

1 **Genetic adaptation to pathogens and increased risk of**
2 **inflammatory disorders in post-Neolithic Europe**

3

4

5 **Authors:**

6 Gaspard Kerner,^{1*} Anna-Lena Neehus,^{2,3} Laurent Abel,^{2,3,4} Jean-Laurent Casanova,^{2,3,4,5,6}

7 Etienne Patin,^{1,8} Guillaume Laval,^{1,8} and Lluís Quintana-Murci,^{1,7,8,9,*}

8

9 **Affiliations:**

10 ¹Institut Pasteur, Université Paris Cité, CNRS UMR2000, Human Evolutionary Genetics
11 Unit, F-75015 Paris, France

12 ²Laboratory of Human Genetics of Infectious Diseases, INSERM UMR 1163, Necker
13 Hospital for Sick Children, Paris, France.

14 ³University Paris Cité, Imagine Institute, Paris, France.

15 ⁴St. Giles Laboratory of Human Genetics of Infectious Diseases, The Rockefeller University,
16 New York, NY, United States.

17 ⁵Howard Hughes Medical Institute, New York, NY, United States

18 ⁶Department of Pediatrics, Necker Hospital for Sick Children, Paris, France

19 ⁷Collège de France, Chair of Human Genomics and Evolution, F-75005 Paris, France

20 ⁸Senior author

21 ⁹Lead contact

22 *Correspondence: gakerner@pasteur.fr and quintana@pasteur.fr

23

1 **ABSTRACT**

2 Ancient genomics can directly detect human genetic adaptation to environmental cues.
3 However, it remains unclear how pathogens have exerted selective pressures on human
4 genome diversity across different epochs and affected present-day inflammatory disease risk.
5 Here, we use an ancestry-aware approximate Bayesian computation framework to estimate
6 the nature, strength, and time of onset of selection acting on 2,879 ancient and modern
7 European genomes from the last 10,000 years. We found that the bulk of genetic adaptation
8 occurred after the start of the Bronze Age, <4,500 years ago, and was enriched in genes
9 relating to host-pathogen interactions. Furthermore, we detected directional selection acting
10 on specific leukocytic lineages and experimentally demonstrated that the strongest negatively
11 selected immunity gene variant — the lipopolysaccharide-binding protein gene (*LBP*) D283G
12 — is hypomorphic. Finally, our analyses suggest that the risk of inflammatory disorders has
13 progressively increased in post-Neolithic Europeans, partly due to antagonistic pleiotropy
14 following genetic adaptation to pathogens.

15

16 **Keywords:** ancient DNA, immunity, host defense, natural selection, local adaptation,
17 inflammatory disorders, approximate Bayesian computation, antagonistic pleiotropy, *LBP*

1 INTRODUCTION

2 Infectious diseases have been the leading cause of human mortality throughout human
3 history.^{1,2} Population genetics studies have provided support for the notion that pathogens are
4 among the strongest selective forces faced by humans since the discovery in the 1950s that
5 heterozygosity for red blood-cell disorders provides some protection against malaria.³ An
6 increasing number of genes involved in host-pathogen interactions have since been identified
7 as targets of natural selection.⁴⁻⁸ However, major questions remain regarding the evolutionary
8 impact of infectious diseases on human genome diversity. First, little is known about the
9 specific epochs during which humans were most exposed to pathogens and pathogen-
10 mediated selection. It has been suggested that the transition to an agriculture-based lifestyle,
11 which began ~10,000 years ago, increased exposure to deadly microbes, including density-
12 dependent viruses and zoonoses,⁹⁻¹¹ but the archaeological and genetic evidence is scarce, and
13 even challenges this view.^{12,13} Second, the extent to which host defenses, including the
14 engagement of leukocytic lineages, the qualitative and quantitative composition of which is
15 associated with common and rare disorders,¹⁴⁻¹⁷ have been affected by such exposure has not
16 been explored. Third, the rising life expectancy in recent centuries¹⁸ has contributed to an
17 increase in the prevalence of inflammatory and autoimmune disorders, but it has been
18 hypothesized that this increase is also the result of long-term pathogen pressures and
19 antagonistic pleiotropy of the selected gene products.¹⁹⁻²¹

20 The antagonistic pleiotropy hypothesis is supported by the overlap between loci
21 underlying infectious and inflammatory traits^{22,23} and the discovery of several pleiotropic
22 variants conferring protection against infectious diseases and susceptibility to some chronic
23 inflammatory or auto-immune conditions.²²⁻²⁷ Classic examples are the human leukocyte
24 antigen (*HLA*) loci, variants of which are thought to be under pathogen-driven positive
25 selection,²⁸ and to increase considerably the risk of autoimmune disease (e.g. *HLA-B27* and
26 spondyloarthritis and *HLA-DQ8* and diabetes).²⁹⁻³³ Another example is the common *TYK2*
27 P1104A variant, which protects against auto-immune phenotypes (OR: 0.1-0.3) but, in the
28 homozygous state, confers a predisposition to mycobacterium-related infectious diseases (OR
29 >10).^{25,26,34-39} Nevertheless, the evidence that antagonistically pleiotropic variants have been
30 selected in humans remains circumstantial.

31 Detection of a legacy of positive selection (rapid increase in the frequency of a beneficial
32 allele) or negative selection (the removal of a deleterious variant) in human populations has
33 long been limited to statistical inference from patterns of genetic variation in contemporary
34 individuals. However, in recent years, the increasing availability of ancient DNA (aDNA)

1 data has greatly facilitated the study of human genetic adaptation over time. Direct
2 comparisons of past and current allele frequencies have provided new insight into the genes
3 and functions involved in human adaptation to environmental changes following cultural
4 transitions.⁴⁰⁻⁴² A pioneering study based on 230 Eurasian samples across the Holocene period
5 detected 12 positively selected loci associated with diet and skin pigmentation, but also host
6 defense against pathogens.⁴³ Another study explored how the arrival of Europeans in the
7 Americas altered the exposure of Native Americans to new pathogens. Comparisons of
8 genomic data from a Canadian First Nation population dating from before and after the first
9 contact with Europeans showed a recent decrease in the frequency of formerly beneficial
10 *HLA-DQA1* alleles in the native population.⁴⁴ These studies highlight how the analysis of
11 ancient genomes can identify specific genetic variants under selection, but we still lack a
12 comprehensive picture of the selective forces affecting host defense genes during human
13 history and of the times at which these forces operated.

14 Ancient genomics provides us with a unique opportunity to determine whether resistance
15 to infectious diseases and susceptibility to inflammatory diseases have changed in the recent
16 past. Furthermore, aDNA data can be used to detect variants subject to complex models of
17 selection, such as time-dependent selection, including selection on *de novo* or standing
18 variation due to sudden environmental changes (e.g., epidemics). For example, a recent study
19 revealed that negative selection has led to a rapid decrease in the frequency of a 30,000-year-
20 old tuberculosis (TB) risk variant over the last 2,000 years, suggesting that TB has recently
21 imposed a heavy burden in Europeans.³⁵ In this study, we sought to reconstruct the history of
22 host-pathogen interactions by detecting immunity variants affected by natural selection,
23 thereby modulating infectious and/or inflammatory disease risk, over the last 10,000 years. To
24 this end, we explored the strength and timing of both positive and negative selection at the
25 genome-wide scale, by analyzing 2,879 ancient and modern European genomes in an
26 ancestry-aware approximate Bayesian computation (ABC) framework.

27

28 **RESULTS**

29 **Inferring the intensity and onset of positive selection from ancient DNA data**

30 We assembled a genome-wide dataset corresponding to 2,376 ancient and 503 modern
31 individuals of western Eurasian ancestry, and computed allele frequency trajectories for
32 1,233,013 polymorphic sites, over a time transect covering the Neolithic, the Bronze Age, the
33 Iron Age, the Middle Ages and the present (**Methods**). Using simulation-based ABC⁴⁵ and
34 the calculated trajectories, we estimated the selection coefficient (s) and the time of selection

1 onset (T) for each derived allele. We considered a variant to be a candidate for positive
2 selection if its s value was higher than that of 99% of s estimates for simulated neutral
3 variants ($p_{\text{sel}} < 0.01$). We used a demographic model³⁵ accounting for (i) the major migratory
4 movements contributing to the genetic diversity of contemporary Europeans, i.e., the arrival
5 of Anatolian farmers $\sim 8,500$ years ago (ya)^{46,47} and that of populations associated with the
6 Yamnaya culture around $\sim 4,500$ ya,^{48,49} (ii) the uneven sampling across time, with the
7 matching of simulated data on the observed numbers of ancient and modern DNA data, and
8 (iii) ancestry variation across epochs, by matching on observed Anatolian and Pontic Steppe
9 ancestries (**Methods**).

10 Our simulations showed that the estimation of s was highly accurate ($r^2 = 0.93$ between
11 simulated and estimated values; **Figure S1A**), and the power to detect selection from s was
12 $>80\%$ for variants with a present-day frequency $>20\%$, regardless of the timing of selection
13 onset (**Figure S2**). The accuracy of estimation was lower for T ($r^2 = 0.76$; **Figure S1B**), but
14 our approach was nevertheless able to distinguish between variants selected before and after
15 the beginning of the Bronze Age, $\sim 4,500$ ya (F1 score = 0.85; **Figure S1C**). The application
16 of ABC to the observed allele frequency trajectories replicated the 12 loci previously shown
17 to be subject to positive selection in Europe⁴³ (based on the same criterion of ≥ 3 candidate
18 variants for positive selection per locus). These loci included genes associated with host
19 defense (*HLA* and *TLR10-TLR1-TLR6*), skin pigmentation (*SLC45A2* and *GRM5*) and
20 metabolism (*MCM6/LCT*, *SLC22A4* and *FADS1*) (**Figure 1; Table S1**). As further evidence
21 of the accuracy of our approach, the s estimate for the lactase persistence allele (rs4988235)
22 was 8.1% (95% CI: [0.06-0.09]) with selection beginning 6,102 ya (3,150-8,683), as
23 previously reported.^{50,51} Our approach expands previous aDNA studies by not only
24 identifying known and novel candidates for selection, but also providing appropriate
25 estimates of the selection parameters, s and T , over the entire genome.

26

27 **Positive selection has pervasively affected host defense genes in Europeans**

28 We identified the strongest signals of time-dependent positive selection, by focusing on the
29 top 2.5% of loci with the highest proportions of candidate variants, i.e., regions defined on the
30 basis of linkage disequilibrium (LD) groups containing $>7.5\%$ of the candidate positively
31 selected, derived variants ($p_{\text{sel}} < 0.01$, **Methods**). We identified 102 candidate loci for positive
32 selection, 89 of which were non-consecutive. These loci were enriched in gene ontology (GO)
33 categories such as *antigen processing and presentation of endogenous antigen* (Bonferroni
34 corrected $p_{\text{adj}} = 0.01$), *viral life cycle* ($p_{\text{adj}} = 0.03$), and *positive regulation of leukocyte*

1 *activation* ($p_{\text{adj}} = 0.05$), as well as *transport vesicle*, *vesicle membrane*, *luminal side of*
2 *endoplasmic reticulum membrane*, *cell surface* and *vesicle-mediated transport*. These 89 loci
3 were also enriched in a curated list of immunity genes (IGs; **Methods**), whether we
4 considered all candidate genes ($\text{OR} = 1.6$, $p = 8.0 \times 10^{-3}$) or a single gene per locus (28/89
5 loci; $\text{OR} = 1.6$, $p = 0.049$). These genes included the *OAS* cluster (*OAS1-OAS2-OAS3-*
6 *RNASEL*), which modulates the response to infection with RNA viruses,⁵² and genes
7 underlying inborn errors of immunity (IEI), such as *AICDA*, for which biallelic loss-of-
8 function mutations cause antibody deficiencies (OMIM:6055258 and OMIM:605257), and
9 *PLCG2*, for which monoallelic gain-of-function mutations underlie autoinflammatory
10 disorders (OMIM:614878 and OMIM:614468).

11 One of the strongest signals identified was close the *ABO* gene (smallest p_{sel} , **Table S2**),
12 which encodes the ABO blood group system. We found that alleles tagging the A and B blood
13 groups⁵³ were candidates for positive selection ($p_{\text{sel}} < 0.01$; **Table S3**), suggesting that the
14 frequency of individuals with the A, AB and, particularly, B blood groups has increased
15 through positive selection over the last few millennia. The A and B blood groups have been
16 shown to confer limited protection ($\text{OR} > 0.94$) against childhood ear infections,^{54,55} and mild
17 susceptibility ($\text{OR} < 1.13$) to malaria and COVID-19,^{53,56} consistent with the action of
18 balancing selection.⁵⁷ Our results show that the bulk of recent genetic adaptation in Europe
19 has primarily concerned genes involved in the host response to infection, suggesting that
20 pathogens have imposed strong selective pressures over the last few millennia.

21

22 **Genetic adaptation has occurred principally since the Neolithic period**

23 We then explored the time at which positive selection began at the 89 loci, by determining
24 whether selection occurred before or after the start of the Bronze Age (**Figure S1C**). For more
25 than 80% of loci, selection was estimated to have begun after the beginning of the Bronze
26 Age (<4,500 ya) (**Figure 1**). Indeed, the distribution of T estimates for the top 89 variants was
27 not consistent with a uniform distribution across time ($p = 1.3 \times 10^{-4}$), with a much more
28 recent time of selection onset than expected ($T_{\text{mean}} = 3,327$ ya; **Figure S3A; Table S2**). This
29 result cannot be explained by differences in detection power, as our approach has a higher
30 power for variants with selection beginning at earlier timepoints (**Figure S2**). Furthermore,
31 using 100 distributions of T estimates obtained from 89 simulated variants drawn with
32 uniform durations of selection and matched for allele frequency and selection intensity, we
33 checked that the distributions were not skewed towards T estimates postdating the beginning
34 of the Bronze Age ($p_{\text{adj}} > 0.05$ for all 100 distributions; **Methods**). Finally, the ancient and

1 modern DNA datasets were processed differently, so batch effects could have resulted in a
2 sudden change in allele frequency in the last epoch and, therefore, a bias of T towards more
3 recent times. However, when estimating T for aDNA only, we again obtained a significantly
4 larger fraction of estimates $< 4,500$ ya ($n = 67/89$; $p < 10^{-3}$; **Figure S3B**). These analyses
5 support a scenario in which most of the positive selection events have occurred since the
6 Neolithic Period.

7 In light of these results, we reasoned that the arrival of populations from the Pontic-
8 Caspian region $\sim 4,500$ ya may have brought beneficial variants that continued to be subject to
9 positive selection in the resulting admixed population, a process known as adaptive
10 admixture.⁵⁸ In this case, we would expect to see a positive correlation between Pontic Steppe
11 ancestry and the probability of carrying positively selected alleles. However, no significant
12 differences were observed between positively selected alleles and matched controls (p_{adj}
13 > 0.05 for all matched samples; **Methods**), and there was no systematic bias in the ancestry of
14 simulated and observed aDNA samples (**Figure S4A**). Our results therefore rule out adaptive
15 admixture as the main driver of positive selection in post-Neolithic Europe, instead
16 suggesting that there has been selection on standing variation; in other words, most of the
17 alleles that became beneficial after the Bronze Age were already present in Neolithic Europe
18 (**Figure S4B**).

19

20 **Regulatory elements of immunity genes have been a prime substrate of positive selection**

21 We then investigated *in-silico* the functional effects of candidate positively selected variants
22 ($n = 1,846$). We found that these variants were strongly enriched in missense variants (OR =
23 3.8 , $p = 8.8 \times 10^{-21}$) (**Figure 2A**). Furthermore, IGs harboring missense variants were found to
24 be significantly enriched in positively selected variants relative to the rest of the candidate
25 loci ($p = 6.0 \times 10^{-3}$). At the genome-wide scale, we detected 11 missense variants of IGs with
26 signals of positive selection (*MYH9*, *CHAT*, *NLRC5*, *INPP5D*, *SELE*, *C3*, *PSMB8*, *CFB*,
27 *PIK3R1*, *RIPK2*, *CD300E*). Eight of these variants were associated with immune cell traits
28 and infectious or autoimmune disorders, such as blood cell counts, viral warts and type I
29 diabetes, in phenome-wide association studies ($p < 10^{-5}$; **Figure 2B**). However, given that
30 most positively selected variants ($>96\%$) are non-coding, we investigated whether variations
31 of gene expression could account for the observed selection signals. Taking LD and derived
32 allele frequency (DAF) ($r^2 < 0.6$) into account, we found that candidate variants were
33 enriched in *cis*-eQTLs in whole blood, particularly for strong eQTL associations (e.g., OR =
34 2.9 ; $p = 3.3 \times 10^{-21}$ for eQTLs with $p < 10^{-50}$; **Figure 2A**; **Methods**). For example, the top

1 candidate variant for the *OAS* cluster, rs1859332>T, was found to be linked ($r^2 = 0.97$) to the
2 *OAS1* splice QTL rs10774679>C (isoform p46), which is of Neanderthal origin.⁵⁹
3 Interestingly, this isoform has been associated with protection against severe COVID-19 (OR
4 = 0.05),^{60,61} suggesting that past coronavirus-like viruses may have driven selection at this
5 locus.

6 Positively selected variants are also enriched in variants affecting transcription factor (TF)
7 binding (OR = 1.75; $p = 1.6 \times 10^{-6}$; **Figure 2A**), based on allele-specific binding (ASB)
8 events detected in ChIP-Seq data covering 1,025 human TFs and 566 cell types (FDR <
9 0.01)⁶². For example, the positively selected rs3771180>T variant ($s = 0.015$, $T = 2,250$, $p_{\text{sel}} =$
10 0.005) at the *ILIR1* locus disturbs the 7 bp binding motif of *JUND* (OR = 3.07; $p_{\text{adj}} = 5.7 \times$
11 10^{-12}) and is associated with a lower risk of asthma⁶³ and higher neutrophil levels.¹⁴
12 Moreover, 17 of the 265 TFs with at least five ASBs have ASB-associated variants enriched
13 in positive selection signals, and the genes closest to these ASBs are enriched in IGs (OR =
14 1.24; $p = 0.04$). Collectively, these results provide insight into the regulatory mechanisms
15 underlying recent human adaptation in the context of immunity to infection.

16

17 **Directional selection has affected leukocytic lineages since the Bronze Age**

18 We analyzed epigenomic data from ENCODE,⁶⁴ to identify the tissues for which the
19 regulatory elements were the most enriched in positively selected variants. In tests with
20 matched controls and correction for multiple testing, we found an enrichment in positively
21 selected variants at DNase-hypersensitive sites in 24 of the 41 tissues and cell types tested,
22 including monocytes ($p_{\text{adj}} = 6.1 \times 10^{-7}$), tonsils ($p_{\text{adj}} = 2.1 \times 10^{-6}$) and blood ($p_{\text{adj}} = 1.0 \times 10^{-5}$)
23 (**Methods**). These results, together with the enrichment in ASB in candidate variants close to
24 IGs, led us to investigate whether positively selected variants affect hematopoiesis by altering
25 blood cell fractions.

26 Taking LD and DAF into account, we found that variants associated with blood cell
27 composition in GWAS were strongly enriched in positively selected variants (OR = 10, $p =$
28 7.2×10^{-65}) (**Figure 2C**). For each of the 36 hematopoietic traits, we analyzed the frequency
29 trajectories over time of all the trait-increasing alleles ($p_{\text{GWAS}} < 5.0 \times 10^{-8}$) with a polygenic
30 score derived from GWAS data,⁴⁰ within the studied candidate positively selected LD groups
31 (**Figure S5; Methods**). This score integrates tens or hundreds of generally small effect sizes,
32 which we assessed over hundreds of generations. We found that polygenic scores for platelet
33 and reticulocyte counts had decreased significantly over the last 10,000 years, whereas scores
34 for mean platelet volume and mean corpuscular (red blood cell) volume, and the mean mass

1 of hemoglobin per red blood cell, had significantly increased (**Figure 2D**). Likewise,
2 polygenic scores for the proportion of eosinophils among granulocytes were found to have
3 significantly decreased over the study period, whereas the proportions of lymphocytes and
4 monocytes among white blood cells and neutrophil counts had significantly increased (**Figure**
5 **2D**). Importantly, the polygenic scores of seven traits, including traits relating to platelets,
6 reticulocytes, lymphocytes and monocytes, were found to have significantly increased or
7 decreased in post-Neolithic Europe, whereas no trait other than the number of granulocytes
8 changed significantly during the Neolithic (**Figure 2D**), consistent with directional selection
9 acting principally on hematopoietic lineages after the Bronze Age.

10

11 **Joint temporal increase in resistance to infection and the risk of inflammation**

12 It has been suggested that past selection relating to host-pathogen interactions has favored
13 host resistance alleles that, today, increase the prevalence of chronic inflammatory disorders
14 due to antagonistic pleiotropy.^{4,19-21,23,24} We investigated the correlation of genetic
15 susceptibility to inflammatory disorders with resistance to infectious disease, by assembling
16 the results for (i) 40 GWAS of infectious diseases, including severe infections, such as TB,
17 hepatitis B and C, AIDS and COVID-19 (COVID-19 HGI, release 7), and surrogate
18 phenotypes, such as *positive TB test* (based on tuberculin skin test) or *tonsillectomy*, and (ii)
19 30 GWAS of inflammatory/auto-immune disorders, including rheumatoid arthritis (RA),
20 systemic lupus erythematosus (SLE) and inflammatory bowel disease (IBD) (**Table S4**). For
21 both independent and randomly matched variants, we found that variants significantly
22 associated with both infectious and inflammatory traits (i.e., pleiotropic variants; $p_{\text{GWAS}} < 5.0$
23 $\times 10^{-8}$) were more prevalent than expected by chance (OR = 132, $p = 7.4 \times 10^{-28}$ and $p < 10^{-3}$,
24 respectively; **Methods**), suggesting a shared genetic architecture. Furthermore, that the
25 estimated s values for these pleiotropic variants are stronger than for randomly matched
26 variants (Wilcoxon $p = 0.02$) supports their adaptive nature.

27 We used polygenic risk scores (PRS) to explore changes in the frequency of risk alleles
28 for infectious or inflammatory disorders ($p_{\text{GWAS}} < 5 \times 10^{-8}$) over time (**Methods**). We found
29 that the PRS for the merged set of all inflammatory/autoimmune disorders had significantly
30 increased over the last 10,000 years, whereas that for infectious diseases had significantly
31 decreased (**Figure 3A**). Taking into account the coverage, location and ancestry of the
32 samples, the risk of Crohn's disease (CD) or of IBD generally has increased significantly,
33 whereas that of severe COVID-19 has decreased significantly, essentially since the Neolithic
34 period (**Figure 3A-B**). Given the recent nature of the emergence of SARS-CoV-2, the

1 infectious agent of COVID-19, this finding suggests that other related pathogens have exerted
2 selection pressure in recent European history.

3 We then searched for the specific genetic variants making the largest contribution to the
4 changes over time in the risks of IBD and severe COVID-19. Taking multiple testing into
5 account, we found that four risk alleles for IBD and one protective allele for COVID-19 had
6 significantly increased in frequency over time (**Figure 3C** and **Table S5**). The risk alleles
7 concerned were the IBD- and/or CD-associated rs1456896>T, rs2188962>T, rs11066188>A
8 and rs492602>G alleles, and the COVID-19-associated rs10774679>C allele. The first three
9 of these variants increase the expression of *IRF1*, *SH2B3*, and *IKZF1*, respectively, in blood
10 (rs2188962, $s = 0.022$, $T = 8,075$; rs11066188, $s = 0.016$, $T = 4,045$; rs1456896>C, $s = 0.007$,
11 $T = 3,264$; **Figure 3D**). These genes have been shown to protect against several infectious
12 agents,^{27,65,66} suggesting that increases in their expression may be beneficial, even if they
13 slightly increase autoimmunity risk. The fourth IBD-associated risk allele, rs492602>G ($s =$
14 0.0015 , $T = 6,823$), is a *FUT2* variant in complete LD ($r^2 = 0.99$) with the null allele
15 rs601338>A, which confers monogenic resistance to infections with intestinal norovirus⁶⁷ and
16 respiratory viruses.⁶⁸ Finally, the COVID-19-associated variant rs10774679>C ($s = 0.013$, $T =$
17 $3,617$) is linked ($r^2 = 0.9$) to the *OAS1* splice variant rs10774671>G, which is also associated
18 with slightly higher susceptibility to IBD (OR = 1.08; $p = 1.4 \times 10^{-3}$;
19 FINNGEN_R5_K11_IBD). Overall, these analyses provide support for a role of selection in
20 increasing autoimmune disease risk over recent millennia, particularly for gastrointestinal
21 inflammatory traits, probably due to antagonistic pleiotropy.

22

23 **Searching for the footprints of time-dependent negative selection**

24 Given the observed links between candidate selected variants and host defense, we then used
25 our approach to identify candidate variants increasing infectious disease risk. Specifically, we
26 searched for variants of genes involved in host-pathogen interactions that have been under
27 time-dependent negative selection, that is, variants that have become deleterious since the
28 Neolithic. A typical example is *TYK2* P1104A, which we have previously shown to underlie
29 clinical TB^{34,39} and to have evolved under negative selection over the last 2,000 years,
30 probably due to an increase in the pressure imposed by *Mycobacterium tuberculosis*.³⁵ We
31 removed the 89 positively selected candidate regions, and loci presenting positive selection
32 signals in contemporary Europeans (**Methods**), and then searched for candidate deleterious
33 variants displaying significant decreases in DAF in the rest of the genome (93% of all LD
34 groups).

1 We first checked that the s and T estimates for alleles under negative selection were as
2 accurate as those for positively selected alleles (**Figure S1D-E**). We confirmed that variants
3 with a low DAF (<5% across epochs) were strongly enriched (OR = 4.8, $p = 2.7 \times 10^{-210}$) in
4 candidate negatively selected variants, as opposed to positively selected alleles (OR = 0.64, p
5 = 4.2×10^{-7}). We found that negative selection had a broader impact across the genome; 25%
6 of LD groups harbor candidate negatively selected variants ($p_{\text{sel}} < 10^{-4}$), whereas only 3%
7 harbor positively selected alleles (**Figures 1A and 4A**). By focusing on negatively selected
8 missense variants at conserved positions (GERP score >4; $n = 50$; **Table S6**), we found that
9 the vast majority of these variants (41 out of the 50; 82%), like the positively selected
10 variants, had an estimated onset of selection <4,500 ya (**Figure 4**). Focusing on immunity
11 genes, we identified six negatively selected missense variants: *LBP* D283G (rs2232607),
12 *TNRC6A* P788S (rs3803716), *CIS* R119H (rs12146727), *IL23R* R381Q (rs11209026), *TLR3*
13 L412F (rs3775291) and *TYK2* P1104A (rs34536443). These time-series analyses provide
14 compelling evidence for negative selection against variants of host defense genes in recent
15 millennia.

16

17 **The D283G variant is hypomorphic and impairs LBP expression**

18 We investigated the functional impact of negatively selected variants, by focusing on the *LBP*
19 D283G variant, which had the strongest negative selection signal ($s = -0.018$, $T = 3,084$). *LBP*
20 encodes the lipopolysaccharide (LPS)-binding protein, which senses LPS — a major
21 component of the outer membrane of Gram-negative bacteria — and initiates immune
22 responses that prime host defense mechanisms against further infection.⁶⁹ Previous studies
23 have shown that a common *LBP* variant (P333L; 8% in Europe) reducing both LPS binding
24 capacity and cytokine response is associated with greater mortality due to sepsis and
25 pneumonia.⁷⁰ We investigated whether *LBP* D283G, which has decreased in frequency from
26 6% to 1.2% over the last 5,000 years, had a similar biochemical impact on LBP function,
27 potentially decreasing host fitness for fighting bacterial infections.

28 We transiently transfected HEK 293T cells with plasmids encoding wild-type (WT) or
29 mutant *LBP* cDNAs, including the previously reported P333L variant. All constructs were C-
30 terminally tagged with DDK and equal transfection efficiencies were confirmed by RT-qPCR
31 (**Figure 5A**). Protein levels were analyzed by western blot on whole-cell lysates and cell
32 culture supernatants. The WT and mutant proteins were detected at the expected molecular
33 weight (~60 kDa) with an *LBP*-specific antibody and an antibody against the C-terminal tag
34 (**Figure 5B**). However, D283G levels were lower than WT protein levels, and P333L was not

1 detectable. The low levels of D283G protein were confirmed by ELISA (**Figure 5C**),
2 suggesting that this variant may be less stable when secreted. Finally, we showed that the
3 D283G and WT proteins had similar LPS binding capacities (**Figure 5D**), demonstrating that
4 this variant does not affect this function. These findings suggest that D283G alters LBP
5 stability, resulting in lower LBP protein levels, potentially impairing the host response to
6 bacterial exposure.

7

8 **DISCUSSION**

9 Based on a time-series of human aDNA data and extensive computer simulations, we have
10 delineated the genomic regions under the strongest selective pressure over the last 10,000
11 years of European history. We found that host defense genes are enriched in positive selection
12 signals, with selected variants primarily involved in regulatory functions. We also found that
13 directional selection has operated on at least four leukocytic traits ($|r| < 0.6$; **Figure S5**) in
14 recent millennia. Leukocytic lineages have undergone positive selection, with the exception
15 of eosinophils, which have decreased in proportion among granulocytes, possibly reflecting
16 an evolutionary trade-off in favor of neutrophils, consistent with the apparently less essential
17 role of eosinophils in immunity to infection.⁷¹ Conversely, non-leukocytic lineages, such as
18 red blood cells, seem to have undergone changes in efficacy, given the observed decrease in
19 the number of reticulocytes but the increase in their mean size and in the concentration of
20 hemoglobin per red blood cell. These results suggest that recent positive selection has targeted
21 the regulatory machinery underlying immune cell variation, possibly as a result of temporal
22 changes in pathogen exposure.^{35,72,73}

23 The estimated times of selection onset highlight the importance of the post-Neolithic
24 period in the adaptive history of Europeans, as most selection events — both positive and
25 negative — postdate the beginning of the Bronze Age (<4,500 ya). Furthermore, our results
26 support a history of selection preferentially targeting variants that were already segregating in
27 Europe before the arrival of Pontic-Caspian groups.⁷⁴ The increase in adaptation following the
28 Neolithic may be due to the population growth that followed the ‘Neolithic decline’,⁷⁵ with
29 higher selection efficacy.⁷⁶ Alternatively, selection pressures may have increased during the
30 Bronze Age; the expansion of urban communities, greater human mobility, animal
31 husbandry⁷⁷ and environmental changes⁷⁸ may have favored the spread of epidemics, such as
32 plague, as suggested by archeological and ancient microbial data.^{13,72,73}

1 Our analyses also provide several lines of evidence supporting a significant contribution
2 of antagonistic pleiotropy to the emergence of modern chronic diseases. First, selection over
3 the last 10,000 years, particularly since the beginning of the Bronze Age, has led to a higher
4 genetic risk of inflammatory gastrointestinal disorders⁷⁹. Second, the main risk variants for
5 IBD/CD are located close to key immunity genes (*IRF1*, *IKZF1*, *FUT2* and *SH2B3*), for
6 which monogenic lesions confer susceptibility or resistance to infectious diseases.^{27,65,67}
7 Third, we found that pleiotropic variants underlying infectious and inflammatory phenotypes
8 have been primary targets of positive selection in recent millennia.

9 Finally, our study highlights the value of adopting an evolutionary genomics approach,
10 not only to determine the legacy of past epidemics in human genome diversity, but also to
11 identify candidate negatively selected variants potentially increasing infectious disease risk.
12 Candidate negatively selected variants included the hypomorphic *LBP* D283G and TB-risk
13 *TYK2* P1104A variants, together with *TLR3* L412F, which is associated with mild protection
14 against autoimmune thyroid disease (OR = 0.93; $p = 7.0 \times 10^{-12}$)⁸⁰ but with a modest increase
15 in the risk of severe COVID-19 pneumonia,⁸¹ and *IL23R* R381Q, which increases IBD risk
16 (OR = 1.93)⁸². Autosomal recessive or dominant *TLR3* deficiency underlies viral diseases of
17 the brain and lungs,^{83,84} whereas autosomal recessive *IL23R* deficiency underlies clinical
18 disease caused by weakly virulent fungi and mycobacteria.⁸⁵ It is tempting to speculate that
19 *TLR3* L412F and *IL23R* R381Q confer predispositions to viral and mycobacterial/fungal
20 infectious diseases, respectively.⁸⁶ Detailed functional characterization of these variants in
21 appropriate cell types, and the detection of new candidates from high-quality ancient
22 genomes, will provide insight into the contribution of other variants to infectious disease
23 susceptibility and severity.

24 In conclusion, this study shows that directional selection has targeted host defense genes
25 over the last ten millennia of European history, particularly since the start of the Bronze Age,
26 probably contributing to present-day disparities in susceptibility to infectious and
27 inflammatory disorders.

28

29 **Limitations of the study**

30 This study assumes population continuity since the Bronze Age, but fine-scale migrations
31 have probably affected the European gene pool in modern times. Nevertheless, ancient DNA
32 data suggest that no major population turnover has occurred over the last three millennia,^{43,87-}
33 ⁸⁹ indicating that our results should be largely robust to unmodeled discontinuity.

34 Furthermore, the spatially heterogeneous nature of the aDNA dataset used here reduces the

1 power to detect loci undergoing local adaptation. Likewise, the array used to generate ancient
2 genomes, originally designed for demographic purposes, does not capture most rare variants,
3 particularly those that became very rare, or even extinct, due to negative selection, providing
4 a partial view of the evolutionary past of west Eurasians. Finally, given that environmental
5 exposures may have differed in nature and intensity across Europe — during and beyond the
6 studied timeframe — larger and denser sequence-based aDNA datasets are required to
7 replicate our results and detect more subtle, region-specific selection events.
8

1 **METHODS**

2 **Ancient DNA analysis of the 1240k capture dataset**

3 We analyzed 2,632 aDNA genomes (**Table S7**) (i) originating from burial sites in western
4 Eurasia ($-9 < \text{longitude } (^\circ) < 42.5$ and $36 < \text{latitude } (^\circ) < 70.1$), (ii) including genotypic
5 information for 1,233,013 polymorphic sites, and (iii) retrieved from the V44.3: January 2021
6 release at [https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-](https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data)
7 [genotypes-present-day-and-ancient-dna-data](https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data). All individuals were treated as pseudo-haploid
8 (i.e., hemizygotes for either the reference or the alternative allele). Using READ,⁹⁰ we
9 manually removed 95 samples that were either annotated as duplicated or found to correspond
10 to first-degree relatives of at least one other individual with higher coverage in the dataset.

11

12 **Variant filtering**

13 We removed variants for which the derived allele was present in fewer than five aDNA
14 samples from the analysis. We also excluded variants absent from gnomAD v2.1.1 (ref⁹¹) and
15 the 1,000 Genomes Project⁹² or for which the ancestral allele was not annotated in the 1,000
16 Genomes Project⁹². Finally, we included only variants with the ‘PASS’ flag in gnomAD
17 v2.1.1. Other filters based on frequency were applied.

18 We controlled for potential artifacts due to undetected technical problems in this ‘capture’
19 dataset, by comparing the allele frequencies obtained with those from shotgun sequencing
20 data. We processed 952 available published aDNA shotgun data (FASTQ files) with a
21 published pipeline⁸⁸ to obtain pseudohaploid data comparable with those of the capture
22 dataset. We combined all samples and retrieved 4,620,071 variants after filtering, accounting
23 for 95% of those present in the capture dataset. Strikingly, nine of the top 10 variants in the
24 capture dataset, ranked by *s* value, had a frequency trajectory different from that of the
25 shotgun dataset, with a single, strong change in frequency in the present generation,
26 suggesting a misestimation of frequency in either the ancient or modern dataset. For these
27 variants, we found that, when both genotypes were called in the capture and shotgun datasets,
28 they were consistent between datasets, but there was a high percentage of missingness in the
29 capture dataset when shotgun-sequenced individuals were called as homozygotes for the
30 derived allele (**Table S8**). This was not the case for genotypes called as homozygous for the
31 ancestral allele in the shotgun dataset, generating a frequency bias at population level. While
32 searching for a pattern common to these variants, we found that only 3% of all variants in the
33 capture dataset were within or close to an indel (within 20 bp) with a frequency $>5\%$ in
34 European populations, 90% and 44% of our top 10 and top 100 variants from the capture

1 dataset, ranked by s value, respectively, were located in such a position. We therefore
2 conservatively removed all SNPs close to (<50 bp) an indel with a frequency >5% in the
3 European population from the capture dataset, and variants for which the increase (or
4 decrease) in frequency over the last 1,000 years exceeded 10% (a highly unlikely scenario),
5 thereby excluding 28,493 variants in total. Our final dataset consisted of 933,781 SNPs from
6 2,537 ancient genomes.

8 **Allele frequency trajectories across epochs**

9 For each SNP, we used the ancient genomes and 503 modern European genomes from the
10 1,000 Genomes Project (CEU, GBR, FIN, TSI and IBR populations),⁹² to compute time-series
11 data corresponding to the trajectory of allele frequencies across various time transects. The
12 ancient genomes were grouped by well-characterized historical periods.³⁵ We considered the
13 Neolithic (8,500-5,000 ya; $n=729$ for the capture dataset), the Bronze Age (5,000-2,500 ya; n
14 = 893), the Iron Age (2,500-1,250 ya; $n = 319$) and the Middle Ages (1,250-750 ya; $n = 435$).
15 We excluded 190 samples dated to before 8,500 ya because of the small number of
16 individuals for such a large time period. The genetic information for the variants in ancient
17 and modern genomes was summarized in a five-dimensional frequency vector.

19 **Ancestry estimation**

20 We used factor analysis⁹³ to estimate ancestry proportions at the individual level. We used a
21 merged dataset consisting of 143,081 SNPs for 363 present-day European individuals from
22 the 1,000 Genomes Project (IBS, TSI, GBR and FIN, available at the V42.4: March 1 2020
23 release) and all ancient samples (see above). Before ancestry analysis, samples were imputed
24 with the 'LEA' R v3.6 package, as recommended elsewhere.⁹³ We imputed individuals with
25 high levels of genome coverage (> 795,475 SNPs covered in the 1240k dataset; $n = 421$) first,
26 to prevent bias due to the inclusion of low-coverage samples. The remaining individuals were
27 then imputed one-by-one and added to the imputed set of high-coverage samples. Factors are
28 interpreted as the principal components of principal component analysis (PCA), but with
29 temporal correction for present-day and ancient samples. We set the drift parameter so as to
30 remove the effect of time on the K^{th} factor ($K = 3$ here), where K is the number of ancestral
31 groups considered. As source populations, we used 41 samples annotated as Mesolithic
32 hunter-gatherers, 25 as Anatolian farmers, and 17 as Yamnaya herders.

34 **Demographic model**

1 The model used includes demographic parameters, such as divergence times, effective
2 population sizes, migration rates and exponential growth rates of continental populations
3 (ancestral African population, and West and East Eurasians)³⁵. This model also accounts for
4 the two major migratory movements that shaped the genetic diversity of current Europeans:
5 the arrival of Anatolian farmers in Europe around 8,500 ya, admixing with the local
6 Mesolithic hunter-gatherers, and that of populations of Yamnaya culture from the north of the
7 Caucasus around 4,500 ya. We accounted for the over/under representation of a particular
8 ancestry at selected epochs, by matching simulated ancestry proportions to the mean ancestry
9 proportions of the observed samples, as described in ref.³⁵, and by performing the analyses at
10 the SNP level taking into account their coverage in each individual.

11

12 **Forward-in-time simulations**

13 Computer simulations of an allele evolving under the aforementioned demographic model
14 were performed with SLiM 3 (ref.⁹⁴), as described elsewhere.³⁵ Briefly, for each simulation,
15 three main evolutionary parameters were randomly sampled from uniform prior distributions:
16 the age of the mutation, the time of selection onset T ($1,000 < T < 10,000$ ya) and selection
17 strength, as measured by the selection coefficient s ($-0.05 < s < 0$ for negative selection and 0
18 $< s < 0.1$ for positive selection), under an additive model ($h = 0.5$)⁹⁴. The age of the mutation
19 was randomly selected over the last million years of human evolution ($1,000$ ya $<$ Age $<$
20 $1,000,000$ ya) and was defined as the point at which the mutation was introduced into the
21 model in a randomly chosen population. Each observed ancient genome was randomly
22 sampled from simulated diploid individuals at the generation corresponding to its calibrated,
23 radiocarbon-based age. For each sampled individual and each polymorphic site, we randomly
24 sampled one allele to generate pseudohaploid data, mirroring the observed pseudohaploid
25 aDNA data used. Simulated present-day European individuals were randomly drawn from the
26 last generation of the simulated population.

27

28 **ABC estimation**

29 We applied the ABC approach⁴⁵ to each of the genetic variants studied, as previously used to
30 estimate the age (T_{age}), strength (s) and the time of onset (T) of selection for *TYK2* P1104A
31 (ref.³⁵). Parameter estimates were obtained from 400,000 simulations (for positive or negative
32 selection) with underlying parameters drawn from predefined uniform prior distributions.
33 Parameters were estimated from computer simulations best fitting the observed time series
34 data for allele frequencies. Simulated and empirical time series data were described by a

1 vector of K allele frequencies over K epochs, used as the summary statistics to fit the
2 observed data. For each parameter, posterior distributions, point estimates (i.e., posterior
3 mode) and the 95% CIs were obtained from the parameter values of the 1,000 simulations
4 with simulated summary statistics the closest to the empirical ones ('abc' R package, method
5 = "Loclinear").

7 **Detection of selection**

8 We used the ABC estimates of the selection coefficient and their 95% CI (under a negative or
9 positive selection model) to detect selection acting on specific genetic variants. The empirical
10 threshold for rejecting neutrality (i.e., type I error estimation) was determined by simulating
11 ~500,000 neutral alleles ($s = 0$) evolving under the same demographic model. We then
12 estimated selection coefficients and their 95% CIs for each simulated neutral variant to obtain
13 the distribution of s under neutrality (i.e., the null distribution). Simulated neutral SNPs were
14 resampled such that the simulated and observed allele frequency spectra were identical.
15 Finally, we determined empirical thresholds at the 1% nominal level, by calculating the 99th
16 quantile (Q_{99}) of the resulting s distributions. Rather than using the distribution of s to
17 determine these thresholds, we used the empirical distribution of the lower bound of the 95%
18 CI of the selection coefficient (s_l), as a more conservative approach to providing empirical p
19 values (p_{sel}) for each SNP. This approach, like all methods for detecting selection at the
20 variant level, is not designed to infer the distribution of fitness effects (DFE)⁹⁵ due to the loss
21 of true selection signals at the high conservative thresholds used to detect robust selection
22 candidates.

24 **Empirical p value computation**

25 The significance threshold varied with variant frequency, as expected given that low-
26 frequency variants are less identifiable in terms of selection than more frequent variants. We
27 therefore excluded the variants with the lowest frequencies (the lower bound of the CI of the
28 allele frequency <2.5%), for which estimation accuracy was poor. We normalized the data by
29 calculating the null distribution of the lower bound of the confidence interval of the selection
30 coefficient by frequency bin, to obtain bin-dependent significance thresholds. For the analysis
31 of negative selection, we used the following bins: [0.025-0.05]; [0.05-0.1]; [0.1-0.2] and [0.2-
32 0.8], whereas, for positive selection, we used the following bins: [0.025-0.2]; [0.2-0.6] and
33 [0.6-0.8]. We identified the bin to which a variant belonged by calculating, for each variant,
34 the CI for allele frequency estimation at each epoch, according to an approximation to the

1 normal distribution of the 95% binomial proportion CI. We obtained the maximum for the
2 lower bound of these CI for each SNP. Finally, if this maximum lay between 2.5% and 5%,
3 the variant was considered to belong to the bin [0.025-0.05]. The same rationale was used for
4 the rest of the bins. We excluded from the analysis of positive selection, any variant for which
5 the minimum higher bound of the 95% CI of the DAF was >80%, as such variants poorly
6 matched the simulated data. We ended up with 21,129 candidate variants for positive
7 selection, and 27,591 for negative selection ($p_{\text{sel}} < 0.01$).

8 Finally, as the lowest level of empirical significance depends on the number of neutral
9 simulations in each frequency bin, we approximated the empirical null distribution with a
10 known theoretical distribution, to improve discrimination between very small p values. Given
11 the shape of the empirical null distribution, we compared the null distribution to a gamma, a
12 beta and a lognormal distribution, for which parameters were estimated with a maximum
13 likelihood approach (R packages ‘fitdistrplus’ and ‘EnvStats’ (v. 3.6.0)). We generated a
14 Cullen and Frey graph (kurtosis vs. skewness) with the R package ‘fitdistrplus’ (v. 3.6.0), to
15 distinguish between our options, and obtained p values for the beta distribution that best
16 adjusted the null empirical distribution.

17

18 **Time of selection onset for positively selected loci**

19 We evaluated the shape of the distribution of the onset of selection estimated for the top 89
20 positively selected variants (**Figure S3A** and **Table S2**, with mean T estimate of 3,327 ya), by
21 simulating sets of 89 independent variants matching the allele frequency and the selection
22 coefficient of the most significant variant for each positively selected LD group, and the
23 estimated onset of selection. We investigated whether the frequency trajectories based on both
24 ancient and modern DNA samples resulted in biased T estimations, due to differences in
25 genotype calling between datasets, by re-estimating T values for the variant with the smallest
26 p_{sel} at each of the 89 candidate positively selected loci using frequencies from aDNA only
27 (**Figure S3B**). We thus repeated the ABC estimation for frequency trajectories, but we
28 excluded the last epoch corresponding to current frequencies.

29 Finally, we assessed the contribution of adaptive admixture, by averaging the Pontic
30 Steppe proportion of all the carriers of each of the selected alleles (for different p_{sel}
31 thresholds) or that of 89,000 random alleles (**Figure S4B**). We also checked that Pontic
32 Steppe ancestry was similar between the carriers of the variant with smallest p_{sel} at each of the
33 89 candidate loci and the simulated carriers of the 1,000 simulated variants used for each
34 estimation of the evolutionary parameters of such variants (**Figure S4A**).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34

LD grouping

We took LD into account, by using a LD map for Europeans constructed from whole-genome sequence data.⁹⁶ This metric map displays additive “linkage disequilibrium unit” (LDU) distances, which can be used to define genetic units in which variants are in strong LD. Genomic windows were then defined as non-overlapping regions of 15 LDUs, referred to as *LD groups*. This grouping generated genomic units with a mean size of 660 kb, consistent with previous studies.⁴³

Enrichment analyses for positively selected loci

We calculated enrichment for genes (IGs or GO annotations), variants (eQTLs, ASBs, GWAS variants) and variant annotations (e.g., “missense”), with 2 x 2 contingency tables with two predefined categories (e.g. variants with $p_{\text{sel}} < 0.01$ vs. variants with $p_{\text{sel}} \geq 0.01$; missense variants vs. others), from which we calculated ORs, 95% CIs and Fisher’s exact test p values (for cells with counts < 20) or Chi-squared p values, with the “oddsratio” function of the R (v 3.6.0) package “epitools”. We used independent variants to determine enrichment, by pruning variants in LD with the plink command `--indep-pairwise 100 10 0.6 --maf 0.01`, on our aDNA dataset, thus removing variants with $r^2 > 0.6$ in 100 kb windows, using sliding windows of 10 variants. For the *HLA* region, considered to lie between hg19 coordinates 27,298,200 and 34,036,446 of chromosome 6, we used a more conservative LD pruning method considering 1,000 kb rather than 100 kb windows (plink command `--indep-pairwise 1000 100 0.6 --maf 0.01`), for variants with a minor allele frequency (MAF) $> 1\%$. Where indicated, we also matched the DAF distribution of the pruned dataset to that of the studied group of variants (e.g., eQTLs or GWAS variants), using 5% frequency bins.

We calculated enrichment in IGs⁹⁷ or GO annotations,⁹⁸ by considering, for each LD group with > 9 variants (4,134 LD groups), a binary variable, indicating whether the locus included an immunity gene or a gene with a given GO annotation, respectively. This was done to eliminate spurious enrichments due to the presence of gene clusters in a given LD group. For eQTLs analyses, we used data from a meta-analysis of whole-blood *cis*-eQTLs.⁹⁹ We also used ENCODE data⁶⁴ and DNase hypersensitive sites¹⁰⁰ to estimate enrichment in positively selected variants for ENCODE tissues. Finally, for the study of hematopoietic traits, we retrieved GWAS data for counts or proportions of different blood cell types (36 hematopoietic traits) from UK Biobank and INTERVAL study data.¹⁴

1 **Calculations of polygenic scores**

2 For the analysis of inflammatory and infectious traits, we calculated genetic values for each
3 ancient individual as proposed elsewhere.⁴⁰ Specifically, we weighted the presence/absence
4 status of the most significant GWAS Bonferroni-significant variant ($p < 5.0 \times 10^{-8}$) by the
5 GWAS-estimated effect size, for each LD group. Coverage was variable across ancient
6 samples, and some SNPs were not present in all samples. We accounted for missing
7 information by calculating a weighted proportion in which the estimated score was divided by
8 the maximum possible score given the SNPs present in the sample. We then used weighted
9 (on coverage) linear regression to investigate the association between polygenic score and
10 ancient sample age. We included ancestry as a covariate in the model, by including the first
11 four Factor components, and the geographic location of each sample (latitude and longitude).
12 We compared the full model to a nested model without sample age, by performing a
13 likelihood ratio test [R anova(nested model, full model, test = 'Chisq')] to obtain a p value.
14 For the stratified analysis, we divided the capture dataset into mutually exclusive ancestry
15 groups. We categorized individuals as western European hunter-gatherers, Anatolian farmers,
16 or Pontic steppe herders if they carried over 75% of the estimated respective ancestry
17 component (for steppe individuals, we also required the individual to be <5,000 years old).
18 We also conducted this analysis for individuals classified as being from before or after the
19 beginning of the Bronze Age.

20 For the analyses of hematological traits, we used the same method, summing alleles
21 increasing the count or the proportion of the studied hematological trait, weighted by their
22 effect. Genetic correlations between all 36 traits were estimated using 'ldsc' (ref.¹⁰¹) (**Figure**
23 **S5**).

24 For all analyses, we checked that, despite the consideration of only one variant per LD
25 group, none of the variants were in LD with each other. We acknowledge that the PRS
26 obtained are proxies for the actual PRS across time because, variants that have reached
27 fixation may not be detectable by GWAS.

28

29 **Overlap between infectious and autoimmune GWAS variants**

30 We looked for significant overlaps between aDNA variants tagging lead GWAS SNPs for
31 infectious and autoimmune/inflammatory diseases or traits, by retrieving summary statistics
32 for (i) 40 GWAS of infectious diseases and (ii) 30 GWAS of inflammatory/autoimmune
33 phenotypes (**Table S4**). We identified lead infectious or autoimmune disease-associated SNPs
34 by retaining the variant with the highest GWAS effect size (OR) in consecutive 200 kb

1 genomic windows. We then used aDNA variants to tag ($r^2 > 0.6$) the lead GWAS-significant
2 variants (usually absent from the aDNA array) with the plink command: `plink --show-tags`
3 `aDNA_variant_list --list-all --tag-r2 0.6 --tag-kb 1000`. We obtained a list of aDNA variants
4 tagging infectious and/or autoimmune disease-associated lead SNPs. We then performed
5 1,000 samplings of random SNPs, matched for DAF and number of LD groups, from aDNA
6 variants tagging either infectious or autoimmune traits. For each replicate, we calculated the
7 number of overlapping variants and found that none was higher than the observed overlap (p
8 $< 10^{-3}$). We also used a different approach to test for an enrichment in pleiotropic variants,
9 based on the pruning of aDNA-tagged SNPs ($r^2 = 0.6$ and windows of 1 Mb) and the
10 calculation of a classic OR for the resulting list of independent tagged SNPs. We assessed the
11 enrichment in selection signals in the observed overlapping variants, by performing a
12 Wilcoxon test to compare the s distribution of the observed overlapping variants to that of the
13 1,000 randomly sampled controls. Finally, the PRS and respective p and beta values of the
14 regression model were obtained for the lead GWAS SNPs present in the aDNA genotyping
15 array, either for all infectious or for all autoimmune diseases considered together. PRS and the
16 corresponding p and beta values for CD, IBD, COVID A2 and COVID B2 were obtained for
17 all GWAS-significant variants, because this information was available for such phenotypes.
18 Of note, whereas no significant genetic correlations were found between non-overlapping
19 infectious phenotypes (e.g. COVID_A2 and COVID B2 were considered as overlapping) nor
20 between non-overlapping inflammatory phenotypes, formal testing for genetic correlation, as
21 done for hematopoietic traits, was not possible since for the majority of traits we only had
22 access to the lead GWAS-significant SNPs.

23

24 **Detection of negatively selected variants**

25 We excluded positively selected variants from our list of candidate negatively selected
26 variants, by calculating four haplotype-based statistics used to detect recent positive selection:
27 iHS^{102} , iHH^{103} , nSL^{104} and $DIND^{105}$. This was done for all SNPs with a DAF > 0.2 in
28 Europeans of the 1,000 Genomes Project,⁹² with *selink* (<http://github.com/h-e-g/selink>) and a
29 100 kb genomic window. We computed the 99% quantiles (Q_{99}) of each of the four
30 distributions and the proportion of SNPs with scores higher than the respective Q_{99} , in 200 kb
31 sliding windows. Only windows with > 9 variants were considered. We then retrieved all the
32 genomic windows enriched for at least one of the four selection statistics, i.e., those with
33 proportions on the top 1% of at least one of the four distributions. LD groups overlapping,

1 completely or partially, at least one of these windows were then removed from the analysis of
2 negative selection.

3

4 **Site-directed mutagenesis and transient transfection for LBP**

5 Site-directed mutagenesis was performed on the LBP-WT pCMV6 plasmid (#RC221961,
6 OriGene) with appropriate primers (**Table S9**) and the Pfu Ultra II Fusion HS DNA
7 (#600674, Agilent) polymerase, followed by digestion with *DpnI* (#R0176L, New England
8 Biolabs). Plasmids were amplified in NEB-10 β competent *E. coli* (#C3019H, New England
9 Biolabs) and purified with the HiSpeed Plasmid Maxi Kit (#12663, Qiagen). Transient
10 transfection was carried out in HEK293T cells transfected with 1 μ g of plasmid DNA in the
11 presence of X-tremeGene9 DNA transfection reagent (#6365809001, Merck), according to
12 the manufacturer's instructions. Transfected cells were cultured at 37°C, under an atmosphere
13 containing 5% CO₂, in Dulbecco's modified Eagle medium (DMEM) supplemented with 10%
14 fetal bovine serum. After 48 h, supernatants and whole-cell lysates were collected for
15 subsequent experiments.

16

17 **RNA isolation and RT-qPCR**

18 Total RNA was extracted with the Quick-RNA MicroPrep Kit (#R1051, Zymo) according to
19 the manufacturer's instructions. Residual genomic DNA was removed by in-column DNase I
20 digestion. We reverse-transcribed 1 μ g of RNA with the High-Capacity RNA-to-cDNA Kit
21 (#4387406, Thermo Fisher Scientific), and performed quantitative qPCR with PowerUp
22 SYBR Green Master Mix (#A25742, Thermo Fisher Scientific) and the ViiA7 system
23 (Thermo Fisher Scientific) with primers for LBP (PrimerBank ID 31652248c1 and
24 31652248c2) and GAPDH (PrimerBank ID 378404907c1) obtained from the PrimerBank
25 database.¹⁰⁶ Normalization of *LBP* mRNA was performed for each sample with *GAPDH*
26 (Δ cT) and values are expressed as $2^{-\Delta$ cT}.

27

28 **Protein isolation and western blotting**

29 Whole-cell protein lysates were extracted in modified radioimmunoprecipitation assay buffer
30 supplemented with protease inhibitors (#5892970001, Merck) and phosphatase inhibitor
31 cocktail (#4906837001, Merck), 0.1 mM dithiothreitol (DTT; Life Technologies), and 1 mM
32 PMSF (#10837091001, Merck). Protein extracts and supernatants were resolved by
33 electrophoresis in Criterion TGX 10% precast gels (Bio-Rad), with the resulting bands
34 transferred onto PVDF membranes (#1704157, Bio-Rad) with the Transblot turbo system

1 (Bio-Rad). Membranes were probed by incubation for 1 hour at room temperature with
2 antibodies against LBP (#AF870-SP, R&D Systems, 1:2,000), DDK (#A8592, Merck,
3 1:10,000) and GAPDH (#sc-47724, Santa Cruz Biotechnology, 1:5,000). Proteins were
4 detected by chemiluminescence with Clarity Western ECL substrate (#1705061, Bio-Rad)
5 reagents.

6

7 **ELISA for LBP and bacterial ligand binding**

8 Supernatants from transfected HEK293T cells were analyzed for their LBP content by Human
9 LBP DuoSet ELISA (#DY870-05, R&D Systems) according to the manufacturer's
10 instructions. Bacterial ligand binding was assessed on microtiter plates coated with 30 µg/mL
11 LPS derived from *E. coli* O111:B4 (#LPS25, Merck) or Pam₂CSK₄ (#tlrl-pm2s-1, InvivoGen)
12 in 100 mM Na₂CO₃ (pH 9.6) for 18 hours at 4°C. The plates were blocked by incubation with
13 0.005% Tween and 1% bovine serum albumin in PBS for 1 hour at room temperature. Plates
14 were then incubated with cell supernatants at various concentrations. Bound LBP was
15 detected with the detection antibodies from the Human LBP DuoSet ELISA kit. Absorbance
16 was read at 450 nm with a Victor X4 plate reader (Perkin Elmer).

17

1 SUPPLEMENTAL INFORMATION

2 Supplemental Data include 5 figures and 9 tables and can be found with this article online at
3 <http://dx.doi.org/XXX>.

5 AUTHOR CONTRIBUTIONS

6 G.K., E.P, G.L and L.Q.M conceived and designed the study. G.K was the lead analyst, with
7 important contributions from E.P. and G.L. A.L.N performed the functional analyses of the
8 *LBP D283G* variant. E.P, G.L and L.Q.M oversaw the study. G.K., E.P, G.L and L.Q.M
9 wrote the manuscript with substantial contributions from A.L.N., L.A and J.L.C.

10

11 DECLARATION OF INTERESTS

12 The authors have no competing interests to declare.

13

14 ACKNOWLEDGMENTS

15 We thank all members of the Human Evolutionary Genetics Laboratory at Institut Pasteur,
16 Paris, and Nicolas Rascovan, Bertrand Boisson and Iain Mathieson for data sharing and
17 helpful discussions. This work was supported by the *Institut Pasteur*, the *Collège de France*,
18 the *Centre Nationale de la Recherche Scientifique* (CNRS), the *Agence Nationale de la*
19 *Recherche* (ANR) grants LIFECHANGE (ANR 17 CE12 0018 02), CNSVIRGEN (ANR-19-
20 CE15-0009-02) and MORTUI (ANR-19-CE35-0005), the French Government's
21 *Investissement d'Avenir* program, *Laboratoires d'Excellence* "Integrative Biology of
22 Emerging Infectious Diseases" (ANR-10- LABX-62-IBEID) and "*Milieu Intérieur*" (ANR-
23 10-LABX-69-01), the *Fondation pour la Recherche Médicale* (Equipe FRM
24 DEQ20180339214), the *Fondation Allianz-Institut de France*, and the *Fondation de France*
25 (no. 00106080). G.K. is supported by a Pasteur-Roux-Cantarini fellowship.

26

27 DATA AND CODE AVAILABILITY

28 Pseudohaploid ancient and modern genome capture data are available from
29 [https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-](https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data)
30 [present-day-and-ancient-dna-data](https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data) (V44.2). *selink* software is available from
31 <https://github.com/h-e-g/selink>. The code for reproducing the figures of the paper is available
32 from https://github.com/h-e-g/SLiM_aDNA_selection.

33

34

1 WEB RESOURCES SECTION

2 OMIM (Online Mendelian Inheritance in Man): <http://www.omim.org>

3 eQTLs (genQTL): <https://www.eqtlgen.org>

4 GWAS atlas: <https://atlas.ctglab.nl/PheWAS>

5 Protein Atlas: <https://www.proteinatlas.org/>

6 ADAstra: <https://adastra.autosome.ru>

7 Open Targets Genetics: <https://genetics.opentargets.org>

8 GWAS catalog: <https://www.ebi.ac.uk/gwas/home>

9

10 REFERENCES

- 11 1. Casanova, J.-L., and Abel, L. (2005). Inborn errors of immunity to infection : the rule
12 rather than the exception. *J Exp Med* 202, 197-201. 10.1084/jem.20050854.
- 13 2. Cairns, J. (1997). *Matters of Life and Death* (Princeton University Press).
- 14 3. Allison, A.C. (1954). Protection afforded by sickle-cell trait against subtertian
15 malarial infection. *Br Med J* 1, 290-294. 10.1136/bmj.1.4857.290.
- 16 4. Barreiro, L.B., and Quintana-Murci, L. (2010). From evolutionary genetics to human
17 immunology: how selection shapes host defence genes. *Nat Rev Genet* 11, 17-30.
18 10.1038/nrg2698.
- 19 5. Quintana-Murci, L. (2019). Human Immunology through the Lens of Evolutionary
20 Genetics. *Cell* 177, 184-199. <https://doi.org/10.1016/j.cell.2019.02.033>.
- 21 6. Fumagalli, M., and Sironi, M. (2014). Human genome variability, natural selection
22 and infectious diseases. *Curr Opin Immunol* 30C, 9-16.
- 23 7. Karlsson, E.K., Kwiatkowski, D.P., and Sabeti, P.C. (2014). Natural selection and
24 infectious disease in human populations. *Nat Rev Genet* 15, 379-393.
25 10.1038/nrg3734.
- 26 8. Quintana-Murci, L., and Clark, A.G. (2013). Population genetic tools for dissecting
27 innate immunity in humans. *Nat Rev Immunol* 13, 280-293. 10.1038/nri3421.
- 28 9. Diamond, J. (2002). Evolution, consequences and future of plant and animal
29 domestication. *Nature* 418, 700-707. 10.1038/nature01019.
- 30 10. Key, F.M., Posth, C., Esquivel-Gomez, L.R., Hübler, R., Spyrou, M.A., Neumann,
31 G.U., Furtwängler, A., Sabin, S., Burri, M., Wissgott, A., et al. (2020). Emergence of
32 human-adapted *Salmonella enterica* is linked to the Neolithization process. *Nat Ecol*
33 *Evol* 4, 324-333. 10.1038/s41559-020-1106-9.

- 1 11. Wolfe, N.D., Dunavan, C.P., and Diamond, J. (2007). Origins of major human
2 infectious diseases. *Nature* 447, 279-283. 10.1038/nature05775.
- 3 12. Harper, K.N., and Armelagos, G.J. (2013). Genomics, the origins of agriculture, and
4 our changing microbe-scape: time to revisit some old tales and tell some new ones.
5 *Am J Phys Anthropol* 152 Suppl 57, 135-152. 10.1002/ajpa.22396.
- 6 13. Fuchs, K., Rinne, C., Drummer, C., Immel, A., Krause-Kyora, B., and Nebel, A.
7 (2019). Infectious diseases and Neolithic transformations: Evaluating biological and
8 archaeological proxies in the German loess zone between 5500 and 2500 BCE. *The*
9 *Holocene* 29, 1545-1557. 10.1177/0959683619857230.
- 10 14. Astle, W.J., Elding, H., Jiang, T., Allen, D., Ruklisa, D., Mann, A.L., Mead, D.,
11 Bouman, H., Riveros-Mckay, F., Kostadima, M.A., et al. (2016). The Allelic
12 Landscape of Human Blood Cell Trait Variation and Links to Common Complex
13 Disease. *Cell* 167, 1415-1429.e1419. 10.1016/j.cell.2016.10.042.
- 14 15. Bao, E.L., Cheng, A.N., and Sankaran, V.G. (2019). The genetics of human
15 hematopoiesis and its disruption in disease. *EMBO Mol Med* 11, e10316-e10316.
16 10.15252/emmm.201910316.
- 17 16. Liggett, L.A., and Sankaran, V.G. (2020). Unraveling Hematopoiesis through the Lens
18 of Genomics. *Cell* 182, 1384-1400. 10.1016/j.cell.2020.08.030.
- 19 17. Notarangelo, L.D., Bacchetta, R., Casanova, J.-L., and Su, H.C. (2020). Human inborn
20 errors of immunity: An expanding universe. *Sci Immunol* 5, eabb1662.
21 10.1126/sciimmunol.abb1662.
- 22 18. Riley, J.C. (2001). *Rising Life Expectancy: A Global History* (Cambridge University
23 Press). DOI: 10.1017/CBO9781316036495.
- 24 19. Barreiro, L.B., and Quintana-Murci, L. (2020). Evolutionary and population
25 (epi)genetics of immunity to infection. *Hum Genet* 139, 723-732. 10.1007/s00439-
26 020-02167-x.
- 27 20. Benton, M.L., Abraham, A., LaBella, A.L., Abbot, P., Rokas, A., and Capra, J.A.
28 (2021). The influence of evolutionary history on human health and disease. *Nat Rev*
29 *Genet* 22, 269-283. 10.1038/s41576-020-00305-9.
- 30 21. Sironi, M., and Clerici, M. (2010). The hygiene hypothesis: an evolutionary
31 perspective. *Microbes and Infection* 12, 421-427.
32 <https://doi.org/10.1016/j.micinf.2010.02.002>.

- 1 22. Fodil, N., Langlais, D., and Gros, P. (2016). Primary Immunodeficiencies and
2 Inflammatory Disease: A Growing Genetic Intersection. *Trends Immunol* 37, 126-
3 140. [10.1016/j.it.2015.12.006](https://doi.org/10.1016/j.it.2015.12.006).
- 4 23. Langlais, D., Fodil, N., and Gros, P. (2017). Genetics of Infectious and Inflammatory
5 Diseases: Overlapping Discoveries from Association and Exome-Sequencing Studies.
6 *Ann Rev Immunol* 35, 1-30. [10.1146/annurev-immunol-051116-052442](https://doi.org/10.1146/annurev-immunol-051116-052442).
- 7 24. Fumagalli, M., Pozzoli, U., Cagliani, R., Comi, G.P., Riva, S., Clerici, M., Bresolin,
8 N., and Sironi, M. (2009). Parasites represent a major selective force for interleukin
9 genes and shape the genetic predisposition to autoimmune conditions. *J Exp Med* 206,
10 1395-1408. [10.1084/jem.20082779](https://doi.org/10.1084/jem.20082779).
- 11 25. Jostins, L., Ripke, S., Weersma, R.K., Duerr, R.H., McGovern, D.P., Hui, K.Y., Lee,
12 J.C., Schumm, L.P., Sharma, Y., Anderson, C.A., et al. (2012). Host-microbe
13 interactions have shaped the genetic architecture of inflammatory bowel disease.
14 *Nature* 491, 119-124. [10.1038/nature11582](https://doi.org/10.1038/nature11582).
- 15 26. Tsoi, L.C., Spain, S.L., Knight, J., Ellinghaus, E., Stuart, P.E., Capon, F., Ding, J., Li,
16 Y., Tejasvi, T., Gudjonsson, J.E., et al. (2012). Identification of 15 new psoriasis
17 susceptibility loci highlights the role of innate immunity. *Nat Genet* 44, 1341-1348.
18 [10.1038/ng.2467](https://doi.org/10.1038/ng.2467).
- 19 27. Zhernakova, A., Elbers, C.C., Ferwerda, B., Romanos, J., Trynka, G., Dubois, P.C., de
20 Kovel, C.G.F., Franke, L., Oosting, M., Barisani, D., et al. (2010). Evolutionary and
21 functional analysis of celiac risk loci reveals SH2B3 as a protective factor against
22 bacterial infection. *Am J Hum Genet* 86, 970-977. [10.1016/j.ajhg.2010.05.004](https://doi.org/10.1016/j.ajhg.2010.05.004).
- 23 28. Prugnolle, F., Manica, A., Charpentier, M., Guegan, J.F., Guernier, V., and Balloux,
24 F. (2005). Pathogen-driven selection and worldwide HLA class I diversity. *Curr Biol*
25 15, 1022-1027. [10.1016/j.cub.2005.04.050](https://doi.org/10.1016/j.cub.2005.04.050).
- 26 29. Chen, H., Hayashi, G., Lai, O.Y., Diltthey, A., Kuebler, P.J., Wong, T.V., Martin,
27 M.P., Fernandez Vina, M.A., McVean, G., Wabl, M., et al. (2012). Psoriasis patients
28 are enriched for genetic variants that protect against HIV-1 disease. *PLoS Genet* 8,
29 e1002514-e1002514. [10.1371/journal.pgen.1002514](https://doi.org/10.1371/journal.pgen.1002514).
- 30 30. Gough, S.C.L., and Simmonds, M.J. (2007). The HLA Region and Autoimmune
31 Disease: Associations and Mechanisms of Action. *Curr Genomics* 8, 453-465.
32 [10.2174/138920207783591690](https://doi.org/10.2174/138920207783591690).

- 1 31. Matzaraki, V., Kumar, V., Wijmenga, C., and Zhernakova, A. (2017). The MHC locus
2 and genetic susceptibility to autoimmune and infectious diseases. *Genome Biol* 18, 76.
3 10.1186/s13059-017-1207-1.
- 4 32. Ritari, J., Koskela, S., Hyvärinen, K., FinnGen, and Partanen, J. (2022). HLA-disease
5 association and pleiotropy landscape in over 235,000 Finns. *Hum Immunol* 83, 391-
6 398. <https://doi.org/10.1016/j.humimm.2022.02.003>.
- 7 33. Ryder, L.P., Svejgaard, A., and Dausset, J. (1981). GENETICS OF HLA DISEASE
8 ASSOCIATION. *Ann Rev Genet* 15, 169-187.
9 10.1146/annurev.ge.15.120181.001125.
- 10 34. Boisson-Dupuis, S., Ramirez-Alejo, N., Li, Z., Patin, E., Rao, G., Kerner, G., Lim,
11 C.K., Kremontsov, D.N., Hernandez, N., Ma, C.S., et al. (2018). Tuberculosis and
12 impaired IL-23-dependent IFN- γ immunity in humans homozygous for a common
13 TYK2 missense variant. *Sci Immunol* 3, eaau8714. 10.1126/sciimmunol.aau8714.
- 14 35. Kerner, G., Laval, G., Patin, E., Boisson-Dupuis, S., Abel, L., Casanova, J.-L., and
15 Quintana-Murci, L. (2021). Human ancient DNA analyses reveal the high burden of
16 tuberculosis in Europeans over the last 2,000 years. *Am J Hum Genet* 108, 517-524.
17 10.1016/j.ajhg.2021.02.009.
- 18 36. Dendrou, C.A., Cortes, A., Shipman, L., Evans, H.G., Attfield, K.E., Jostins, L.,
19 Barber, T., Kaur, G., Kuttikkatte, S.B., Leach, O.A., et al. (2016). Resolving TYK2
20 locus genotype-to-phenotype differences in autoimmunity. *Sci Transl Med* 8,
21 363ra149-363ra149. 10.1126/scitranslmed.aag1974.
- 22 37. Diogo, D., Bastarache, L., Liao, K.P., Graham, R.R., Fulton, R.S., Greenberg, J.D.,
23 Eyre, S., Bowes, J., Cui, J., Lee, A., et al. (2015). TYK2 protein-coding variants
24 protect against rheumatoid arthritis and autoimmunity, with no evidence of major
25 pleiotropic effects on non-autoimmune complex traits. *PLoS One* 10, e0122271-
26 e0122271. 10.1371/journal.pone.0122271.
- 27 38. International Genetics of Ankylosing Spondylitis, C., Cortes, A., Hadler, J., Pointon,
28 J.P., Robinson, P.C., Karaderi, T., Leo, P., Cremin, K., Pryce, K., Harris, J., et al.
29 (2013). Identification of multiple risk variants for ankylosing spondylitis through
30 high-density genotyping of immune-related loci. *Nat Genet* 45, 730-738.
31 10.1038/ng.2667.
- 32 39. Kerner, G., Ramirez-Alejo, N., Seeleuthner, Y., Yang, R., Ogishi, M., Cobat, A.,
33 Patin, E., Quintana-Murci, L., Boisson-Dupuis, S., Casanova, J.-L., and Abel, L.
34 (2019). Homozygosity for TYK2 P1104A underlies tuberculosis in about 1% of

- 1 patients in a cohort of European ancestry. *Proc Natl Acad Sci U S A* *116*, 10430-
2 10434. [10.1073/pnas.1903561116](https://doi.org/10.1073/pnas.1903561116).
- 3 40. Ju, D., and Mathieson, I. (2021). The evolution of skin pigmentation-associated
4 variation in West Eurasia. *Proc Natl Acad Sci U S A* *118*, e2009227118.
5 [10.1073/pnas.2009227118](https://doi.org/10.1073/pnas.2009227118).
- 6 41. Key, F.M., Fu, Q., Romagné, F., Lachmann, M., and Andrés, A.M. (2016). Human
7 adaptation and population differentiation in the light of ancient genomes. *Nat*
8 *Commun* *7*, 10775-10775. [10.1038/ncomms10775](https://doi.org/10.1038/ncomms10775).
- 9 42. Mathieson, I. (2020). Limited Evidence for Selection at the FADS Locus in Native
10 American Populations. *Mol Biol Evol* *37*, 2029-2033. [10.1093/molbev/msaa064](https://doi.org/10.1093/molbev/msaa064).
- 11 43. Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S.A.,
12 Harney, E., Stewardson, K., Fernandes, D., Novak, M., et al. (2015). Genome-wide
13 patterns of selection in 230 ancient Eurasians. *Nature* *528*, 499-503.
14 [10.1038/nature16152](https://doi.org/10.1038/nature16152).
- 15 44. Lindo, J., Huerta-Sánchez, E., Nakagome, S., Rasmussen, M., Petzelt, B., Mitchell, J.,
16 Cybulski, J.S., Willerslev, E., DeGiorgio, M., and Malhi, R.S. (2016). A time transect
17 of exomes from a Native American population before and after European contact. *Nat*
18 *Commun* *7*, 13175-13175. [10.1038/ncomms13175](https://doi.org/10.1038/ncomms13175).
- 19 45. Beaumont, M.A., and Rannala, B. (2004). The Bayesian revolution in genetics. *Nat*
20 *Rev Genet* *5*, 251-261. [10.1038/nrg1318](https://doi.org/10.1038/nrg1318).
- 21 46. Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K.,
22 Sudmant, P.H., Schraiber, J.G., Castellano, S., Lipson, M., et al. (2014). Ancient
23 human genomes suggest three ancestral populations for present-day Europeans. *Nature*
24 *513*, 409-413. [10.1038/nature13673](https://doi.org/10.1038/nature13673).
- 25 47. Skoglund, P., Malmström, H., Raghavan, M., Storå, J., Hall, P., Willerslev, E.,
26 Gilbert, M.T.P., Götherström, A., and Jakobsson, M. (2012). Origins and Genetic
27 Legacy of Neolithic Farmers and Hunter-Gatherers in Europe. *Science* *336*, 466-469.
28 [10.1126/science.1216304](https://doi.org/10.1126/science.1216304).
- 29 48. Allentoft, M.E., Sikora, M., Sjögren, K.-G., Rasmussen, S., Rasmussen, M.,
30 Stenderup, J., Damgaard, P.B., Schroeder, H., Ahlström, T., Vinner, L., et al. (2015).
31 Population genomics of Bronze Age Eurasia. *Nature* *522*, 167-172.
32 [10.1038/nature14507](https://doi.org/10.1038/nature14507).
- 33 49. Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt,
34 G., Nordenfelt, S., Harney, E., Stewardson, K., et al. (2015). Massive migration from

- 1 the steppe was a source for Indo-European languages in Europe. *Nature* 522, 207-211.
2 10.1038/nature14317.
- 3 50. Burger, J., Link, V., Blöcher, J., Schulz, A., Sell, C., Pochon, Z., Diekmann, Y.,
4 Žegarac, A., Hofmanová, Z., Winkelbach, L., et al. (2020). Low Prevalence of Lactase
5 Persistence in Bronze Age Europe Indicates Ongoing Strong Selection over the Last
6 3,000 Years. *Curr Biol* 30, 4307-4315.e4313.
7 <https://doi.org/10.1016/j.cub.2020.08.033>.
- 8 51. Segurel, L., Guarino-Vignon, P., Marchi, N., Lafosse, S., Laurent, R., Bon, C., Fabre,
9 A., Hegay, T., and Heyer, E. (2020). Why and when was lactase persistence selected
10 for? Insights from Central Asian herders and ancient DNA. *PLoS Biol* 18, e3000742-
11 e3000742. 10.1371/journal.pbio.3000742.
- 12 52. Kristiansen, H., Scherer, C.A., McVean, M., Iadonato, S.P., Vends, S., Thavachelvam,
13 K., Steffensen, T.B., Horan, K.A., Kuri, T., Weber, F., et al. (2010). Extracellular 2'-5'
14 oligoadenylate synthetase stimulates RNase L-independent antiviral activity: a novel
15 mechanism of virus-induced innate immunity. *J Virol* 84, 11898-11904.
16 10.1128/JVI.01003-10.
- 17 53. Shelton, J.F., Shastri, A.J., Ye, C., Weldon, C.H., Filshtein-Sonmez, T., Coker, D.,
18 Symons, A., Esparza-Gordillo, J., Chubb, A., Fitch, A., et al. (2021). Trans-ancestry
19 analysis reveals genetic and nongenetic associations with COVID-19 susceptibility
20 and severity. *Nat Genet* 53, 801-808. 10.1038/s41588-021-00854-7.
- 21 54. Pickrell, J.K., Berisa, T., Liu, J.Z., Ségurel, L., Tung, J.Y., and Hinds, D.A. (2016).
22 Detection and interpretation of shared genetic influences on 42 human traits. *Nat*
23 *Genet* 48, 709-717. 10.1038/ng.3570.
- 24 55. Tian, C., Hromatka, B.S., Kiefer, A.K., Eriksson, N., Noble, S.M., Tung, J.Y., and
25 Hinds, D.A. (2017). Genome-wide association and HLA region fine-mapping studies
26 identify susceptibility loci for multiple common infections. *Nat Commun* 8, 599-599.
27 10.1038/s41467-017-00257-5.
- 28 56. Malaria Genomic Epidemiology, N. (2019). Insights into malaria susceptibility using
29 genome-wide data on 17,000 individuals from Africa, Asia and Oceania. *Nat*
30 *Commun* 10, 5732-5732. 10.1038/s41467-019-13480-z.
- 31 57. Ségurel, L., Thompson, E.E., Flutre, T., Lovstad, J., Venkat, A., Margulis, S.W.,
32 Moyse, J., Ross, S., Gamble, K., Sella, G., et al. (2012). The ABO blood group is a
33 trans-species polymorphism in primates. *Proc Natl Acad Sci U S A* 109, 18493-18498.
34 10.1073/pnas.1210603109.

- 1 58. Cuadros-Espinoza, S., Laval, G., Quintana-Murci, L., and Patin, E. (2022). The
2 genomic signatures of natural selection in admixed human populations. *Am J Hum*
3 *Genet* *109*, 710-726. <https://doi.org/10.1016/j.ajhg.2022.02.011>.
- 4 59. Zeberg, H., and Pääbo, S. (2021). A genomic region associated with protection against
5 severe COVID-19 is inherited from Neandertals. *Proc Natl Acad Sci U S A* *118*,
6 e2026309118. doi:10.1073/pnas.2026309118.
- 7 60. Huffman, J.E., Butler-Laporte, G., Khan, A., Pairo-Castineira, E., Drivas, T.G.,
8 Peloso, G.M., Nakanishi, T., Initiative, C.-H.G., Ganna, A., Verma, A., et al. (2022).
9 Multi-ancestry fine mapping implicates OAS1 splicing in risk of severe COVID-19.
10 *Nat Genet* *54*, 125-127. 10.1038/s41588-021-00996-8.
- 11 61. Zhou, S., Butler-Laporte, G., Nakanishi, T., Morrison, D.R., Afilalo, J., Afilalo, M.,
12 Laurent, L., Pietzner, M., Kerrison, N., Zhao, K., et al. (2021). A Neanderthal OAS1
13 isoform protects individuals of European ancestry against COVID-19 susceptibility
14 and severity. *Nat Med* *27*, 659-667. 10.1038/s41591-021-01281-1.
- 15 62. Abramov, S., Boytsov, A., Bykova, D., Penzar, D.D., Yevshin, I., Kolmykov, S.K.,
16 Fridman, M.V., Favorov, A.V., Vorontsov, I.E., Baulin, E., et al. (2021). Landscape of
17 allele-specific transcription factor binding in the human genome. *Nat Commun* *12*,
18 2751-2751. 10.1038/s41467-021-23007-0.
- 19 63. Ferreira, M.A.R., Mathur, R., Vonk, J.M., Szwajda, A., Brumpton, B., Granell, R.,
20 Brew, B.K., Ullemer, V., Lu, Y., Jiang, Y., et al. (2019). Genetic Architectures of
21 Childhood- and Adult-Onset Asthma Are Partly Distinct. *Am J Hum Genet* *104*, 665-
22 684. 10.1016/j.ajhg.2019.02.022.
- 23 64. Davis, C.A., Hitz, B.C., Sloan, C.A., Chan, E.T., Davidson, J.M., Gabdank, I., Hilton,
24 J.A., Jain, K., Baymuradov, U.K., Narayanan, A.K., et al. (2018). The Encyclopedia
25 of DNA elements (ENCODE): data portal update. *Nucleic Acids Res* *46*, D794-D801.
26 10.1093/nar/gkx1081.
- 27 65. Nunes-Santos, C.J., Kuehn, H.S., and Rosenzweig, S.D. (2020). IKAROS Family Zinc
28 Finger 1-Associated Diseases in Primary Immunodeficiency Patients. *Immunol*
29 *Allergy Clin North Am* *40*, 461-470. 10.1016/j.iac.2020.04.004.
- 30 66. Panda, D., Gjinaj, E., Bachu, M., Squire, E., Novatt, H., Ozato, K., and Rabin, R.L.
31 (2019). IRF1 Maintains Optimal Constitutive Expression of Antiviral Genes and
32 Regulates the Early Antiviral Response. *Front Immunol* *10*.
33 10.3389/fimmu.2019.01019.

- 1 67. Lindesmith, L., Moe, C., Marionneau, S., Ruvoen, N., Jiang, X., Lindblad, L., Stewart,
2 P., LePendu, J., and Baric, R. (2003). Human susceptibility and resistance to Norwalk
3 virus infection. *Nat Med* 9, 548-553. 10.1038/nm860.
- 4 68. Raza, M.W., Blackwell, C.C., Molyneaux, P., James, V.S., Ogilvie, M.M., Inglis,
5 J.M., and Weir, D.M. (1991). Association between secretor status and respiratory viral
6 illness. *BMJ* 303, 815-818. 10.1136/bmj.303.6806.815.
- 7 69. Park, B.S., and Lee, J.-O. (2013). Recognition of lipopolysaccharide pattern by TLR4
8 complexes. *Exp Mol Med* 45, e66-e66. 10.1038/emm.2013.97.
- 9 70. Eckert, Jana K., Kim, Young J., Kim, Jung I., Gürtler, K., Oh, D.-Y., Sur, S.,
10 Lundvall, L., Hamann, L., van der Ploeg, A., Pickkers, P., et al. (2013). The Crystal
11 Structure of Lipopolysaccharide Binding Protein Reveals the Location of a Frequent
12 Mutation that Impairs Innate Immunity. *Immunity* 39, 647-660.
13 <https://doi.org/10.1016/j.immuni.2013.09.005>.
- 14 71. Gleich, G.J., Klion, A.D., Lee, J.J., and Weller, P.F. (2013). The consequences of not
15 having eosinophils. *Allergy* 68, 829-835. 10.1111/all.12169.
- 16 72. Andrades Valtueña, A., Neumann Gunnar, U., Spyrou Maria, A., Musralina, L., Aron,
17 F., Beisenov, A., Belinskiy Andrey, B., Bos Kirsten, I., Buzhilova, A., Conrad, M., et
18 al. (2022). Stone Age *Yersinia pestis* genomes shed light on the early evolution,
19 diversity, and ecology of plague. *Proc Natl Acad Sci U S A* 119, e2116722119.
20 10.1073/pnas.2116722119.
- 21 73. Rascovan, N., Sjögren, K.-G., Kristiansen, K., Nielsen, R., Willerslev, E., Desnues,
22 C., and Rasmussen, S. (2019). Emergence and Spread of Basal Lineages of
23 *Yersinia pestis* during the Neolithic Decline. *Cell* 176, 295-305.e210.
24 <https://doi.org/10.1016/j.cell.2018.11.005>.
- 25 74. Skoglund, P., and Mathieson, I. (2018). Ancient Genomics of Modern Humans: The
26 First Decade. *Ann Rev Genom Hum Genet* 19, 381-404. 10.1146/annurev-genom-
27 083117-021749.
- 28 75. Kristiansen, K., Fowler, C., Harding, J., & Hofmann, D. (2014). The Decline of the
29 Neolithic and the Rise of Bronze Age Society. *Oxford Handbooks Online*.
30 <https://doi.org/10.1093/OXFORDHOB/9780199545841.013.057>.
- 31 76. Charlesworth, B. (2009). Effective population size and patterns of molecular evolution
32 and variation. *Nat Rev Genet* 10, 195-205. 10.1038/nrg2526.
- 33 77. Scott, A., Reinhold, S., Hermes, T., Kalmykov, A.A., Belinskiy, A., Buzhilova, A.,
34 Berezina, N., Kantorovich, A.R., Maslov, V.E., Guliyev, F., et al. (2022). Emergence

- 1 and intensification of dairying in the Caucasus and Eurasian steppes. *Nat Ecol Evol.*
2 10.1038/s41559-022-01701-6.
- 3 78. Racimo, F., Woodbridge, J., Fyfe, R.M., Sikora, M., Sjögren, K.-G., Kristiansen, K.,
4 and Vander Linden, M. (2020). The spatiotemporal spread of human migrations
5 during the European Holocene. *Proc Natl Acad Sci U S A* *117*, 8989-9000.
6 10.1073/pnas.1920051117.
- 7 79. Song, W., Shi, Y., Wang, W., Pan, W., Qian, W., Yu, S., Zhao, M., and Lin, G.N.
8 (2021). A selection pressure landscape for 870 human polygenic traits. *Nat Hum*
9 *Behav* *5*, 1731-1743. 10.1038/s41562-021-01231-4.
- 10 80. Saevarsdottir, S., Olafsdottir, T.A., Ivarsdottir, E.V., Halldorsson, G.H.,
11 Gunnarsdottir, K., Sigurdsson, A., Johannesson, A., Sigurdsson, J.K., Juliusdottir, T.,
12 Lund, S.H., et al. (2020). FLT3 stop mutation increases FLT3 ligand level and risk of
13 autoimmune thyroid disease. *Nature* *584*, 619-623. 10.1038/s41586-020-2436-0.
- 14 81. Croci, S., Venneri, M.A., Mantovani, S., Fallerini, C., Benetti, E., Picchiotti, N.,
15 Campolo, F., Imperatore, F., Palmieri, M., Daga, S., et al. (2021). The polymorphism
16 L412F in TLR3 inhibits autophagy and is a marker of severe COVID-19 in males.
17 *Autophagy*, 1-11. 10.1080/15548627.2021.1995152.
- 18 82. de Lange, K.M., Moutsianas, L., Lee, J.C., Lamb, C.A., Luo, Y., Kennedy, N.A.,
19 Jostins, L., Rice, D.L., Gutierrez-Achury, J., Ji, S.-G., et al. (2017). Genome-wide
20 association study implicates immune activation of multiple integrin genes in
21 inflammatory bowel disease. *Nat Genet* *49*, 256-261. 10.1038/ng.3760.
- 22 83. Guo, Y., Audry, M., Ciancanelli, M., Alsina, L., Azevedo, J., Herman, M., Anguiano,
23 E., Sancho-Shimizu, V., Lorenzo, L., Pauwels, E., et al. (2011). Herpes simplex virus
24 encephalitis in a patient with complete TLR3 deficiency: TLR3 is otherwise redundant
25 in protective immunity. *J Exp Med* *208*, 2083-2098. 10.1084/jem.20101568.
- 26 84. Zhang, S.Y., Jouanguy, E., Ugolini, S., Smahi, A., Elain, G., Romero, P., Segal, D.,
27 Sancho-Shimizu, V., Lorenzo, L., Puel, A., et al. (2007). TLR3 deficiency in patients
28 with herpes simplex encephalitis. *Science* *317*, 1522-1527. 317/5844/1522
29 [pii]10.1126/science.1139522.
- 30 85. Martínez-Barricarte, R., Markle, J.G., Ma, C.S., Deenick, E.K., Ramírez-Alejo, N.,
31 Mele, F., Latorre, D., Mahdavian, S.A., Aytikin, C., Mansouri, D., et al. (2018).
32 Human IFN- γ immunity to mycobacteria is governed by both IL-12 and IL-23. *Sci*
33 *Immunol* *3*, eaau6759. 10.1126/sciimmunol.aau6759.

- 1 86. Sun, R., Hedl, M., and Abraham, C. (2020). IL23 induces IL23R recycling and
2 amplifies innate receptor-induced signalling and cytokines in human macrophages,
3 and the IBD-protective IL23R R381Q variant modulates these outcomes. *Gut* 69, 264-
4 273. 10.1136/gutjnl-2018-316830.
- 5 87. Brunel, S., Bennett, E.A., Cardin, L., Garraud, D., Emam, H.B., Beylier, A.,
6 Boulestin, B., Chenal, F., Ciesielski, E., Convertini, F., et al. (2020). Ancient genomes
7 from present-day France unveil 7,000 years of its demographic history. *Proc Natl*
8 *Acad Sci U S A* 117, 12791-12798. doi:10.1073/pnas.1918034117.
- 9 88. Margaryan, A., Lawson, D.J., Sikora, M., Racimo, F., Rasmussen, S., Moltke, I.,
10 Cassidy, L.M., Jørsboe, E., Ingason, A., Pedersen, M.W., et al. (2020). Population
11 genomics of the Viking world. *Nature* 585, 390-396. 10.1038/s41586-020-2688-8.
- 12 89. Olalde, I., Mallick, S., Patterson, N., Rohland, N., Villalba-Mouco, V., Silva, M.,
13 Duliias, K., Edwards, C.J., Gandini, F., Pala, M., et al. (2019). The genomic history of
14 the Iberian Peninsula over the past 8000 years. *Science* 363, 1230-1234.
15 doi:10.1126/science.aav4040.
- 16 90. Monroy Kuhn, J.M., Jakobsson, M., and Günther, T. (2018). Estimating genetic kin
17 relationships in prehistoric populations. *PLoS One* 13, e0195491-e0195491.
18 10.1371/journal.pone.0195491.
- 19 91. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alfoldi, J., Wang, Q.,
20 Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The
21 mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*
22 581, 434-443. 10.1038/s41586-020-2308-7.
- 23 92. Fairley, S., Lowy-Gallego, E., Perry, E., and Flicek, P. (2020). The International
24 Genome Sample Resource (IGSR) collection of open human genomic variation
25 resources. *Nucleic Acids Res* 48, D941-D947. 10.1093/nar/gkz836.
- 26 93. François, O., and Jay, F. (2020). Factor analysis of ancient population genomic
27 samples. *Nat Commun* 11, 4661-4661. 10.1038/s41467-020-18335-6.
- 28 94. Haller, B.C., and Messer, P.W. (2019). SLiM 3: Forward Genetic Simulations Beyond
29 the Wright-Fisher Model. *Mol Biol Evol* 36, 632-637. 10.1093/molbev/msy228.
- 30 95. Gutenkunst, R.N., Hernandez, R.D., Williamson, S.H., and Bustamante, C.D. (2009).
31 Inferring the joint demographic history of multiple populations from multidimensional
32 SNP frequency data. *PLoS Genet* 5, e1000695-e1000695.
33 10.1371/journal.pgen.1000695.

- 1 96. Vergara-Lope, A., Jabalameli, M.R., Horscroft, C., Ennis, S., Collins, A., and
2 Pengelly, R.J. (2019). Linkage disequilibrium maps for European and African
3 populations constructed from whole genome sequence data. *Sci Data* 6, 208-208.
4 10.1038/s41597-019-0227-y.
- 5 97. Deschamps, M., Laval, G., Fagny, M., Itan, Y., Abel, L., Casanova, J.-L., Patin, E.,
6 and Quintana-Murci, L. (2016). Genomic Signatures of Selective Pressures and
7 Introgression from Archaic Hominins at Human Innate Immunity Genes. *Am J Hum*
8 *Genet* 98, 5-21. 10.1016/j.ajhg.2015.11.014.
- 9 98. Young, M.D., Wakefield, M.J., Smyth, G.K., and Oshlack, A. (2010). Gene ontology
10 analysis for RNA-seq: accounting for selection bias. *Genome Biol* 11, R14.
11 10.1186/gb-2010-11-2-r14.
- 12 99. Võsa, U., Claringbould, A., Westra, H.-J., Bonder, M.J., Deelen, P., Zeng, B., Kirsten,
13 H., Saha, A., Kreuzhuber, R., Yazar, S., et al. (2021). Large-scale cis- and trans-eQTL
14 analyses identify thousands of genetic loci and polygenic scores that regulate blood
15 gene expression. *Nat Genet* 53, 1300-1310. 10.1038/s41588-021-00913-z.
- 16 100. Schmidt, E.M., Zhang, J., Zhou, W., Chen, J., Mohlke, K.L., Chen, Y.E., and Willer,
17 C.J. (2015). GREGOR: evaluating global enrichment of trait-associated variants in
18 epigenomic features using a systematic, data-driven approach. *Bioinformatics* 31,
19 2601-2606. 10.1093/bioinformatics/btv201.
- 20 101. Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.R.,
21 ReproGen, C., Psychiatric Genomics, C., Genetic Consortium for Anorexia Nervosa
22 of the Wellcome Trust Case Control, C., Duncan, L., et al. (2015). An atlas of genetic
23 correlations across human diseases and traits. *Nat Genet* 47, 1236-1241.
24 10.1038/ng.3406.
- 25 102. Voight, B.F., Kudravalli, S., Wen, X., and Pritchard, J.K. (2006). A map of recent
26 positive selection in the human genome. *PLoS Biol* 4, e72-e72.
27 10.1371/journal.pbio.0040072.
- 28 103. Grossman Sharon, R., Shylakhter, I., Karlsson Elinor, K., Byrne Elizabeth, H.,
29 Morales, S., Frieden, G., Hostetter, E., Angelino, E., Garber, M., Zuk, O., et al.
30 (2010). A Composite of Multiple Signals Distinguishes Causal Variants in Regions of
31 Positive Selection. *Science* 327, 883-886. 10.1126/science.1183863.
- 32 104. Ferrer-Admetlla, A., Liang, M., Korneliussen, T., and Nielsen, R. (2014). On
33 detecting incomplete soft or hard selective sweeps using haplotype structure. *Mol Biol*
34 *Evol* 31, 1275-1291. 10.1093/molbev/msu077.

- 1 105. Barreiro, L.B., Ben-Ali, M., Quach, H., Laval, G., Patin, E., Pickrell, J.K., Bouchier,
2 C., Tichit, M., Neyrolles, O., Gicquel, B., et al. (2009). Evolutionary dynamics of
3 human Toll-like receptors and their different contributions to host defense. *PLoS*
4 *Genet* 5, e1000562-e1000562. [10.1371/journal.pgen.1000562](https://doi.org/10.1371/journal.pgen.1000562).
- 5 106. Wang, X., Spandidos, A., Wang, H., and Seed, B. (2012). PrimerBank: a PCR primer
6 database for quantitative gene expression analysis, 2012 update. *Nucleic Acids Res*
7 *40*, D1144-1149. [10.1093/nar/gkr1013](https://doi.org/10.1093/nar/gkr1013).
- 8

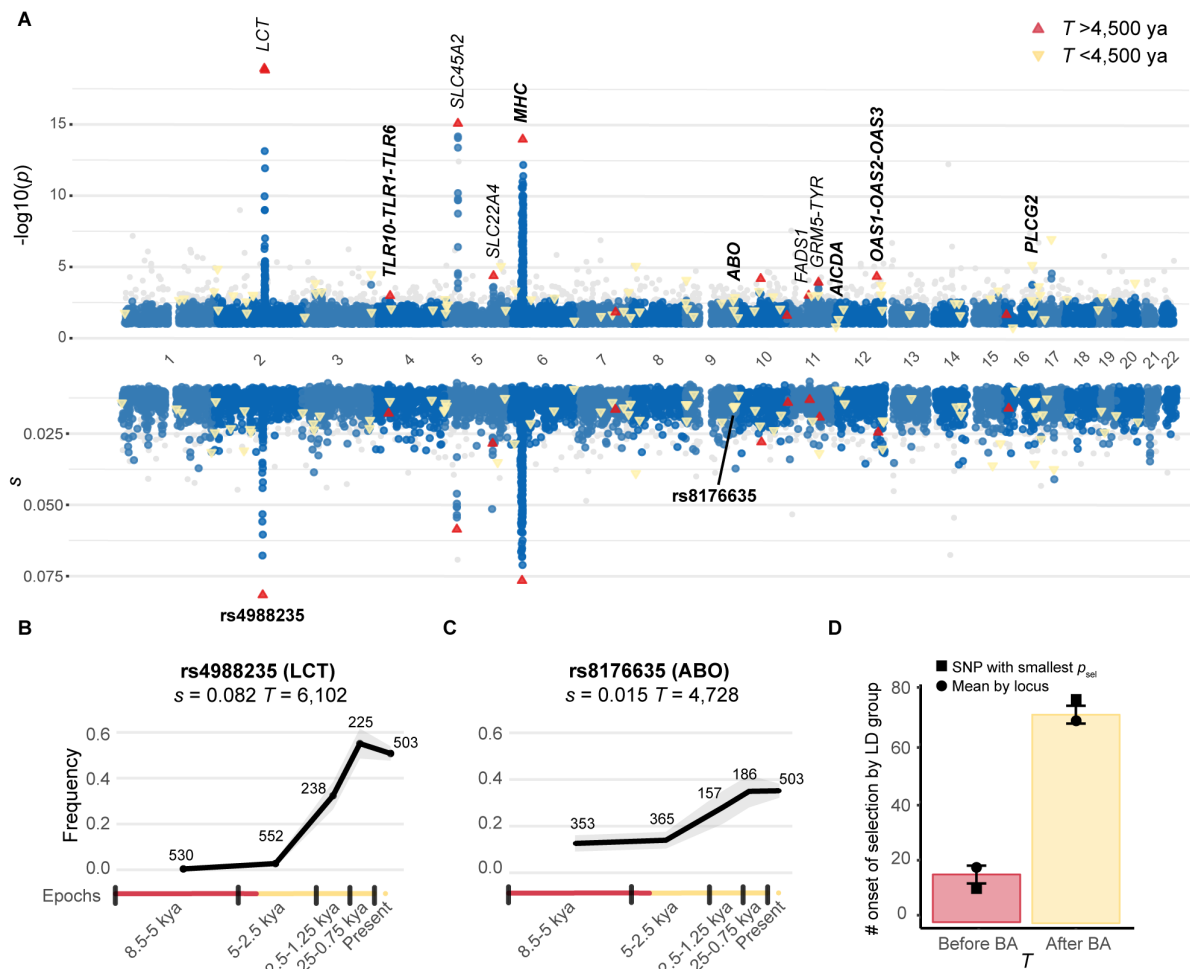


Figure 1. Genome-wide detection of positive selection in Europe since the Neolithic

(A) Test for positive selection ($-\log_{10}(p_{sel})$) and selection coefficient estimates (s) for each genomic marker in the capture dataset with $p_{sel} < 0.1$. The empirical null distribution was approximated by a beta distribution (**Methods**). Variants with $p_{sel} < 10^{-4}$ outside the 89 enriched loci are colored in gray. Several candidate genes for positive selection are indicated, and host defense genes are highlighted in bold. Upward-pointing pink triangles and downward-pointing yellow triangles correspond to the 89 SNPs with the smallest p_{sel} at each of the 89 candidate loci selected before and after the beginning of the Bronze Age, respectively. The time of selection onset was estimated as the mean T across all SNPs with $p_{sel} < 0.01$ at each locus.

(B-C) Frequency trajectory of the most significant variant at (B) the *LCT* locus (rs4988235) and (C) the *ABO* locus (rs8176635). Gray shading around the frequency trajectory indicates the lower and upper bounds of the 95% CI for variant frequency estimation in each epoch.

(D) Mean number of times a single randomly selected candidate SNP per LD group had an estimated T before or after the beginning of the Bronze Age, across 1,000 replicates. The error bar for each bar plot indicates the standard deviation of the distribution obtained. We also included the estimate based on the mean T of all SNPs with $p_{sel} < 0.01$ (as for the Manhattan plot) for each of the 89 LD groups (rectangles), or based on the SNP with the smallest p_{sel} for each LD group (circles).

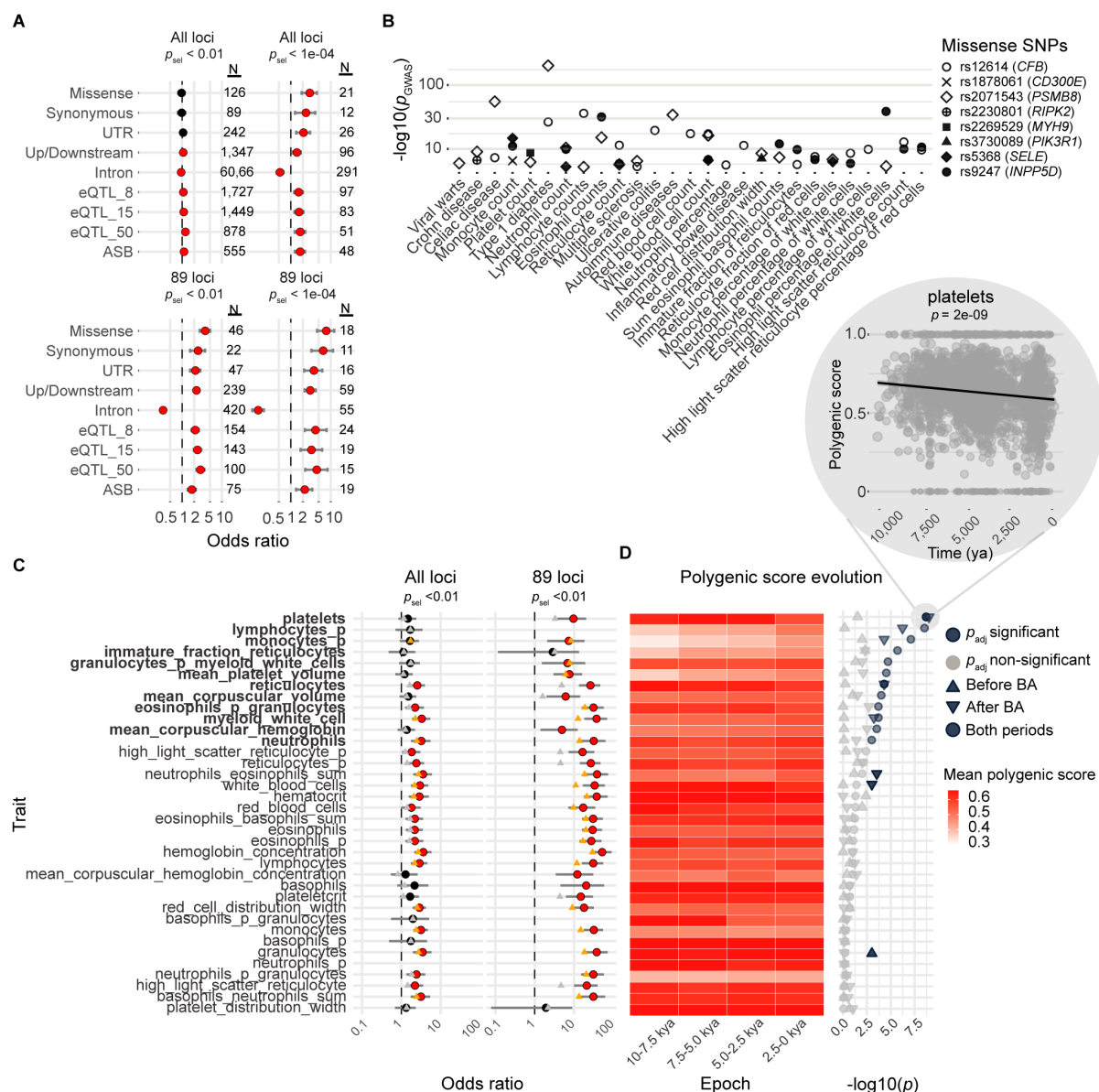


Figure 2. Genetic variants and host defense traits targeted by positive selection

(A) Enrichments in missense, whole blood *cis*-eQTL or ASB variants for variants with $p_{sel} < 10^{-2}$ and $p_{sel} < 10^{-4}$, across the genome ('All loci') or within the 89 candidate positively selected loci ('89 loci'). Only variants of transcripts annotated as 'protein-coding' were considered for the analysis of variant annotations. The three eQTL groups were defined on the basis of the statistical significance of the eQTL association, $p_{eQTL} < 5 \times 10^{-8}$, $< 10^{-15}$ or $< 10^{-50}$.

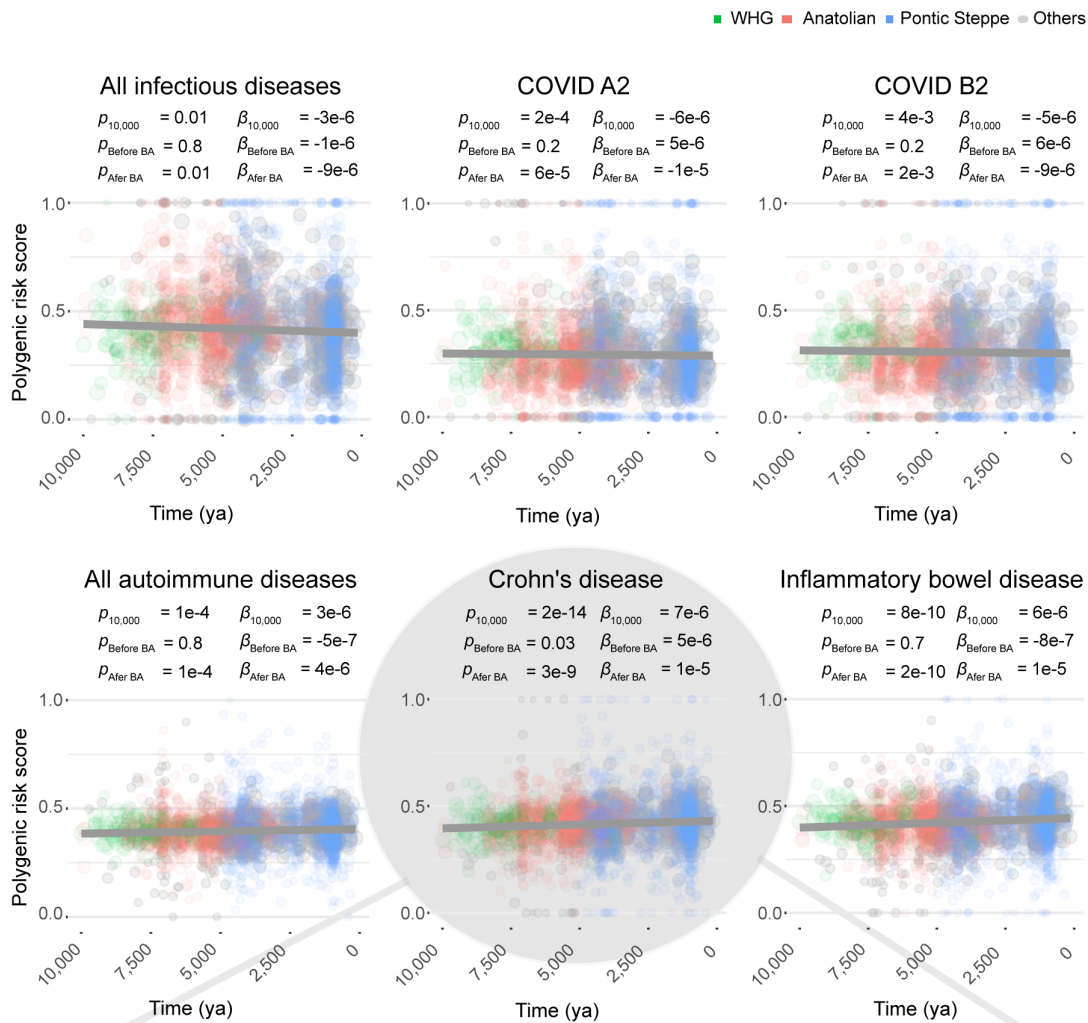
(B) Significant associations ($p < 10^{-5}$; <https://genetics.opentargets.org>) in PheWAS between positively selected missense variants overlapping host defense genes and hematopoietic traits, infections or inflammatory disorders.

(C) Enrichments in variants associated with the 36 hematopoietic traits studied for variants with $p_{sel} < 10^{-2}$, across the genome ('All loci') or within the 89 candidate positively selected loci ('89 loci'). Red and black circles indicate significant and non-significant enrichments, respectively. Orange and gray triangles indicate significant and non-significant enrichments, respectively, after exclusion of the *HLA* locus.

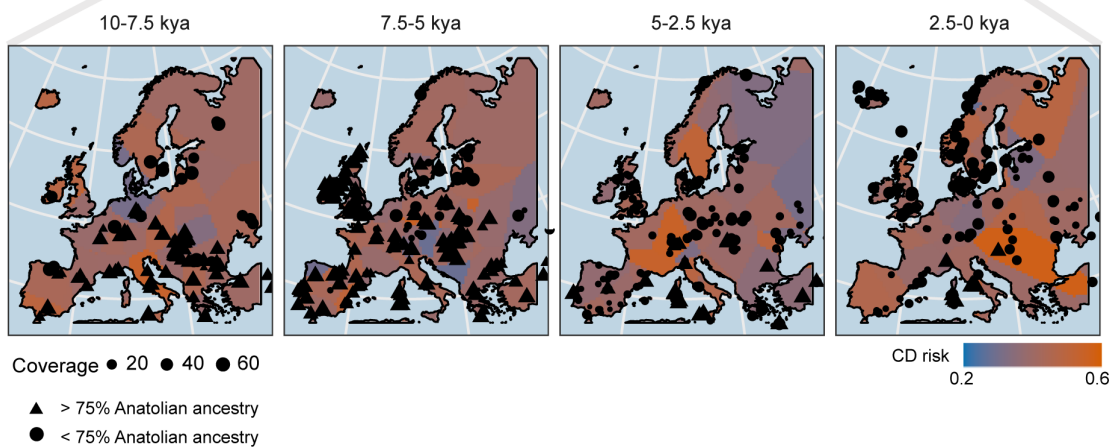
(D) Left: mean polygenic score (**Methods**) across four equally spaced epochs for each of the hematopoietic traits. Right: $-\log_{10} p$ values for the significant increases or decreases in each hematopoietic trait over the last 10,000 years (dot), before (upward-pointing triangle) or after

(downward-pointing triangle) the beginning of the Bronze Age. Light gray symbols represent non-significant trajectories. The inset shows the decrease in polygenic score over time for platelet counts, which has the highest $-\log_{10} p$ value. The black line is the regression line. (C-D) Traits followed by ‘_p’ indicate the percentage of the cell type considered, whereas other traits are absolute counts.

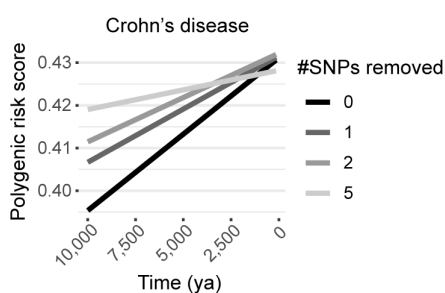
A



B



C



D

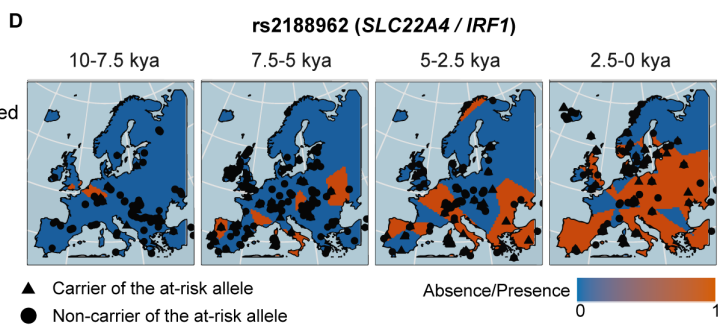


Figure 3. Resistance to infection and risk of inflammation have increased since the Neolithic

(A) Polygenic scores for infectious and autoimmune traits as a function of time, over the last 10,000 years. COVID A2 and B2 indicate critical COVID-19 cases vs. the general population and hospitalized COVID-19 cases vs. the general population, respectively. Dot size scale with the number of SNPs genotyped in the individual. Dark gray lines are the regression lines for a model adjusted for ancestry and geographic location. *P*-values indicate the significance of the regression model, over the last 10 millennia (*p*), before the Bronze Age ($p_{\text{Before BA}}$) or since the beginning of the Bronze Age ($p_{\text{After BA}}$) (**Methods**). Beta values for the regression model considering only samples dating from before ($\beta_{\text{Before BA}}$) or after ($\beta_{\text{After BA}}$) the beginning of the Bronze Age are shown. Green, pink and blue dots indicate individuals with >75% Western hunter-gatherer, Anatolian or Pontic Steppe ancestry, respectively. Individuals with mixed ancestries are shown in gray.

(B) Polygenic score for Crohn's disease as a function of the geographic location and age of the samples, obtained with the “bleiglas” package of R version 3.6.2. Individuals with >75% Anatolian ancestry are represented by triangles, and the others are represented by circles.

(C) Polygenic risk score for Crohn's disease as a function of time, after removal of the SNPs most significantly associated with sample age.

(D) Presence or absence of the CD-risk allele rs2188962>T as a function of the geographic location and the age of ancient samples.

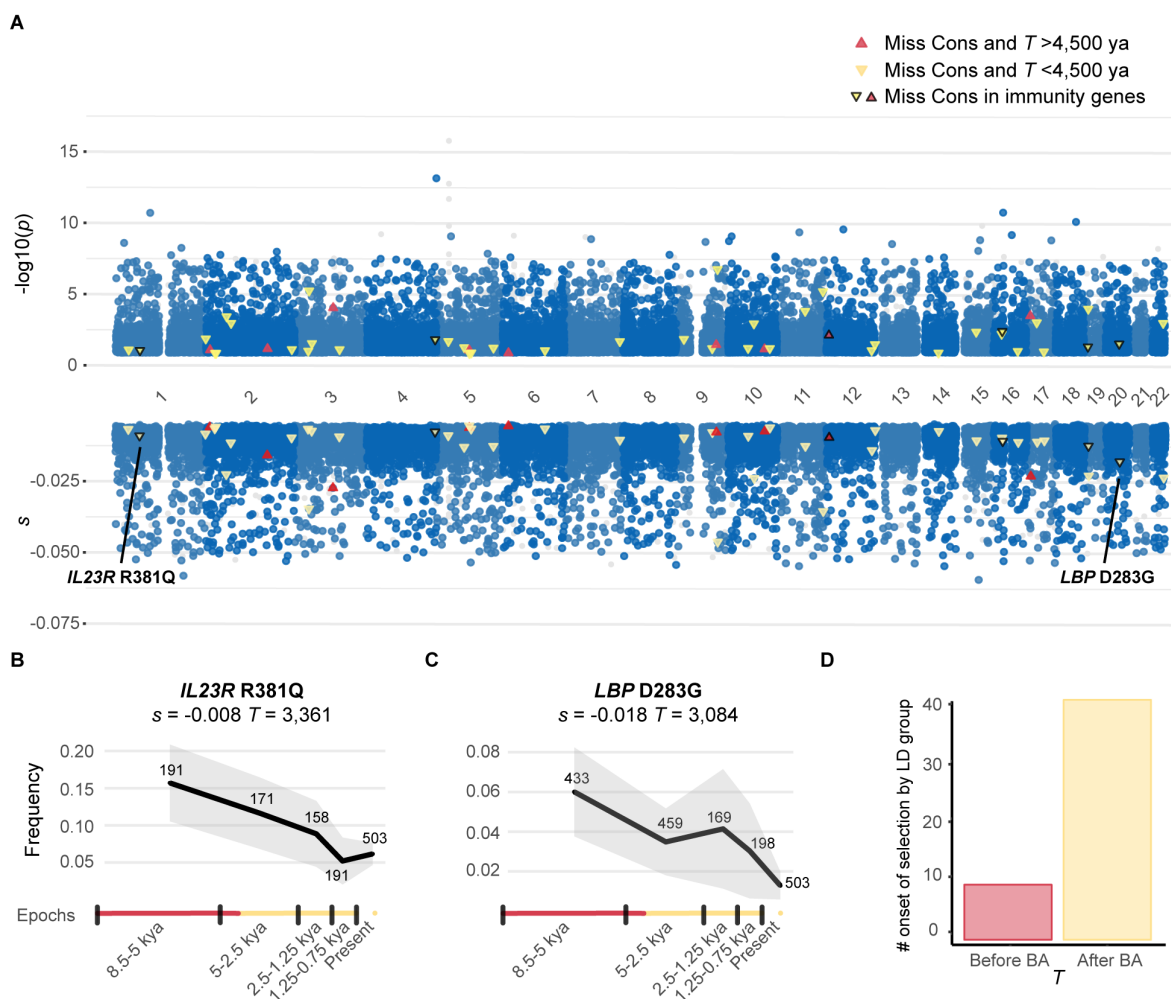


Figure 4. Genome-wide detection of negative selection since the Neolithic

(A) Test for negative selection ($-\log_{10}(p)$) and selection coefficient estimates (s) for each genomic marker in the capture dataset. Red and yellow triangles indicate missense variants at conserved positions (GERP score >4) under negative selection, starting before or after the beginning of the Bronze Age, respectively. The triangles for the six candidate negatively selected variants of immunity genes (IGs) are outlined in black.

(B-C) Frequency trajectory for the strongest candidate negatively selected variant, *LBP D283G*, and the missense variant *IL23R R381Q*.

(D) Number of candidate variants under negative selection beginning before or after the beginning of the Bronze Age.

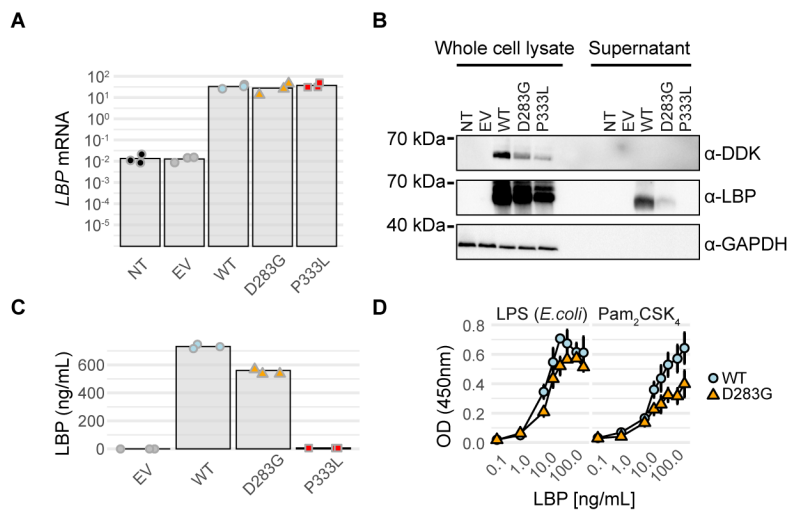


Figure 5. *LBP* D283G is hypomorphic in an overexpression system

(A) RT-qPCR for *LBP* on cDNA from HEK293T cells non-transfected (NT) or transfected with an empty plasmid (EV), or plasmids encoding wild-type (WT), or mutated *LBP*. Dots indicate three independent experiments and the height of each bar their mean values.

(B) Western Blot of whole cell lysates or cell culture supernatants from HEK293T cells either left NT, transfected with an EV, or C-terminally tagged plasmids expressing WT or mutated *LBP*. *LBP* was detected with a polyclonal anti-*LBP* antibody and an antibody against the C-terminal DDK tag. An antibody against GAPDH was used as loading control. The results shown are representative of three independent experiments.

(C) *LBP* concentration in cell culture supernatants from transfected HEK293T cells as measured by ELISA. Dots indicate three independent experiments and the height of each bar their mean values.

(D) Binding of WT or mutant *LBP* collected from cell culture supernatant at increasing concentrations to LPS and Pam₂CSK₄ assessed by a binding assay. Each point represents the mean of two biological replicates ± SD.