A Neanderthal haplotype introgressed into the human genome confers protection against membranous nephropathy

OPEN

Catalin D. Voinescu[1], Monika Mozere[2], Giulio Genovese[3,4,5], Mallory L. Downie[1], Sanjana Gupta[1], Daniel P. Gale[1], Detlef Bockenhauer[1], Robert Kleta[1,7], Mauricio Arcos-Burgos[6,7] and Horia C. Stanescu[1,7]

[1]Centre for Genetics and Genomics, Department of Renal Medicine, UCL Division of Medicine, University College London, London, UK; [2]Department of Human and Medical Genetics, Faculty of Medicine, Vilnius University, Vilnius, Lithuania; [3]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; [4]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; [5]Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA; and [6]Grupo de Investigación en Psiquiatría, Departamento de Psiquiatría, Instituto de Investigaciones Médicas, Facultad de Medicina, Universidad de Antioquia, Medellín, Colombia

Class 2 HLA and PLA2R1 alleles are exceptionally strong genetic risk factors for membranous nephropathy (MN), leading, through an unknown mechanism, to a targeted autoimmune response. Introgressed archaic haplotypes (introduced from an archaic human genome into the modern human genome) might influence phenotypes through gene dysregulation. Here, we investigated the genomic region surrounding the PLA2R1 gene. We reconstructed the phylogeny of Neanderthal and modern haplotypes in this region and calculated the probability of the observed clustering being the result of introgression or common descent. We imputed variants for the participants in our previous genome-wide association study and we compared the distribution of Neanderthal variants between MN cases and controls. The region associated with the lead MN risk locus in the PLA2R1 gene was confirmed and showed that, within a 507 kb region enriched in introgressed sequence, a stringently defined 105 kb haplotype, intersecting the coding regions for PLA2R1 and ITGB6, is inherited from Neanderthals. Thus, introgressed Neanderthal haplotypes overlapping PLA2R1 are differentially represented in MN cases and controls, with enrichment In controls suggesting a protective effect.

Kidney International (2024) ■, ■–■; https://doi.org/10.1016/j.kint.2024.01.017

KEYWORDS: chronic kidney disease; membranous nephropathy; nephrotic syndrome

Correspondence: *Horia C. Stanescu, Centre for Genetics and Genomics, Department of Renal Medicine, UCL Division of Medicine, University College London, First Floor Royal Free Hospital, Rowland Hill Street, London NW3 2PF, UK. E-mail: h.stanescu@ucl.ac.uk*

[7]RK, MA-B and HCS contributed equally to this work.

Received 19 January 2023; revised 27 December 2023; accepted 8 January 2024

Translational Statement

Membranous nephropathy (MN) is an important example of a disorder in which careful molecular analysis furthers the understanding of an autoimmune pathogenetic mechanism. Much progress has been done regarding the qualitative (structural) aspects of antigenicity of the autoantibody, but we believe that this mechanism is complemented by a quantitative (level of expression) component, which still needs further investigation. Our work provides insight into the structure of the genomic region putatively involved in the control of the expression of the major autoantigen in MN.

Idiopathic or primary membranous nephropathy (MN) is a rare kidney-specific autoimmune disease. The incidence of MN is 10 to 12 per million persons per year.[1,2] Despite its rarity, it is the leading cause of nephrotic syndrome in European adults.[3] MN progresses to end-stage kidney disease in 30% to 40% of cases within 5 years.[4–6]

MN is an adult-onset disease with a peak age of onset between the fifth and sixth decades of life. Over 80% of patients are aged >40 years at presentation, and MN is uncommon in children.[7] In contrast to most other autoimmune diseases, males are more commonly affected by MN than females.[8]

MN is an autoimmune disorder in which a structure of the self (i.e., a fragment of phospholipase A2 receptor [PLA2R], the most frequent autoimmune target in MN) is recognized as an antigen with consequent transfer onto the human leukocyte antigen (HLA)-II antigen-presenting groove.[9–11] This triggers a targeted autoimmune response against the molecule from which the intrinsic epitope originates. The probability of this occurring likely depends on the level of expression of *PLA2R1*: variants predisposing to the development of MN have been shown to be associated with an increase of *PLA2R1* glomerular expression (probably through the modification of regulatory elements).[11]

Homo sapiens neanderthalensis is an extinct subspecies of humans who populated Eurasia 200,000 years ago.[12] They extended toward Asia and were present in Europe between 200,000 and 40,000 years ago.[13] This allowed plenty of time for them to become well adapted to the local (geoclimatic and biological) environment. In contrast, *Homo sapiens sapiens* started populating Europe ≈50,000 years ago.[14,15] They encountered the local population of *H sapiens neanderthalensis*, and the interaction between the 2 groups seems to have followed all possible avenues: avoidance, conflict, and mixing.[13,15,16]

Sapiens-neanderthalensis interbreeding was a contentious aspect of the history of the 2 hominin groups, debated in the scientific community for a long time.[17] More recently, compelling evidence has gathered in its favor, mainly starting with the sequencing of the Neanderthal genome.[12,13,18]

As a result of the interbreeding, ≈1% to 3% of the *H sapiens sapiens* genome is believed to be of introgressed Neanderthal origin.[12(p201),13,19,20] with specific loci being particularly prone to originate from the ancient group. Neanderthal DNA is not evenly distributed throughout the genome—particularly rich in some areas, almost absent in others—thus strongly suggesting positive selection pressure as a mechanism involved in preserving the ancient DNA stretches.[13]

This scenario would make sense given that the newcomers (*H sapiens sapiens*) were less adapted to their new environment, and that those DNA stretches of the older inhabitants coding for phenotypic traits advantageous for survival in European conditions would enhance their ability to survive. Examples of traits/loci of Neanderthal origin that fit the positive selection hypothesis include skin thickness and hair density, which might have been advantageous in the colder climatic context of Europe versus Africa.[21]

Immunity has previously been shown to be 1 of the aspects of the modern human phenotype influenced by Neanderthal DNA introgression, with links between ancient variants and infections and autoimmune disorders having been convincingly established.[20] It was shown that Neanderthal haplotypes influence our susceptibility to coronavirus disease 2019 (COVID-19),[22,23] HIV,[24] and *Helicobacter pylori*[25] infections. Variants inherited from Neanderthals were also linked to autoimmune conditions, such as rheumatoid arthritis, systemic lupus erythematosus, Crohn disease, Grave disease, biliary cirrhosis, and type 2 diabetes, among others.[13,20,26–29] We, therefore, investigated if susceptibility to MN could also be influenced by Neanderthal variants retained in our modern genomes.

We define our region of interest to be the 1-Mb genomic region surrounding the *PLA2R1* gene (chromosome 2:160500000-161500000; reference genome: hg19). Previously published ancestral introgression maps suggested Neanderthal sequence enrichment in the *PLA2R1* gene region.[18,20,27,30,31]

METHODS

Definition of the target genomic region

We investigate a genomic region of ≈1 million bases centered around rs4664308, the lead single-nucleotide polymorphism (SNP) on chromosome 2 from the 2011 MN genome-wide association study (GWAS).[9] This region is defined by coordinates chromosome 2:160500000-161500000 (hg19) and harbors the *PLA2R1* gene and its 5′ upstream and 3′ downstream regions (Figure 1).

Archaic DNA

Data for 3 Neanderthal genomes (Chagyrskaya 8,[32] Vindija 33.19,[33] and Altai Denisova 5[12]) were located through the Allen Ancient DNA Resource version 44.3[34] and downloaded from the Max Planck Institute for Evolutionary Anthropology (see Data Statement).

Genotyping

The genotyping for the British cohort of the study by Stanescu *et al.* in 2011 has been performed by deCODE genetics on DNA samples from 335 case patients and 349 ethnically matched controls, as described in the article.[9]

Imputation of human variants

We used Beagle 5.1 to phase[35] and to impute[36] variants for the UK participants from our previously published study[9] (the parameters used were: burnin = 6, iterations = 12, phase states = 280, imputed states = 1600, imputed segment = 6.0, step = 0.1, imputed nsteps = 7, cluster = 0.005). We used all 2504 genomes from the 1000 Genomes Project (phase 3 release)[37] as the reference panel (see Data Statement).

For the preimputation quality control, we filtered the samples by call rate > 0.95, and variants by call rate > 0.99, minor allele frequency > 0.01, and Hardy-Weinberg equilibrium for controls $P > 0.0001$. We also removed variants that had no variation in either cases or controls. Imputation quality control was performed by removing variants below an R^2 threshold of 0.8.

Definition of Neanderthal variation and consensus haplotype

We used bcftools 1.10.2[38] and tabix 1.10.2[39] to extract the variants within the region of interest, both from the British cohort of the study by Stanescu *et al.* in 2011 and from the 3 Neanderthal genomes. Variants with known archaic alleles were selected, and linkage disequilibrium (LD) with rs4664308 (the lead SNP from the article by Stanescu *et al.* in 2011) was estimated with LDproxy (Ldlink 5.1)[40] using the British participants to the 1000 Genomes Project as a reference. We used bcftools to extract haplotypes for the region of high LD from the 1000 Genomes Project and the 3 Neanderthal genomes, including only variants for which data are available in both the 1000 Genomes data set and at least 1 of the Neanderthal genomes, and we converted haplotype data to PHYLIP format with an in-house Python script.

We created 2 consensus Neanderthal haplotypes for the variants in the region of interest, at different levels of stringency: a majority consensus by selecting the allele present in most of the known Neanderthal haplotypes at each locus, and a unanimous consensus by selecting only variants where all 3 Neanderthal samples are available, homozygous, and in unanimous agreement.

Evolutionary tree reconstruction

The reconstruction of the phylogeny of the haplotypes from the 1000 Genomes Project in the region of high LD was performed using phyml version 3.3.3 with the default parameters (neighbor joining using the Hasegawa-Kishino-Yano substitution matrix),[41] and the resulting trees were visualized with the ETE toolkit[42] and with iTOL.[43] Ancestral allele information from Ensembl, as supplied in the 1000 Genomes Project data set, was used for rooting the trees.

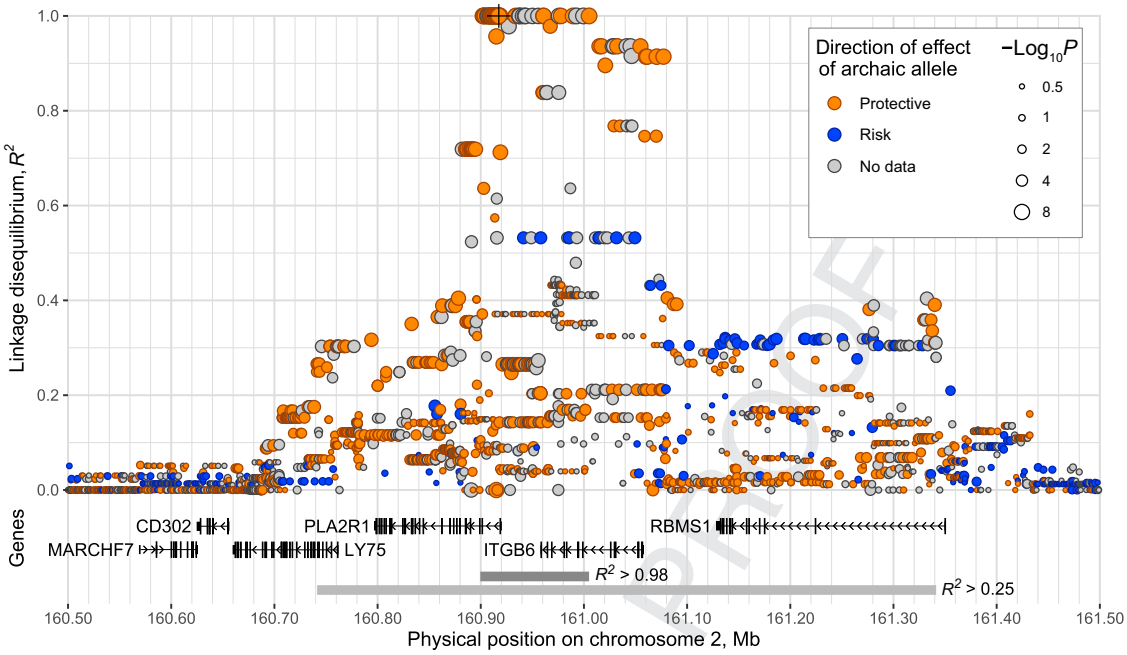

Figure 1 | Linkage disequilibrium (LD) plot of the region of interest on chromosome 2, indicating the locations of the genes contained therein. The *x* axis shows genomic coordinates (GRCh37); the *y* axis represents R^2 as a measure of LD for the British population in the 1000 Genomes Project, with respect to the lead membranous nephropathy risk single-nucleotide polymorphism (SNP) on chromosome 2 (rs4664308; marked with crosshairs). The darker gray bar marks the region of high LD ($R^2 > 0.98$) with this SNP, and the lighter gray bar marks the region of moderate LD ($R^2 > 0.25$). Only variants for which the genome-wide association study identifies a direction of effect are included. Circle colors show the origin and direction of effect of the alleles: blue means the consensus Neanderthal allele is the risk allele and the modern allele is protective; orange means the Neanderthal allele is protective and the modern one is the risk allele. Variants are colored gray where the archaic allele is unknown or ambiguous (no Neanderthal consensus). The size of the circles represents the strength of association (*P* value). Mb, xxx. Q14Q15

Comparing introgression with common ancestry

We used the method published by Huerta-Sánchez *et al.*[44] to investigate the likelihood of having an extended haplotype originating from a common ancestor of Neanderthals and present-day humans, as opposed to the introgression hypothesis. We interpolated the local recombination rate to 0.471 cM per Mb (between markers D2S156 and D2S2395) from the data set published by Kong *et al.*,[45] assuming the divergence between Neanderthals and modern humans took place 550,000 years ago,[12] interbreeding between Neanderthals and modern humans took place ≈50,000 years ago, and the average generational time is 29 years.[46] Q7

Direction of effect and archaic/modern origin of alleles

For the cases and controls from the British cohort of the study by Stanescu *et al.* in 2011, we performed a basic allele test (additive model) on biallelic variants with minor allele frequency > 0.05, as implemented in PLINK 2.0 (www.cog-genomics.org/plink/2.0/).[47] No adjustment for confounders was used, as the cases and controls are ethnically matched. This was verified by primary component analysis (see Supplementary Figure S1). The direction of effect of the alleles was determined using the odds ratio (odds ratio > 1 [risk] and odds ratio < 1 [protective]) for each variant on the imputed data.

We grouped variants with a known direction of effect in 3 categories: variants where the risk allele is the consensus Neanderthal allele; variants where the risk allele is the modern allele (i.e., the protective allele is the consensus Neanderthal allele); and variants where the consensus Neanderthal allele is unknown or undefined (no majority).

We created an LD plot based on the R^2, color coded according to the 3 categories described above.

Statistical assessment of haplotype overrepresentation

We counted the number of positions where the haplotypes in the study differed from the Neanderthal consensus haplotype and used R version 4.3.1 to decide on the statistical significance of the difference in distribution between cases and controls.

Functional annotation of the variants in the introgressed sequence

We performed detailed functional annotations for all the 1237 variants identified in the introgressed sequence. We used 3 data sets sourced from the University of California, Santa Cruz, Genome Browser[48] as follows: PhastCons[49] Conservation Elements (at a conservation score > 500; range, 0–1000). This allowed us to detect queried SNPs that fall within conserved genomic elements. DnaseI from ENCODE[50]/OpenChrom was used for identifying queried SNPs in accessible, open chromatin regions. Transcription factor chromatin immunoprecipitation–sequencing clusters (spanning 338 factors and 130 cell types) from ENCODE 3[51] were used to determine intersections between queried SNPs and transcription factor binding sites.

We employed the wANNOVAR[52] web tool to annotate the functional consequences (missense, synonymous) of all coding variants. The tool also predicted the potential impact on protein sequences based on metrics from SIFT, PolyPhen, CADD, and GERP++ pathogenicity scores. Q8

We queried general (GTEx[53]) and kidney-specific expression quantitative trait loci (eQTL) data sets (NephQTL[54] and the Human Kidney eQTL Atlas[55]).

RESULTS

Preimputation/postimputation coverage

In our region of interest, chromosome 2:160500000-161500000 (hg19), we have 102 genotyped variants (post quality control, preimputation) and 4355 variants post-imputation at an R^2 threshold of 0.8.

Regions of LD

Q9 We found a 105-kb region, chromosome 2:160899566-161004983 (hg19), to be in high ($R^2 > 0.98$) LD with rs4664308, the lead SNP from the 2011 GWAS. A 507-kb region, chromosome 2:160741415-161341364 (hg19), is in moderate LD ($R^2 > 0.25$) with the same SNP (Figure 1, dark and light gray bars).

Presence of a Neanderthal-like haplotype in the modern population

In the phylogenetic tree reconstructed from all 5008 haplotypes from the 1000 Genomes Project and the 3 Neanderthal sequences, we observed clustering of all 3 archaic sequences along with a subset of 1011 modern haplotypes, with a distinctly long branch length and excellent bootstrap support of 1.0 (Figure 2, shaded in gray). The self-declared ancestry of the 1000 Genomes Project participants in this 1011-haplotype cluster was mainly European (411 haplotypes) and East Asian (323 haplotypes). There were 43 haplotypes belonging to individuals of self-declared African ancestry.

The clustering and branch lengths show that, in the region of high LD, Neanderthals and some modern humans share a nearly identical haplotype, which is distinct from other haplotypes present in most modern humans.

Evidence of introgression

We calculated the expected length of an ancestral haplotype shared by inheritance over a total branch length of 1,050,000 years (t_0 = 36,207 generations): L_0 = 5864 base pairs. For comparison, the expected length of a haplotype following introgression 50,000 years ago, t_1 = 1724 generations, is L_1 = 123,142 base pairs. The probability of shared inheritance of an ancestral 105.4-kb haplotype is $P_0 = 2.96 \times 10^{-7}$; the shared haplotype is thus almost certainly the result of recent introgression. This is a conservative estimate of the probability, as it does not account for the fact that shared inheritance would imply that variants shared between non-Africans and Neanderthals had to have been lost in Africans.

Direction of effect of Neanderthal variants

After imputing the 2011 study data and performing the GWAS on the imputed genotypes, we found 361 biallelic single-nucleotide substitutions within the high LD interval, chromosome 2:160899566-161004983 (hg19), for which the GWAS identifies a direction of effect. For 199 of these, we have data in at least 1 of the 3 Neanderthal genomes, and a Neanderthal majority consensus can be determined (see Supplementary Figure S2 and Supplementary Tables S1 and S2).

Within the larger genomic region, chromosome 2:160500000-161500000 (hg19), Figure 1 indicates the origin of the risk and protective alleles, showing enrichment of Neanderthal variants with protective effect. Within the high LD interval, using the majority consensus, 193 Neanderthal alleles are protective and 6 are deleterious with respect to MN (see Supplementary Figure S3 and Supplementary Table S3 for the result of the more stringent unanimous Neanderthal consensus).

Evidence of a protective Neanderthal haplotype

There are 320 biallelic single-nucleotide variants within the high LD interval where a Neanderthal majority consensus can be determined, and there is variation within the 2011 study cohort (these include the 199 variants discussed above). The number of differences between the haplotypes in the study and the Neanderthal majority consensus sequence at these loci follows a strongly bimodal distribution. We defined haplotypes with 0 to 10 differences to be Neanderthal-like and those with >10 differences to be modern, but any cutoff between 7 and 80 gives essentially the same result (see Supplementary Figure S4). We found that 166 of the 668 haplotypes belonging to cases (24.9%) and 279 of the 698 belonging to controls (40.0%) are Neanderthal-like, and 502 haplotypes belonging to cases (75.1%) and 419 belonging to controls (60.0%) are modern (Fisher exact test $P = 2.5 \times 10^{-9}$) (see Supplementary Table S3 for the result of the more stringent unanimous Neanderthal consensus).

Functional annotation of the variants in the introgressed sequence

Among the 1237 variants identified in the introgressed sequence, there are 11 exonic variants, 5 of which are non-synonymous, with only 1 having a CADD score slightly higher than the suggested cutoff of 20 (rs2271381; CADD phred-like score = 22.8; in the exon 5 of LY75).

We also identified 2 glomerulus-specific eQTLs in the extended Neanderthal haploblock (rs13408963 and rs72957586). These 2 variants are part of the set of 303 genome-wide significant glomerular eQTLs (at FPR Q10 threshold < 0.05). The first variant, rs13408963, is included in the high LD region. It lies 37 kb upstream of the transcription start site of PLA2R1 and is not annotated by ENCODE as a known regulatory site. There is strong evidence that the "G" allele of the rs13408963 variant has an enhancing effect (all 4 studies listed in Human Kidney eQTL Atlas). There are no direct Neanderthal data for this variant, but it is in strong LD (R^2 = 0.93) with 3 protective Neanderthal variants (rs4076844, rs4470295, and rs4522556). We notice that the direction of effect of the Neanderthal introgressed allele ("A"), that is, decreased expression of *PLA2R1*, is consistent with the protective effect observed in the GWAS ($-\log_{10} P = 5.78$). The second variant, rs72957586, is situated

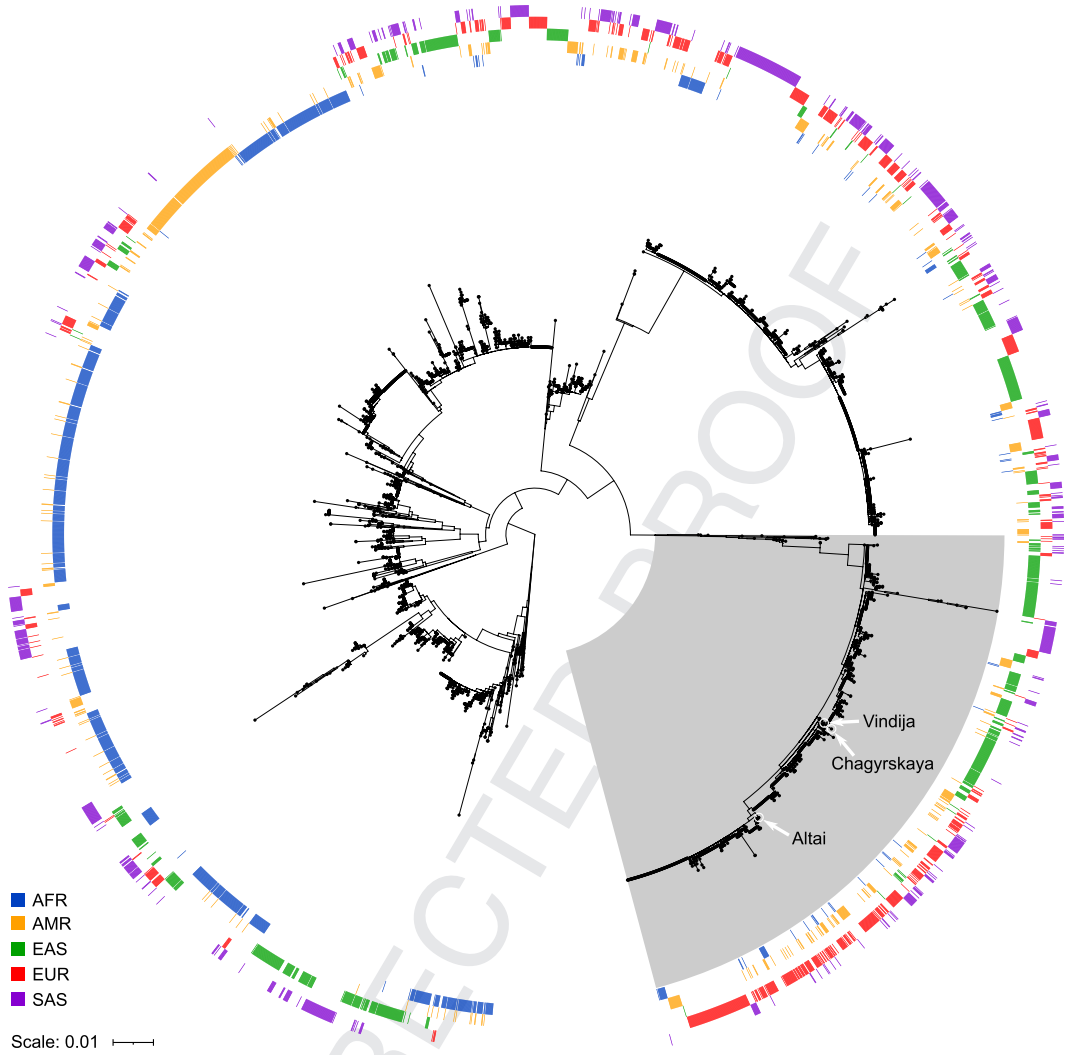

Figure 2 | Phylogenetic tree of modern and Neanderthal haplotypes (all 5008 haplotypes from the 2504 participants in the 1000 Genomes Project, plus the 3 Neanderthal sequences available to date), based on the 105-kb region of chromosome 2 in association with membranous nephropathy risk. The tree represents the reconstructed structure of the relationships between the individuals considered (black dots), including their putative ancestors (branch points). The inferred common ancestor is in the center. Arrows indicate the 3 Neanderthal sequences (Altai, Chagyrskaya, and Vindija). Branch lengths (radial lines) represent genetic distance (number of substitutions per variant, see bar); longer radial paths imply a lower degree of relatedness. To convey the frequencies of the haplotypes, duplicates have not been removed: sets of identical haplotypes appear as trains of partially overlapping dots. The surrounding color tracks depict the self-reported ancestry from the 1000 Genomes Project participants; from the innermost track outward as follows: blue, African (AFR); yellow, American (AMR); green, East Asian (EAS); red, European (EUR); and violet, South Asian (SAS). Gray shading highlights a cluster including all 3 Neanderthals joined together by short branches and connected to the rest of the tree by a long branch. It clearly indicates that these haplotypes are similar to each other and distinct from the rest. Note the underrepresentation of African ancestry and the overrepresentation of European and East Asian ancestries among these Neanderthal-like haplotypes.

1.5 kb upstream of *LY75*, in the predicted promoter flanking region of *LY75* (for the HepG2 cell line). It also has no direct Neanderthal data, but it is in perfect ($R^2 = 1$) LD with 9 protective ($-\log_{10} P = 5.26$) Neanderthal variants (rs3806603, rs3792198, rs925407, rs57771571, rs6735923, rs72957578, rs72957576, rs7601374, and rs55977890). It is also an eQTL for whole blood according to GTEx. In this case, the assumed Neanderthal allele ("A") enhances the expression of LY75. All the functional annotations for the variants identified in the introgressed sequence can be found in the Supplementary Data S1.

DISCUSSION

Here we describe a genomic region associated with MN,[9] where clustering in the phylogenetic tree reconstruction shows Neanderthal sequence enrichment in modern-day humans. Although previously published ancestral introgression maps suggested archaic sequence enrichment in the *PLA2R1* gene region,[18,20,27,30,31] we show that the Neanderthal-like haplotypes in this region are enriched for variants protective against MN.

The most stringently defined haploblock in this region is 105 kb long. On the basis of the length of this haploblock and

the recombination rate in the region, we show that this is far more likely to be the result of relatively recent introgression (≈50,000 years ago), than being inherited from a common ancestor of Neanderthals and modern humans.

This is also reinforced by the near absence of the Neanderthal-like haplotype in the 661 individuals of self-reported African ancestry in the 1000 Genomes Project (1322 haplotypes). Of the 43 such occurrences, 31 are in individuals from the Caribbean and the US South-West, where partial non-African ancestry is likely.

The larger region of introgression contains 4 genes: the beginning of *LY75* (5' end: exons 1 to 6 of 34), *PLA2R1*, *ITGB6*, and *RBMS1* (with the exception of its first exon).

The PLA2R protein (Uniprot Q13018), a member of the C-type lectin superfamily of still unclear function, is the most important autoantigen involved in MN pathogenesis.[56] LY75 (Uniprot O60449), also a member of the C-type lectin superfamily, is a myeloid C-type lectin receptor, putatively involved in pathogen recognition and modulation of the immune host response. Its ligands are HIV, *Yersinia pestis*, and endogenous dead cells.[57] The *ITGB6* gene encodes the integrin β 6 protein (Uniprot P18564), 1 of the 2 subunits of the integrin α-v:β-6 heterodimer. Integrin αvβ6 is a receptor with roles in cell adhesion and innate immune surveillance modulation.[58] It is also an attachment factor for viruses of the Coxsackie/Enterovirus B family[59] and of the Herpes family.[60]

RBMS1 (Uniprot P29558) is a protein that binds single-stranded DNA/RNA that can potentially suppress transcription of target genes.[61] It is considered to be 1 of the key regulators of the *c-myc* proto-oncogene expression[62] and appears to be involved in a large variety of biological functions (replication, transcription, cell cycle control, and apoptosis).

The stringently defined introgression region contains the beginning of *PLA2R1* (5′ end: first and second exon) and the end of *ITGB6* (3′ end: exons 8 to 15, with both genes being encoded on the negative strand), as well as the respective intergenic region.

Traits related to most chronic kidney disease are not under direct evolutionary pressure because they typically manifest later in life, after the reproductive years, and therefore do not directly impact reproductive fitness. Thus, where renal loci appear to be under selection, this is more likely to be caused by neighboring nonrenal loci.[63]

The phylo-ontogenetic development of the immune system, by creating either resistance or tolerance, has resulted in a finely balanced system that, if disrupted, can lead to immunodeficiency or autoimmunity. One of the most substantial adaptive pressures a biological population faces in its interaction with the environment is the 1 from infectious agents. The permanent arms race between invaders (nonself) and defenders (self; i.e., their immune systems) leads to the constant refining of both defense mechanisms and defense evasion mechanisms. From an evolutionary point of view, it is a relatively fast process.[64]

Neanderthals, by long preceding *H sapiens sapiens* in Europe, were probably better adapted to the local regional pathogen spectrum; hence, introgression might have contributed to the adaptation of the latter to the newly encountered infections.[65] Thus, risk factors for autoimmune diseases could have originated as adaptations to the infectious diseases of the time, but what we see today might be an unfortunate rebalancing of the immune system.

A variant on an archaic haplotype could have conferred an adaptive advantage in the arms race between pathogens and ancient humans, leading to the enrichment of the archaic variant at the corresponding locus. The sequence introgressed into the modern human genome also contains the regulatory region upstream of *PLA2R1*. Transcriptional regulatory elements, either directly adaptive or "hitchhiking" on the archaic haplotype, could have modified the expression of PLA2R in a way that made it less prone to act as an autoantigen under the conditions that might otherwise have triggered MN.

We have been able to identify 2 variants within the extended Neanderthal haploblock that are glomerulus-specific eQTLs: rs13408963, an eQTL for *PLA2R1* (situated upstream of it); and rs72957586, an eQTL for *LY75* (situated in its promoter flanking region). The *PLA2R1* variant is consistent with the protective effect noticed for the introgressed sequence, with the Neanderthal allele corresponding to lower *PLA2R1* expression.

We are aware of some general limitations of paleogenomic studies, such as the difficulty of distinguishing between introgression and positive selection. Introgressed material in noncoding, regulatory regions has been shown to be mainly under negative selection pressure, whereas positive selection pressure tends to act on coding regions.[66] As our region of near-perfect LD includes both, a neutral assumption ruling out positive selection in this case appears justified. Another limitation of human introgression studies is archaic sequence availability. This is partially alleviated by the growing understanding of the fact that archaic populations were less variable than modern ones and will hopefully improve with the publication of more high-quality sequenced ancient genomes. In the same vein of sequence availability and quality, it has to be mentioned that imputation (as opposed to direct sequencing) could be a potential source of errors.

Our study intended to begin the exploration of the structure of archaic introgression in the *PLA2R1* region. As such, it did not explore a more diverse set of populations, and it did not consider epistatic interactions between the *PLA2R1* gene and other loci, such as *HLA*. We also acknowledge that functional annotation is not a substitute for experimental confirmation, and the proposed mechanisms are speculative.

Human evolution has recently been the focus of detailed scientific investigation involving the acquisition, storage, dating, and curation of referenced genome sequences and their variation across archaic humans.[13,67–71]

Until recently, the relationship between introgression and autoimmunity has been assumed to involve variants that

enhance fitness in reproductive age at the cost of an altered immune system balance later in life.

Ours is an interesting example where autoimmunity-related introgressed variants have the opposite effect (protective in later life) in modern humans, as has been recently shown for other immune-modulating traits by McArthur *et al.*[67]

DISCLOSURE

All the authors declared no competing interests.

DATA STATEMENT

Imputed variants for the region of interest on chromosome 2, British cohort of the study by Stanescu *et al.* in 2011—EGAD00010002641 MN2011UKcontrol and EGAD00010002642 MN2011UKcase Genome-wide association study (GWAS) summary statistics, chromosome 2: EBI GWAS catalog GCST90302187.

Neanderthal genomes links are as follows: ftp.eva.mpg.de/neandertal/altai/AltaiNeandertal/, ftp.eva.mpg.de/neandertal/Chagyrskaya/, and ftp.eva.mpg.de/neandertal/Vindija/. 1000 Genomes Reference is ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase3/. The Beagle Imputation Reference Panel is https://bochet.gcc.biostat.washington.edu/beagle/1000_Genomes_phase3_v5a/.

ACKNOWLEDGMENTS

RK is supported by the Potter Foundation, and DPG is supported by the St. Peter's Trust. Part of this work was supported by grants from St. Peter's Trust and Kidney Research UK.

SUPPLEMENTARY MATERIAL

Supplementary File (PDF)

Supplementary Figure S1. Principal component analysis of the British cohort from the study by Stanescu *et al.* in 2011.[9]

Supplementary Figure S2. Venn diagrams of variants employed in the determination of direction of effect of Neanderthal alleles and the identification of Neanderthal-like haplotypes in cases and controls.

Supplementary Figure S3. Linkage disequilibrium (LD) plots comparing choices of methods for determining the Neanderthal consensus haplotype (majority or unanimous) and the direction of effect of variants (with or without 95% confidence interval threshold).

Supplementary Figure S4. Distribution of the number of differences between the haplotypes of cases and controls in the region of high linkage disequilibrium (LD) ($R^2 > 0.98$) and the consensus Neanderthal haplotype.

Supplementary Table S1. Neanderthal variants in the region of high linkage disequilibrium (LD) ($R^2 > 0.98$) and the determination of the majority allele.

Supplementary Table S2. Summary of the distribution of Neanderthal alleles in the region of high linkage disequilibrium (LD) ($R^2 > 0.98$).

Supplementary Table S3. Comparison of the majority allele method and the homozygous unanimity method in determining the Neanderthal consensus: risk/protective variant count and haplotype counts.

Supplementary File (XLSX)

Supplementary Data S1. Functional annotation of variants.

REFERENCES

1. Couser WG. Primary membranous nephropathy. *Clin J Am Soc Nephrol*. 2017;12:983–997.
2. McGrogan A, Franssen CF, de Vries CS. The incidence of primary glomerulonephritis worldwide: a systematic review of the literature. *Nephrol Dial Transplant*. 2011;26:414–430.
3. Ronco P, Beck L, Debiec H, et al. Membranous nephropathy. *Nat Rev Dis Primer*. 2021;7:69.
4. Hogan SL, Muller KE, Jennette JC, Falk RJ. A review of therapeutic studies of idiopathic membranous glomerulopathy. *Am J Kidney Dis*. 1995;25:862–875.
5. Lai WL, Yeh TH, Chen PM, et al. Membranous nephropathy: a review on the pathogenesis, diagnosis, and treatment. *J Formos Med Assoc Taiwan Yi Zhi*. 2015;114:102–111.
6. Jha V, Ganguli A, Saha TK, et al. A randomized, controlled trial of steroids and cyclophosphamide in adults with nephrotic syndrome caused by idiopathic membranous nephropathy. *J Am Soc Nephrol*. 2007;18:1899–1904.
7. Glassock RJ. Diagnosis and natural course of membranous nephropathy. *Semin Nephrol*. 2003;23:324–332.
8. Gupta S, Köttgen A, Hoxha E, et al. Genetics of membranous nephropathy. *Nephrol Dial Transplant*. 2018;33:1493–1502.
9. Stanescu HC, Arcos-Burgos M, Medlar A, et al. Risk HLA-DQA1 and PLA(2) R1 alleles in idiopathic membranous nephropathy. *N Engl J Med*. 2011;364:616–626.
10. Sanchez-Rodriguez E, Southard CT, Kiryluk K. GWAS-based discoveries in IgA nephropathy, membranous nephropathy, and steroid-sensitive nephrotic syndrome. *Clin J Am Soc Nephrol*. 2021;16:458–466.
11. Xie J, Liu L, Mladkova N, et al. The genetic architecture of membranous nephropathy and its potential to improve non-invasive diagnosis. *Nat Commun*. 2020;11:1600.
12. Prüfer K, Racimo F, Patterson N, et al. The complete genome sequence of a Neanderthal from the Altai mountains. *Nature*. 2014;505:43–49.
13. Green RE, Krause J, Ptak SE, et al. Analysis of one million base pairs of Neanderthal DNA. *Nature*. 2006;444:330–336.
14. Lazaridis I, Nadel D, Rollefson G, et al. Genomic insights into the origin of farming in the ancient Near East. *Nature*. 2016;536:419–424.
15. Fu Q, Posth C, Hajdinjak M, et al. The genetic history of Ice Age Europe. *Nature*. 2016;534:200–205.
16. Noonan JP, Coop G, Kudaravalli S, et al. Sequencing and analysis of Neanderthal genomic DNA. *Science*. 2006;314:1113–1118.
17. Hodgson JA, Disotell TR. No evidence of a Neanderthal contribution to modern human diversity. *Genome Biol*. 2008;9:206.
18. Vernot B, Akey JM. Resurrecting surviving Neandertal lineages from modern human genomes. *Science*. 2014;343:1017–1021.
19. Wall JD, Yang MA, Jay F, et al. Higher levels of neanderthal ancestry in East Asians than in Europeans. *Genetics*. 2013;194:199–209.
20. Sankararaman S, Mallick S, Dannemann M, et al. The genomic landscape of Neanderthal ancestry in present-day humans. *Nature*. 2014;507:354–357.
21. Gittelman RM, Schraiber JG, Vernot B, et al. Archaic hominin admixture facilitated adaptation to out-of-Africa environments. *Curr Biol*. 2016;26:3375–3382.
22. Zeberg H, Pääbo S. The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. *Nature*. 2020;587:610–612.
23. Huffman JE, Butler-Laporte G, Khan A, et al. Multi-ancestry fine mapping implicates OAS1 splicing in risk of severe COVID-19. *Nat Genet*. 2022;54:125–127.
24. Kist NC, Lambert B, Campbell S, et al. HIV-1 p24Gag adaptation to modern and archaic HLA-allele frequency differences in ethnic groups contributes to viral subtype diversification. *Virus Evol*. 2020;6:veaa085.
25. Dannemann M, Andrés AM, Kelso J. Introgression of Neandertal- and Denisovan-like haplotypes contributes to adaptive variation in human toll-like receptors. *Am J Hum Genet*. 2016;98:22–33.
26. Almarri MA, Bergström A, Prado-Martinez J, et al. Population structure, stratification, and introgression of human structural variation. *Cell*. 2020;182:189–199.e15.
27. Dannemann M, Kelso J. The contribution of Neanderthals to phenotypic variation in modern humans. *Am J Hum Genet*. 2017;101:578–589.
28. Nédélec Y, Sanz J, Baharian G, et al. Genetic ancestry and natural selection drive population differences in immune responses to pathogens. *Cell*. 2016;167:657–669.e21.
29. McCoy RC, Wakefield J, Akey JM. Impacts of Neanderthal-introgressed sequences on the landscape of human gene expression. *Cell*. 2017;168:916–927.e12.
30. Deschamps M, Laval G, Fagny M, et al. Genomic signatures of selective pressures and introgression from archaic hominins at human innate immunity genes. *Am J Hum Genet*. 2016;98:5–21.
31. Skov L, Coll Macià M, Sveinbjörnsson G, et al. The nature of Neanderthal introgression revealed by 27,566 Icelandic genomes. *Nature*. 2020;582:78–83.

32. Mafessoni F, Grote S, de Filippo C, et al. A high-coverage Neandertal genome from Chagyrskaya Cave. *Proc Natl Acad Sci U S A*. 2020;117:15132–15136.
33. Prüfer K, de Filippo C, Grote S, et al. A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science*. 2017;358:655–658.
34. Reich D. Allen Ancient DNA Resource (AADR): downloadable genotypes of present-day and ancient DNA data | David Reich Lab. Published October 10, 2021. Accessed December 9, 2021. https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data
35. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet*. 2007;81(5):1084–1097.
36. Browning BL, Zhou Y, Browning SR. A One-penny imputed genome from next-generation reference panels. *Am J Hum Genet*. 2018;103:338–348.
37. 1000 Genomes Project Consortium, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature*. 2015;526:68–74.
38. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinforma Oxf Engl*. 2011;27:2987–2993.
39. Li H. Tabix: fast retrieval of sequence features from generic TAB-delimited files. *Bioinforma Oxf Engl*. 2011;27:718–719.
40. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinforma Oxf Engl*. 2015;31:3555–3557.
41. Guindon S, Dufayard JF, Lefort V, et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 2010;59:307–321.
42. Huerta-Cepas J, Serra F, Bork P. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol Biol Evol*. 2016;33:1635–1638.
43. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res*. 2021;49:W293–W296.
44. Huerta-Sánchez E, Jin X, Asan, et al. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature*. 2014;512:194–197.
45. Kong A, Gudbjartsson DF, Sainz J, et al. A high-resolution recombination map of the human genome. *Nat Genet*. 2002;31:241–247.
46. Zeberg H, Dannemann M, Sahlholm K, et al. A Neanderthal sodium channel increases pain sensitivity in present-day humans. *Curr Biol*. 2020;30:3465–3469.e4.
47. Chang CC, Chow CC, Tellier LC, et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*. 2015;4:7.
48. Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. *Genome Res*. 2002;12:996–1006.
49. Siepel A, Bejerano G, Pedersen JS, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res*. 2005;15:1034–1050.
50. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489:57–74.
51. Wang J, Zhuang J, Iyer S, et al. Factorbook.org: a Wiki-based database for transcription factor-binding data generated by the ENCODE consortium. *Nucleic Acids Res*. 2013;41:D171–D176.
52. Chang X, Wang K. wANNOVAR: annotating genetic variants for personal genomes via the web. *J Med Genet*. 2012;49:433–436.
53. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;45:580–585.
54. Gillies CE, Putler R, Menon R, et al. An eQTL landscape of kidney tissue in human nephrotic syndrome. *Am J Hum Genet*. 2018;103:232–244.
55. Qiu C, Huang S, Park J, et al. Renal compartment-specific genetic variation analyses identify new pathways in chronic kidney disease. *Nat Med*. 2018;24:1721–1731.
56. Beck LH Jr, Bonegio RG, Lambeau G, et al. M-type phospholipase A2 receptor as target antigen in idiopathic membranous nephropathy. *N Engl J Med*. 2009;361:11–21.
57. Osorio F, Reis e Sousa C. Myeloid C-type lectin receptors in pathogen recognition and host defense. *Immunity*. 2011;34:651–664.
58. Koivisto L, Bi J, Häkkinen L, Larjava H. Integrin αvβ6: structure, function and role in health and disease. *Int J Biochem Cell Biol*. 2018;99:186–196.
59. Agrez MV, Shafren DR, Gu X, et al. Integrin alpha v beta 6 enhances coxsackievirus B1 lytic infection of human colon cancer cells. *Virology*. 1997;239:71–77.
60. Gianni T, Salvioli S, Chesnokova LS, et al. αvβ6- and αvβ8-integrins serve as interchangeable receptors for HSV gH/gL to promote endocytosis and activation of membrane fusion. *PLoS Pathog*. 2013;9:e1003806.
61. Zhang W, Sun Y, Bai L, et al. RBMS1 regulates lung cancer ferroptosis through translational control of SLC7A11. *J Clin Invest*. 2021;131:e152067.
62. Aggarwal P, Bhavesh NS. Hinge like domain motion facilitates human RBMS1 protein binding to proto-oncogene c-myc promoter. *Nucleic Acids Res*. 2021;49:5943–5955.
63. Adeyemo AA, Shriner D, Bentley AR, et al. Evolutionary genetics and acclimatization in nephrology. *Nat Rev Nephrol*. 2021;17:827–839.
64. Roy BA, Kirchner JW. Evolutionary dynamics of pathogen resistance and tolerance. *Evol Int J Org Evol*. 2000;54:51–63.
65. Kerner G, Patin E, Quintana-Murci L. New insights into human immunity from ancient genomics. *Curr Opin Immunol*. 2021;72:116–125.
66. Petr M, Pääbo S, Kelso J, Vernot B. Limits of long-term selection against Neandertal introgression. *Proc Natl Acad Sci U S A*. 2019;116:1639–1644.
67. McArthur E, Rinker DC, Capra JA. Quantifying the contribution of Neanderthal introgression to the heritability of complex traits. *Nat Commun*. 2021;12:4481.
68. Green RE, Krause J, Briggs AW, et al. A draft sequence of the Neandertal genome. *Science*. 2010;328:710–722.
69. Reich D, Green RE, Kircher M, et al. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*. 2010;468:1053–1060.
70. Disotell TR. Archaic human genomics. *Am J Phys Anthropol*. 2012;149:24–39.
71. Houldcroft CJ, Underdown SJ. Neanderthal genomics suggests a pleistocene time frame for the first epidemiologic transition. *Am J Phys Anthropol*. 2016;160:379–388.