.

Kist, N.C., Lambert, B., Campbell, S., Katzourakis, A., Lunn, D., Lemey, P., Iversen, A.K. N., 2020. HIV-1 p24Gag adaptation to modern and archaic HLA-allele frequency differences in ethnic groups contributes to viral subtype diversification. Virus Evol. 6 https://doi.org/10.1093/ve/veaa085 veaa085.

Le, T.D., Nguyen, P.D., Korkin, D., Thieu, T., 2022. PHILM2Web: a high-throughput database of macromolecular host-pathogen interactions on the Web. Database (Oxford) 2022. https://doi.org/10.1093/database/baac042 baac042.

Lee, Y.H., Choi, S.J., Ji, J.D., Song, G.G., 2012. Genome-wide pathway analysis of a genome-wide association study on psoriasis and Behcet's disease. Mol. Biol. Rep. 39, 5953–5959. https://doi.org/10.1007/s11033-011-1407-9.

Leffler, E.M., Gao, Z., Pfeifer, S., Ségurel, L., Auton, A., Venn, O., Bowden, R., Bontrop, R., Wall, J.D., Sella, G., Donnelly, P., McVean, G., Przeworski, M., 2013. Multiple instances of ancient balancing selection shared between humans and chimpanzees. Science 339 (6127), 1578–1582. https://doi.org/10.1126/science.1234070.

Maggadottir, S.M., Li, J., Glessner, J.T., Li, Y.R., Wei, Z., Chang, X., Mentch, F.D., Thomas, K.A., Kim, C.E., Zhao, Y., Hou, C., Wang, F., Jorgensen, S.F., Perez, E.E., Sullivan, K.E., Orange, J.S., Karlsen, T.H., Chapel, H., Cunningham-Rundles, C., Hakonarson, H., 2015. Rare variants at 16p11.2 are associated with common variable immunodeficiency. J. Allergy Clin. Immunol. 135, 1569–1577. https://doi.org/10.1016/j.jaci.2014.12.1939.

Mangan, R.J., Alsina, F.C., Mosti, F., Sotelo-Fonseca, J.E., Snellings, D.A., Au, E.H., Carvalho, J., Sathyan, L., Johnson, G.D., Reddy, T.E., Silver, D.L., Lowe, C.B., 2022. Adaptive sequence divergence forged new neurodevelopmental enhancers in humans. Cell 185, 4587–4603.e23. https://doi.org/10.1016/j.cell.2022.10.016.

Meyer, D., Thomson, G., 2001. How selection shapes variation of the human major histocompatibility complex: a review. Ann. Hum. Genet. 65, 1–26. https://doi.https://doi.org/10.1046/j.1469-1809.2001.6510001.x.

Meyer, M., Kircher, M., Gansauge, M.T., Li, H., Racimo, F., Mallick, S., Schraiber, J.G., Jay, F., Prufer, K., de Filippo, C., Sudmant, P.H., Alkan, C., Fu, Q., Do, R., Rohland, N., Tandon, A., Siebauer, M., Green, R.E., Bryc, K., Briggs, A.W., Stenzel, U., Dabney, J., Shendure, J., Kitzman, J., Hammer, M.F., Shunkov, M.V., Derevianko, A.P., Patterson, N., Andres, A.M., Eichler, E.E., Slatkin, M., Reich, D., Kelso, J., Paabo, S., 2012. A high-coverage genome sequence from an archaic Denisovan individual. Science 338, 222–226. https://doi.org/10.1126/science.1224344.

Murga-Moreno, J., Coronado-Zamora, M., Bodelon, A., Barbadilla, A., Casillas, S., 2019. PopHumanScan: the online catalog of human genome adaptation. Nucleic Acids Res. 47, D1080–D1089. https://doi.org/10.1093/nar/gky959.

Naumann, A., Soderhall, C., Folster-Holst, R., Baurecht, H., Harde, V., Muller-Wehling, K., Rodriguez, E., Ruether, A., Franke, A., Wagenpfeil, S., Novak, N., Mempel, M., Kalali, B.N., Allgaeuer, M., Koch, J., Gerhard, M., Melen, E., Wahlgren, C.F., Kull, I., Stahl, C., Pershagen, G., Lauener, R., Riedler, J., Doekes, G., Scheynius, A., Illig, T., von Mutius, E., Schreiber, S., Kere, J., Kabesch, M., Weidinger, S., 2011. A comprehensive analysis of the COL29A1 gene does not support a role in eczema. J. Allergy Clin. Immunol. 127, 1187–1194.e7. https://doi.org/10.1016/j.jaci.2010.12.1123.

Nedelec, Y., Sanz, J., Baharian, G., Szpiech, Z.A., Pacis, A., Dumaine, A., Grenier, J.C., Freiman, A., Sams, A.J., Hebert, S., Sabourin, A.P., Luca, F., Blekhman, R., Hernandez, R.D., Pique-Regi, R., Tung, J., Yotova, V., Barreiro, L.B., 2016. Genetic ancestry and natural selection drive population differences in immune responses to pathogens. Cell 167, 657–669.e21. https://doi.org/10.1016/j.cell.2016.09.025.

Orange, J.S., Glessner, J.T., Resnick, E., Sullivan, K.E., Lucas, M., Ferry, B., Kim, C.E., Hou, C., Wang, F., Chiavacci, R., Kugathasan, S., Sleasman, J.W., Baldassano, R., Perez, E.E., Chapel, H., Cunningham-Rundles, C., Hakonarson, H., 2011. Genome-wide association identifies diverse causes of common variable immunodeficiency. J. Allergy Clin. Immunol. 127, 1360–1367.e6. https://doi.org/10.1016/j.jaci.2011.02.039.

Owers, K.A., Sjodin, P., Schlebusch, C.M., Skoglund, P., Soodyall, H., Jakobsson, M., 2017. Adaptation to infectious disease exposure in indigenous Southern African populations. Proc. Biol. Sci. 284, 20170226. https://doi.org/10.1098/rspb.2017.0226.

Prufer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., Renaud, G., Sudmant, P.H., 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, J.C., Vohr, S.H., Green, R.E., Hellmann, I., Johnson, P.L., Blanche, H., Cann, H., Kitzman, J.O., Shendure, J., Eichler, E.E., Lein, E.S., Bakken, T.E., Golovanova, L.V., Doronichev, V. B., Shunkov, M.V., Derevianko, A.P., Viola, B., Slatkin, M., Reich, D., Kelso, J., Paabo, S., 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature 505, 43–49. https://doi.org/10.1038/nature12886.

Prufer, K., de Filippo, C., Grote, S., Mafessoni, F., Korlevic, P., Hajdinjak, M., Vernot, B., Skov, L., Hsieh, P., Peyregne, S., Reher, D., Hopfe, C., Nagel, S., Maricic, T., Fu, Q., Theunert, C., Rogers, R., Skoglund, P., Chintalapati, M., Dannemann, M., Nelson, B. J., Key, F.M., Rudan, P., Kucan, Z., Gusic, I., Golovanova, L.V., Doronichev, V.B., Patterson, N., Reich, D., Eichler, E.E., Slatkin, M., Schierup, M.H., Andres, A.M., Kelso, J., Meyer, M., Paabo, S., 2017. A high-coverage Neandertal genome from Vindija Cave in Croatia. Science 358, 655–658. https://doi.org/10.1126/science.aao1887.

Quintana-Murci, L., 2019. Human immunology through the lens of evolutionary genetics. Cell 177, 184–199. https://doi.org/10.1016/j.cell.2019.02.033.

Quintana-Murci, L., Clark, A.G., 2013. Population genetic tools for dissecting innate immunity in humans. Nat. Rev. Immunol. 13, 280–293. https://doi.org/10.1038/nri3421.

Racimo, F., Sankararaman, S., Nielsen, R., Huerta-Sanchez, E., 2015. Evidence for archaic adaptive introgression in humans. Nat. Rev. Genet. 16, 359–371. https://doi.org/10.1038/nrg3936.

Reich, D., Green, R.E., Kircher, M., Krause, J., Patterson, N., Durand, E.Y., Viola, B., Briggs, A.W., Stenzel, U., Johnson, P.L., Maricic, T., Good, J.M., Marques-Bonet, T., Alkan, C., Fu, Q., Mallick, S., Li, H., Meyer, M., Eichler, E.E., Stoneking, M., Richards, M., Talamo, S., Shunkov, M.V., Derevianko, A.P., Hublin, J.J., Kelso, J., Slatkin, M., Paabo, S., 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature 468, 1053–1060. https://doi.org/10.1038/nature09710.

Ruault, M., Ventura, M., Galtier, N., Brun, M.E., Archidiacono, N., Roizes, G., De Sario, A., 2003. BAGE genes generated by juxtacentromeric reshuffling in the Hominidae lineage are under selective pressure. Genomics 81, 391–399. https://doi.org/10.1016/s0888-7543(03)00025-9.

Russell, T.J., Schultes, L.M., Kuban, D.J., 1972. Histocompatibility (HL-A) antigens associated with psoriasis. N. Engl. J. Med. 12;287 (15), 738–740. https://doi.org/10.1056/NEJM197210122871503.

Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z., Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platko, J.V., Patterson, N.J., McDonald, G.J., Ackerman, H.C., Campbell, S.J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R., Lander, E.S., 2002. Detecting recent positive selection in the human genome from haplotype structure. Nature 419, 832–837. https://doi.org/10.1038/nature01140.

Sankararaman, S., Mallick, S., Dannemann, M., Prufer, K., Kelso, J., Paabo, S., Patterson, N., Reich, D., 2014. The genomic landscape of Neanderthal ancestry in present-day humans. Nature 507, 354–357. https://doi.org/10.1038/nature12961.

Scheibye-Knudsen, M., Mitchell, S.J., Fang, E.F., Iyama, T., Ward, T., Wang, J., Dunn, C. A., Singh, N., Veith, S., Hasan-Olive, M.M., et al., 2014. A high-fat diet and NAD(+) activate Sirt1 to rescue premature aging in cockayne syndrome. Cell Metab. 20, 840–855. https://doi. https://doi.org/10.1016/j.cmet.2014.10.005.

Sewerin, P., Brinks, R., Schneider, M., Haase, I., Vordenbäumen, S., 2019. Prevalence and incidence of psoriasis and psoriatic arthritis. Ann. Rheum. Dis. 78, 286–287. https://doi.org/10.1136/annrheumdis-2018-214065.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504. https://doi.org/10.1101/gr.1239303.

Shaw, SW., Chen, CP., Cheng, PJ., et al., 2008. Gene dosage change of *TPTE* and *BAGE2* and breakpoint analysis in Robertsonian Down syndrome. J Hum Genet 53, 136–143. https://doi.org/10.1007/s10038-007-0229-z.

Shen, C., Ge, Z., Dong, C., Wang, C., Shao, J., Cai, W., Huang, P., Fan, H., Li, J., Zhang, Y., Yue, M., 2021. Genetic Variants in KIR/HLA-C Genes are Associated with the Susceptibility to HCV Infection in a High-Risk Chinese Population. Front. Immunol. 12, 632353 https://doi.org/10.3389/fimmu.2021.632353.

Sherry, S.T., Ward, M., Sirotkin, K., 1999. dbSNP—database for Single Nucleotide Polymorphisms and other classes of Minor Genetic Variation. Genome Res. 9, 677–679. https://doi.org/10.1093/nar/29.1.308.

Spracklen, C.N., Shi, J., Vadlamudi, S., Wu, Y., Zou, M., Raulerson, C.K., Davis, J.P., Zeynalzadeh, M., Jackson, K., Yuan, W., Wang, H., Shou, W., Wang, Y., Luo, J., Lange, L.A., Lange, E.M., Popkin, B.M., Gordon-Larsen, P., Du, S., Huang, W., Mohlke, K.L., 2018. Identification and functional analysis of glycemic trait loci in the China health and nutrition survey. PLoS Genet. 14, e1007275 https://doi.org/10.1371/journal.pgen.1007275.

Spurgin, L.G., Richardson, D.S., 2010. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. Proc. Biol. Sci. 277 (1684), 979–988. https://doi.org/10.1098/rspb.2009.2084.

Urnikyte, A., Flores-Bello, A., Mondal, M., Molyte, A., Comas, D., Calafell, F., Bosch, E., Kučinskas, V., 2019. Patterns of genetic structure and adaptive positive selection in the Lithuanian population from high-density SNP data. Sci. Rep. 9, 9163. https://doi.org/10.1038/s41598-019-45746-3.

Urnikyte, A., Pranckeniene, L., Domarkiene, I., Dauengauer-Kirliene, S., Molyte, A., Matuleviciene, A., Pilypiene, I., Kucinskas, V., 2022. Inherited and de novo variation in lithuanian genomes: introduction to the analysis of the generational shift. Genes (Basel) 13, 569. https://doi.org/10.3390/genes13040569.

Vasseur, E., Quintana-Murci, L., 2013. The impact of natural selection on health and disease: uses of the population genetics approach in humans. Evol. Appl. 6, 596–607. https://doi.org/10.1111/eva.12045.

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

Voight, B.F., Kudaravalli, S., Wen, X., Pritchard, J.K., 2006. A map of recent positive selection in the human genome. PLoS Biol. 4, e72 https://doi.org/10.1371/journal.pbio.0040072.

Wang, K., Li, M., Hakonarson, H., 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 38, e164 https://doi.org/10.1093/nar/gkq603.

Wang, X., Pandey, A.K., Mulligan, M.K., Williams, E.G., Mozhui, K., Li, Z., Jovaisaite, V., Quarles, L.D., Xiao, Z., Huang, J., Capra, J.A., Chen, Z., Taylor, W.L., Bastarache, L., Niu, X., Pollard, K.S., Ciobanu, D.C., Reznik, A.O., Tishkov, A.V., Zhulin, I.B., Peng, J., Nelson, S.F., Denny, J.C., Auwerx, J., Lu, L., Williams, R.W., 2016. Joint mouse-human phenome-wide association to test gene function and disease risk. Nat. Commun. 7, 10464. https://doi.org/10.1038/ncomms10464.

WHO Regional Office for Europe, 2022. European Centre for Disease Prevention and Control. HIV/AIDS Surveillance in Europe 2022–2021 data. WHO Regional Office for Europe, Copenhagen.

Yan, S.M., Sherman, R.M., Taylor, D.J., Nair, D.R., Bortvin, A.N., Schatz, M.C., McCoy, R. C., 2021. Local adaptation and archaic introgression shape global diversity at human structural variant loci. Elife 10, e67615. https://doi.org/10.7554/eLife.67615.

Yang, G.H., Fontaine, D.A., Lodh, S., Blumer, J.T., Roopra, A., Davis, D.B., 2021. TCF19 impacts a network of inflammatory and DNA damage response genes in the Pancreatic beta-Cell. Metabolites 11, 513. https://doi.org/10.3390/metabo11080513.

Zhou, Y., Browning, S.R., 2021. Protocol for detecting introgressed archaic variants with SPrime. STAR Protoc. 2, 100550 https://doi.org/10.1016/j.xpro.2021.100550.