

# 1 **1,000 ancient genomes uncover 10,000 years of natural** 2 **selection in Europe**

3 Megan K. Le<sup>1</sup>, Olivia S. Smith<sup>2</sup>, Ali Akbari<sup>3,4,7</sup>, Arbel Harpak<sup>2,5+</sup>, David Reich<sup>3,4,6,7+</sup> & Vagheesh M.  
4 Narasimhan<sup>2,8+</sup>

5  
6 <sup>1</sup> Department of Computer Science, The University of Texas at Austin

7 <sup>2</sup> Department of Integrative Biology, The University of Texas at Austin

8 <sup>3</sup> Department of Genetics, Harvard Medical School

9 <sup>4</sup> Department of Human Evolutionary Biology, Harvard University

10 <sup>5</sup> Department of Population Health, Dell Medical School

11 <sup>6</sup> Howard Hughes Medical Institute, Harvard Medical School

12 <sup>7</sup> Broad Institute of MIT and Harvard

13 <sup>8</sup> Department of Statistics and Data Science, The University of Texas at Austin

14 <sup>+</sup> Co-corresponding authors

## 15 **Abstract**

16 Ancient DNA has revolutionized our understanding of human population history. However, its potential  
17 to examine how rapid cultural evolution to new lifestyles may have driven biological adaptation has not  
18 been met, largely due to limited sample sizes. We assembled genome-wide data from 1,291 individuals  
19 from Europe over 10,000 years, providing a dataset that is large enough to resolve the timing of selection  
20 into the Neolithic, Bronze Age, and Historical periods. We identified 25 genetic loci with rapid changes  
21 in frequency during these periods, a majority of which were previously undetected. Signals specific to the  
22 Neolithic transition are associated with body weight, diet, and lipid metabolism-related phenotypes. They  
23 also include immune phenotypes, most notably a locus that confers immunity to *Salmonella* infection at a  
24 time when ancient *Salmonella* genomes have been shown to adapt to human hosts, thus providing a  
25 possible example of human-pathogen co-evolution. In the Bronze Age, selection signals are enriched near  
26 genes involved in pigmentation and immune-related traits, including at a key human protein interactor of  
27 SARS-CoV-2. Only in the Historical period do the selection candidates we detect largely mirror  
28 previously-reported signals, highlighting how the statistical power of previous studies was limited to the  
29 last few millennia. The Historical period also has multiple signals associated with vitamin D binding,  
30 providing evidence that lactase persistence may have been part of an oligogenic adaptation for efficient  
31 calcium uptake and challenging the theory that its adaptive value lies only in facilitating caloric  
32 supplementation during times of scarcity. Finally, we detect selection on complex traits in all three  
33 periods, including selection favoring variants that reduce body weight in the Neolithic. In the Historical  
34 period, we detect selection favoring variants that increase risk for cardiovascular disease plausibly  
35 reflecting selection for a more active inflammatory response that would have been adaptive in the face of  
36 increased infectious disease exposure. Our results provide an evolutionary rationale for the high  
37 prevalence of these deadly diseases in modern societies today and highlight the unique power of ancient  
38 DNA in elucidating biological change that accompanied the profound cultural transformations of recent  
39 human history.

## 40 **Main**

41 Gene-culture co-evolution—whereby cultural adaptations including technological developments  
42 lead to new lifestyles that change selection pressures—have been widely discussed as a potential major  
43 driver of genetic adaptation<sup>1</sup>. To date, however, there have been few empirical examples, possibility due  
44 to the lack of ancient DNA data in sufficient sample sizes to reveal changes in allele frequencies before  
45 and after cultural change. This deficiency can be addressed with large ancient DNA datasets. Several  
46 central hypotheses have been put forward regarding how human cultural evolution may have driven  
47 human biological evolution<sup>2</sup>.

48  
49 The first hypothesis relates to metabolic traits. The advent of agriculture induced a shift toward  
50 starch-rich and less diverse diets, which would be expected to lead to selection for loci that more  
51 effectively metabolize such diets and address their deficiencies of key nutrients<sup>3</sup>. Farming may have  
52 paradoxically also contributed to food scarcity. In times of plenty and food stability, population growth  
53 occurred at much faster rates than in the hunting and gathering period. However, these larger populations  
54 could also have been subject to periods of famine due to drought, agricultural disease outbreaks, or poor  
55 food distribution which might lead to additional selection for reduced caloric demand or more efficient  
56 energy metabolism.

57  
58 The second hypothesis relates to gene-culture co-evolution associated with immunity. As humans  
59 began living in closer proximity to domesticated animals in the Neolithic, they would have been exposed  
60 to disease affecting those animals. In the Bronze Age and Historical periods, larger increases in  
61 population size as well as population movement occurred due to improved technology and mobility.  
62 However, this would also have radically increased the opportunity for transmission of infectious disease  
63 and pressures on the immune system to more effectively combat them. The immune system has innate  
64 aspects associated with inflammatory processes and adaptive aspects associated with recognition of  
65 specific antigens. Making both these arms of the immune system more active can have deleterious  
66 consequences, for example a propensity to inflammatory processes such as atherosclerosis and  
67 autoimmune disease.

68  
69 A third hypothesis relates to behavior. As population sizes became larger, societies became more  
70 complex, hierarchical, and inter-dependent. Selection could plausibly have occurred on genetic variation  
71 affecting traits such as individualism and sociability. This could plausibly have had impacts on neuro-  
72 psychiatric traits, including autism, schizophrenia, and bipolar disorder.

73  
74 Ancient DNA provides time series data regarding human evolution, making it possible to directly  
75 study past selection by tracking allele frequency changes over time. Such data provides information about  
76 when and where selection occurred that cannot be obtained through analysis of present-day populations  
77 and should make it possible to study the hypotheses about gene culture co-evolution in practice. Until  
78 recently, the large sample sizes required to carry out these studies with high statistical precision have not  
79 been available. The earliest efforts to study natural selection using ancient DNA data have therefore been  
80 limited<sup>4-6</sup>, often focusing on candidate loci or single traits<sup>7-10</sup>. More recent approaches have looked at  
81 selection genome-wide but focus on obtaining evidence of selection across the full range of time from the  
82 Paleolithic leading to modern Europeans<sup>11-13</sup>. Such analyses may miss out on selective events that might

83 be operating only for short bursts in pre-history in response to cultural change. Some other approaches  
84 look at specific time slices in the data but require comparisons with simulations of demographic models  
85 that might not always be available for ancient genomes<sup>14</sup>. Other approaches utilize haplotype approaches  
86 that are unable to precisely identify the targets of selection<sup>14,15</sup>.

87  
88 Here, to examine selection acting across several time intervals in human history, we assembled a  
89 large sample-size time transect from Holocene Europe comprising published data generated using the  
90 same technology that has been the source of more than 70% of published ancient DNA data to date: in-  
91 solution enrichment for about 1.2 million single nucleotide polymorphisms (SNPs). Studying this period  
92 and geographical region is interesting not only from the limited perspective of this place and time, but  
93 also for understanding the processes of natural selection over ten millennia of profound change in human  
94 lifestyle. These include the transition from hunting and gathering to farming, which resulted in major  
95 changes in diet as well as increased population density and proximity to animals. This period also  
96 includes the transition to state-level societies facilitated by metal-working, which led to large population  
97 densities, long-distance exchange of goods, and division of labor. Several ancient DNA studies have also  
98 sequenced bacterial and viral pathogens that caused epidemics in the last few millennia, including  
99 smallpox, the black death, and tuberculosis, suggesting that studying ancient DNA in a time transect  
100 might provide insights into human adaptation to these new infectious diseases<sup>16-18</sup>.

101  
102 The complicated demographic history of human populations, which includes migration and  
103 mixture with neighbors, makes it challenging to determine whether natural selection or population  
104 mixture is the driving force behind changes in allele frequencies that occurred in the past. However, in  
105 Europe, multiple ancient DNA studies have provided excellent models for demographic history<sup>4</sup>. Here,  
106 we identify individual genetic loci as well as sets of alleles whose changes in frequency are inconsistent  
107 with the expectation under neutral evolution and these demographic models, and are therefore suggestive  
108 of selection. Given the large sample sizes spanning this time transect that provide a nearly gapless record  
109 of human populations in Europe in the Holocene, we are further able to estimate the timing of selection  
110 and generate hypotheses about its correspondence with major demographic and cultural changes.

## 111 **A time transect through Holocene Europe**

112  
113 We assembled genome-wide data from a total of 1,291 individuals from Holocene Europe dated  
114 to between 13,000 and 1,000 years before present (BP) ([Supplementary Table 1](#)). We restricted to  
115 individuals with at least 15,000 SNPs<sup>19,20</sup>. We only included unrelated (up to the third degree) individuals  
116 without significant contamination as assessed on the mtDNA or, in males, the X chromosome. We chose  
117 to only analyze data from libraries that were treated with the enzyme uracil-DNA glycosylase (UDG)  
118 prior to library preparation, which reduces characteristic cytosine-to-thymine errors associated with  
119 ancient DNA data, and that were then enriched in-solution at about 1.2 million SNP positions. For  
120 population history analysis, we generated pseudo-haploid calls at every location. For natural selection  
121 analysis, we retained read counts of the reference and alternate allele at every site for our likelihood  
122 calculation of allele frequencies ([Methods](#)). To be conservative and avoid false-positive signals of  
123 selection, we did not impute genotypes at untargeted positions due to potential biases associated with  
124 using a modern reference panel to phase and impute ancient genomes that are of low coverage (median

125 coverage  $\sim 0.9x$ ) and could have different haplotype structure<sup>21</sup>. To avoid additional biases associated with  
126 misestimating allele frequencies with heterogeneous data, we did not include ancient shotgun data or  
127 modern data in our analysis.

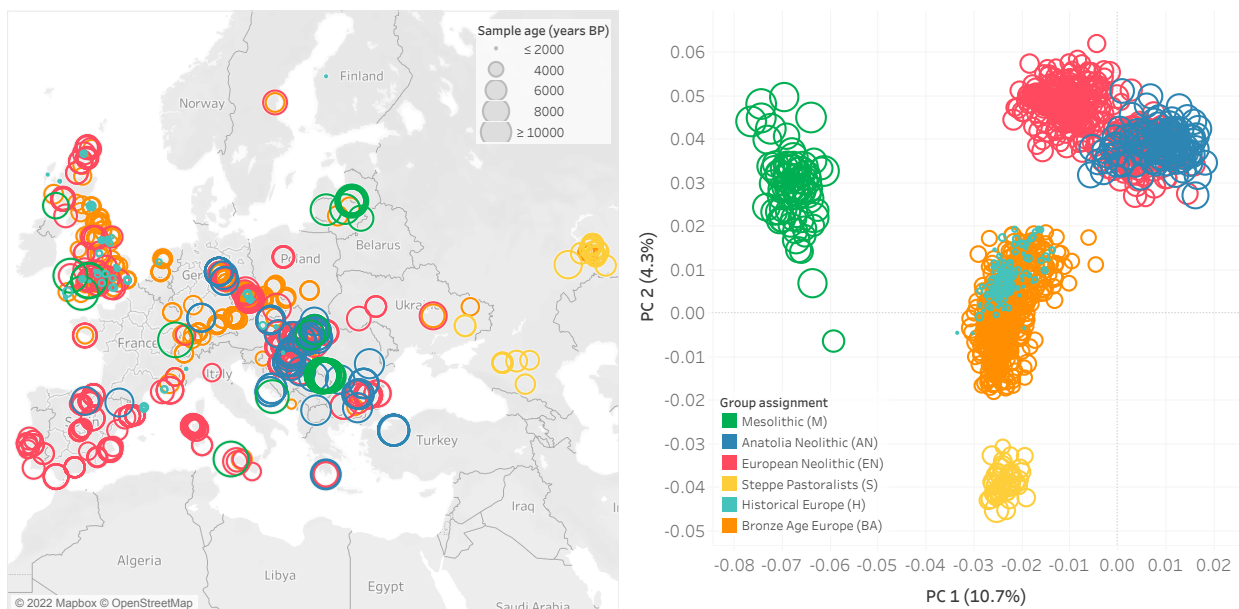
128  
129 We leveraged a model of the demographic history of Europe over the past 10,000 years that has  
130 been inferred from ancient DNA studies<sup>4,5,22–29</sup>. Broadly, these studies conclude that most Europeans in  
131 this period derive the great majority of their ancestry from three primary ancestral sources which came  
132 together in the course of two major demographic transitions corresponding to significant shifts in the  
133 archeological record. The first is the transition from hunting and gathering to farming, which was  
134 accompanied by a major population transition in central and western Europe. In this period, ancestry from  
135 the Mesolithic inhabitants of central and western Europe was largely displaced by ancestry from farmers  
136 whose ancestors originated in Anatolia, and who were amongst the first peoples in the world to use  
137 agriculture a few thousand years before. This economic and demographic shift began in southeastern  
138 Europe after around 6,500 BCE but had spread to the far reaches of Europe as well as Britain by 4,000  
139 BCE. The second major demographic transition occurred during the shift from the Neolithic to the Bronze  
140 Age with the arrival of Steppe Pastoralists from the Eurasian Steppe. In the subsequent millennia leading  
141 to the Historical period, there were subtle shifts in the proportion of Steppe ancestry that largely arose  
142 from the homogenizing of populations with different Steppe ancestry proportions.

143  
144 We assigned individuals to different groupings based on  $f_4$ -statistics, time period (based on direct  
145 radiocarbon dates or well understood archaeological contexts), and geographic location. We removed  
146 individuals that were outliers from each time period and were found to have atypical ancestry of that  
147 period based on  $f_4$ -statistics. The groupings of individuals were:

- 148  
149 M (n=73) Hunter-Gatherers from Europe (abbreviated by the first letter of Mesolithic). The majority  
150 of these individuals were from Eastern Europe, all of whom had no evidence of any  
151 admixture from Anatolian Farmers dated to a mean age of  $\sim 8,600$  BP.  
152  
153 AN (n=111) Anatolians Neolithic farmers and their European direct descendants (abbreviated as the first  
154 letters of Anatolia Neolithic). These individuals were from early agricultural settlements,  
155 with a mean age of  $\sim 7,400$  BP, primarily from western Anatolia, the Balkans, Aegean, and  
156 Central Europe. They all had little to no evidence of mixture from European Hunter-  
157 Gatherers.  
158  
159 EN (n=398) European Farmers from the Middle to Late Neolithic (abbreviated as the first letters of  
160 European Neolithic). These individuals were from across Europe, are dated to a mean age  
161 of  $\sim 5,400$  BP, and are modeled as mixtures of Mesolithic European Hunter-Gatherer and  
162 Anatolian Neolithic ancestry. Sampled across a large geographic region, they have  
163 differences in their hunter-gatherer ancestry proportion, but we average across them to  
164 obtain an allele frequency estimate of each position in that time period.  
165  
166 S (n=47) Steppe Pastoralists (abbreviated as the first letter of Steppe). These individuals are from the  
167 Yamnaya and Afanasievo cultures of Central Asia dated to  $\sim 4,800$  BP. They are genetically  
168 homogenous and have little to no mixture from European Farmers.

169  
170 BA (n=517) Bronze Age Europeans (abbreviated as the first letters of Bronze Age). These individuals  
171 from the Bell Beaker and succeeding cultures of Western and Central Europe are modeled  
172 as having formed as a mixture between the incoming Steppe Pastoralists and European  
173 Farmers, with a mean age of ~4,000 BP.  
174  
175 H (n=145) Historical era individuals from the Roman and Late Antique periods, primarily from Britain  
176 (abbreviated as the first letter of Historical). These individuals were genetically relatively  
177 homogenous, dated to a mean age of ~2,000 BP, and we excluded individuals with ancestry  
178 from additional population sources that began to have major impacts in eastern and  
179 southeastern Europe from the Bronze Age onward. Thus we did not include in our analysis  
180 Scythians and Sarmatians, people likely descended from migrations of Uralic speakers into  
181 Hungary and Fennoscandia, and people with Iranian or Caucasus related ancestry whose  
182 ancestry occurs in relatively high proportion in the Mediterranean, especially in Greece and  
183 southern Italy ([Fig. 1](#), [Extended Data Fig. 1](#)).  
184

185 Full lists of all individuals, their assignments, and additional metadata can be found in [Supplementary](#)  
186 [Table 1](#).  
187



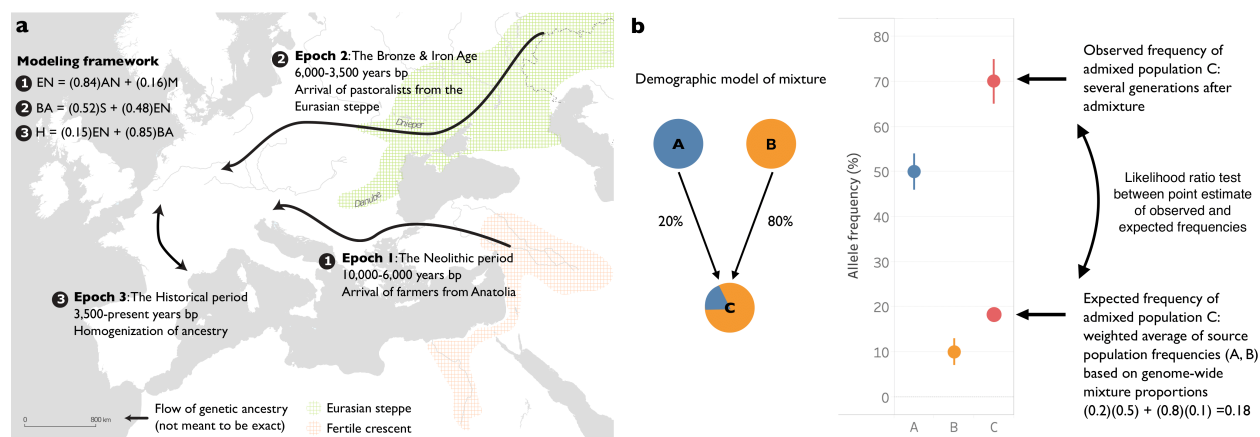
188  
189 **Fig. 1: Geographic and temporal distribution of analyzed individuals. a,** Geographic locations and  
190 group assignments (in color) for all individuals along with sample age represented by the size of the  
191 circular points. **b,** Principal Components Analysis of samples with the same grouping and coloring  
192 scheme as in **a**.

193  
194 To model these demographic changes with our combined dataset, we used *qpAdm*, which  
195 evaluates demographic fit of a target population to various source populations genome wide and then  
196 estimates proportions of ancestry for each source<sup>22</sup>. We divided the roughly 10,000 year period into three  
197 non-overlapping time epochs, each of which spans just over 3,000 years: (1) the transition from hunting



198 and gathering to farming (the Neolithic period), (2) the transition to the Bronze Age, and (3) the transition  
199 to large-scale state-level societies during the Historical period. To capture the major sources of admixture  
200 in epoch (1), we modeled European Farmers (EN) as a 16:84% mixture of European Hunter-Gatherers  
201 and Anatolian Farmers. For epoch (2), we modeled Bronze Age Europeans as a 48:52% mixture of  
202 European Farmers and Steppe Pastoralists. For epoch (3), we modeled Historical European samples as a  
203 15:85% mixture of Bronze Age Europeans and earlier Neolithic Farmers (Fig. 2).

204



205

206 **Fig. 2: Description of approach to detect selection.** a, Demographic changes in Europe over the past  
207 10,000 years are driven by admixture between various populations across three major time  
208 epochs/periods. b, Visual description of our methodology. Under neutrality, the expected frequency of an  
209 allele is the weighted average of the source population frequencies. Large deviations from this genome-  
210 wide expectation can be identified as evidence for selection. In this case, the frequency of the allele in  
211 population C has risen to a frequency that is ~50% above the expected frequency based on mixture  
212 proportions and suggests that natural selection has elevated the frequency of this allele in the time period  
213 since admixture.

## 214 A scan for selection at individual loci

215 To identify candidate selected loci in our dataset, we used our three epoch model and applied a  
216 method that utilizes the admixture events that occurred in each epoch. Under neutrality, the allele  
217 frequency of an admixed population is expected to be the weighted average of the allele frequencies in the  
218 source populations that contributed to the admixture. Significant deviations from this expectation suggest  
219 that natural selection has acted at a particular locus (Fig. 2). After correcting for inflation of the test  
220 statistic independently in each of the three epochs, we used a cutoff of  $5 \times 10^{-8}$  as a genome-wide  
221 significance threshold. This is a common significance threshold in genome-wide association studies  
222 (GWAS)<sup>30</sup>, and also roughly corresponds to a *P* value of <0.05 after Bonferroni correction for a 1.2  
223 million SNP target set (Methods). Previous work has examined the impact of sample size, the strength of  
224 selection, the time that selection has acted, mis-specification of the mixture proportions, and additional  
225 unmodeled mixtures in detecting selection using this method and has shown that, after applying a  
226 correction for genomic inflation, these issues result in reduced power but not an increased rate of false-  
227 positives<sup>4,31</sup>. Additional work on the same method with slightly different statistical formulation has  
228 confirmed this robustness to deviations from the model<sup>32</sup>. To further study the effect of model  
229 misspecification as well as the effect of sample size on our power to detect signals, we carried out two

230 additional analyses. First, we examined our model's robustness to mis-estimates of the admixture  
231 proportions and found that deviations on the order of 15% resulted in little reduction in power ([Extended](#)  
232 [Data Fig. 2](#)). Second, we found that reduced sample size below 80% of the dataset size used for analysis  
233 has a major effect on power to detect selection signals ([Extended Data Fig. 3](#)).

234  
235 Following a previous strategy used to mitigate false-positives in ancient DNA scans of selection  
236 due to biases affecting the sequences aligning to a particular variant, we considered loci to be candidates  
237 for selection if at least 3 alleles within 1 Mb of each other and the causal gene significantly deviated from  
238 their expected frequency<sup>4</sup>. This distance is also in agreement with a recent study examining the optimal  
239 window size for linking GWAS-associated SNPs to causal genes<sup>33</sup>. To determine if functional categories  
240 of genes were significantly associated with selection signals, we carried out enrichment analysis using  
241 FUMA<sup>34</sup>, which maps SNPs to genes and performs gene set enrichment analysis for GWAS and GO  
242 annotations incorporating LD information as well as gene matching by length and conservation scores  
243 ([Methods](#)).

244

## 245 **25 time-resolved candidate signals of natural selection**

246

247 Across all epochs, we discovered a total of 25 regions containing alleles with frequency changes  
248 that significantly deviated from genome-wide expectation ([Fig. 3](#), [Extended Data Fig. 4](#), [Extended Data](#)  
249 [Fig. 5](#), [Extended Data Fig. 6](#), [Table 1](#)). The only locus that contained alleles with significant evidence of  
250 selection across all time periods was the Major Histocompatibility Complex (HLA) on chromosome 6,  
251 which encodes cell surface proteins that are a critical part of the human adaptive immune response.

252

## 253 **Candidate selective signals that were most intense during the early phases of** 254 **the transition to agriculture**

255

256 In the first epoch representing the transition from hunting and gathering to farming, we  
257 discovered individual signals that were plausibly associated with a transition to a high starch,  
258 carbohydrate heavy diet, to which the genomes of the two ancestral populations were not yet fully  
259 adapted.

260

261 First, we observed several alleles at the *FTO/IRX3* locus, the locus that has the largest effect on  
262 predisposition to obesity in humans<sup>35</sup>. Reduction in gene expression of this gene has shown 30%  
263 reduction in weight in humans and model organisms<sup>36</sup>. The region and variants that were significant in  
264 our scan are in the promoter region of the *IRX3* gene and are in high LD with variants that are expression  
265 quantitative trait loci (eQTLs) in human adipose/subcutaneous tissues reported by the GTEx  
266 consortium<sup>37</sup>. *IRX3* expression is known to increase body weight<sup>36</sup>, and variants that decrease the  
267 expression of *IRX3* increased in the Neolithic transition, which suggests that there may have been  
268 selection for reduced body weight specifically during this time.

269

270 We also found alleles that were significant in the gene *PTPRV*. Mice homozygous for a knock-  
271 out allele at this gene exhibit increased resistance to diet-induced obesity and decreased circulating  
272 glucose levels<sup>38</sup>. Other studies have shown that *PTPRV* also contributes to *FTO*'s role in adipogenesis  
273 with simultaneous knockdown of both genes restoring adipogenesis activity that is lost when just *FTO*  
274 alone was knocked down<sup>39</sup>. Selection for these variants affecting adipogenesis could be adaptive in the  
275 course of an economic transition between a hunting and gathering lifestyle to a farming-based lifestyle,  
276 which would have involved a greater reliance on starch-based diets and different patterns of feast-and-  
277 famine.

278  
279 We detected another candidate in the gene *ENSA*, which acts as a stimulator of insulin secretion  
280 by interacting with the protein encoded by *ABCC8*, a sulfonylurea receptor which plays a key role in the  
281 control of insulin release in pancreatic beta cells. We also observed variants with likely similar function in  
282 the regulatory region of the gene *MAF* (rs4073089) which promotes pancreatic development and regulates  
283 insulin gene transcription<sup>40</sup>. Another candidate of selection is on the missense variant rs6265 that occurs  
284 at around 19% frequency in modern Europeans on the *BDNF* (Brain-derived neutrophil factor) gene,  
285 which has been associated with regulation of body weight and has a mechanistic role in Type 2-diabetes  
286 in humans and model organisms<sup>41,42</sup>.

287  
288 Several signals of selection in this period are associated with immune-related functions. We  
289 detect a signal in the gene *FUT2*, the human secretor locus that encodes  $\alpha(1,2)$ -fucosyltransferase, and  
290 determines the secretion status of the ABO blood group antigens. Individuals homozygous for the *FUT2*  
291 non-secretor genotype are resistant to infection with norovirus<sup>43</sup>, suggesting that individuals homozygous  
292 for non-secretor status may be unable to mediate host-microbe interactions. The variants that are  
293 significant in *FUT2* have also been associated with plasma B12 levels<sup>44</sup>, a vitamin that is largely  
294 unavailable from plant-based food sources—in particular, it is virtually absent in wheat and barley, which  
295 make up the bulk of the Neolithic agricultural package—but plentiful in animal products.

296  
297 Another significant signal is at the Interleukin 1 receptor, type II (*IL-1R2*) which is expressed on  
298 lymphoid and myeloid cells including monocytes, neutrophils, macrophages, B, and T cells, and has been  
299 implicated as a susceptibility locus for a number of autoimmune diseases<sup>45</sup>. We also found signals of  
300 selection at alleles in the gene *PPIL2* which encodes a cyclophilin, a class of proteins that bind to  
301 ciclosporin (cyclosporin A), an immunosuppressant which is used to suppress rejection after internal  
302 organ transplants. *PPIL2* and other cyclophilins are also recruited by the Gag polyprotein during HIV-1  
303 infection, and its incorporation into new virus particles is essential for HIV-1 infectivity<sup>46</sup>, suggesting that  
304 selection at this locus may reflect selection against HIV-like retroviruses. Another gene that is under  
305 selection is *CACNB1*, a regulator of T-cell function. Mice lacking in the *CACNB1* gene have been shown  
306 to be immune-deficient to viral infection<sup>47</sup>.

307  
308 Finally, we detect evidence for selection at variants in *FAM49B*. *In-vitro* as well as *in-vivo* studies  
309 have recently shown this gene is a T-cell regulator and confers host resistance to *Salmonella* infection<sup>48,49</sup>.  
310 *FAM49B* negatively regulates *RAC1* signaling, thereby attenuating processes such as macropinocytosis,  
311 phagocytosis, and cell migration. This enables the protein to counteract *Salmonella* at various stages of  
312 infection, including bacterial entry into non-phagocytic and phagocytic cells as well as phagocyte-  
313 mediated bacterial dissemination<sup>48</sup>. Evidence for *Salmonella enterica* adaptation to the human host



314 through rise of specific pathogenic mutations has been detected through bacterial sequencing of ancient  
315 DNA time transects from Europe and has been timed to the Neolithic period<sup>50</sup>. Our observation of  
316 candidate regions of selection in humans (the host) at this locus during the same time period is compatible  
317 with human-pathogen coevolution at a time of major cultural change.

318

319 Consistent with the associated gene function of individual variants, we found an enrichment of  
320 candidates near genes related to fatty acid metabolism and digestion, serum metabolite levels, and  
321 diseases of the digestive system such as Crohn's disease and ulcerative colitis ([Supplementary Table 2](#)).

322

## 323 **Candidate selective signals most intense during the transition to the Bronze** 324 **Age**

325

326 In the Bronze Age, we do not detect evidence for continued selection on candidate variants that  
327 are directly associated with a change in diet. Instead, we found evidence for selection at or near genes that  
328 affect skin and eye pigmentation.

329

330 The strongest signal is at the allele rs16891982, in the gene *SLC45A2*, which is known to play a  
331 major role in light skin pigmentation, and for which there has been previous evidence for selection<sup>51</sup>. The  
332 second strongest signal based on our analysis is in the allele rs11636232 in *OCA2/HERC2*, which is a  
333 primary determinant of light eye color in Europeans<sup>4,52</sup>.

334

335 As in the Neolithic period, we also found several candidate genes involved in immunity beyond  
336 those seen in the HLA region. We observed selection at rs10797666 in the major histocompatibility  
337 complex class I-related gene *MRI*, which is an immune sensor of microbial ligands, including  
338 *Mycobacterium tuberculosis*, *Streptococcus pyogenes*, *Salmonella enterica* and *Escherichia coli*<sup>53</sup>. We  
339 also find evidence of selection at a number of alleles in genes in the killer-cell immunoglobulin-like  
340 receptor locus (*KIR* gene family), which are expressed on the cell membrane of natural killer cells. *KIR*  
341 receptors interact with major histocompatibility molecules to detect pathogen-infected cells and have a  
342 crucial role in host defense. This locus is highly polymorphic across human populations worldwide, and  
343 diversity at this locus has been correlated with pathogen load across populations<sup>54</sup>. We also find evidence  
344 for selection at the *MRGPRX3-4* locus, which includes genes that are key physiological and pathological  
345 mediators of itch and related mast cell-mediated hypersensitivity reactions, as well as potential targets to  
346 reduce drug-induced adverse reactions<sup>55,56</sup>. A final immune-related candidate is the gene *MARK3*, which  
347 is a host protein that is one of the key interactors with SARS-CoV-2 and is important in mediating the  
348 maladaptive host response to COVID-19. The allele under selection appears to be linked to a lead signal  
349 for monocyte count, which has now been shown to be important in the pathology of COVID-19<sup>57-59</sup>.  
350 There is direct ancient DNA evidence for pathogen infection being a major source of mortality in the  
351 Bronze Age. The earliest evidence for *Yersinia pestis* infections in Europe ascertained from ancient DNA  
352 comes from the Bronze Age at times particularly close to or after the arrival of pastoralists from the  
353 Eurasian Steppe, from where both of these pathogens have been recovered from humans several millennia  
354 prior to their first evidence in Europe<sup>60,61</sup>. Thus, pathogens entering Europe along with Steppe Pastoralists  
355 in the Bronze Age could have been a driving force behind changes in these immune related genes.

356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371

We also observed significant frequency deviation from the expectation due to genetic drift in alleles lying on several genes related to cardiovascular disease. One candidate is in the angiotensin gene, *AGT*, which causes vasoconstriction and increases blood pressure<sup>62</sup>. The protein encoded by *AGT* is a frequent antagonist in drugs that treat heart disease. Additionally, we also observed another locus that reached significance, rs915843, which is a missense allele in *ABCG1*, a gene that controls tissue lipid levels and the efflux of cellular cholesterol to HDL<sup>63</sup>.

Finally, we observed candidates in genes where mutations have been linked to reproduction. One of our significant variants is at rs7188473, a splice mutation in the gene *HYDIN*. Homozygous carriers of this allele suffer from primary ciliary dyskinesia-5, which affects sperm motility and leads to male infertility<sup>64</sup>.

More broadly, across this epoch, we find a statistically significant enrichment of signals near genes related to skin, hair, and eye pigmentation ([Supplementary Table 3](#)).

## 372 **Candidate selective signals most intense during the transition to the Historical** 373 **period**

374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397

The variant with the strongest significant deviation from expectation is in the *LCT* gene, which is responsible for conferring the ability to digest lactose in adulthood in Europeans. This is also consistently the strongest signal of natural selection detected in scans in modern Europeans, and in line with findings in previous publications<sup>65</sup>, this allele appears to have experienced its major change in frequency primarily in the past few thousand years, and not during the Bronze Age when the allele was first introduced in central and western Europe.

We found a selective signal in *DHCR7* (the focal SNP that deviates most from expectation is in an enhancer region several kb upstream of the gene), which governs availability of 7-dehydrocholesterol for conversion to vitamin D<sub>3</sub> by the action of sunlight on the skin. Milk is rich in 7-dehydrocholesterol, suggesting that selection on this locus as well as *LCT* might have been related to the need for increased production of vitamin D<sup>66</sup>. This locus has also been linked to several auto-immune diseases. We also detect evidence for deviation in allele frequency from expectation in the missense variant rs653178 in the gene *SH2B3*. This allele doubled in frequency from the Bronze Age to the Historical period and is a major risk locus for Celiac disease. The allele we identify as a signal of selection has recently also been shown to be fine-mapped in a GWAS for vitamin D binding<sup>67</sup>. Functional investigation of the effect of the *SH2B3* genotype in response to lipopolysaccharide and muramyl dipeptide revealed that carriers of the *SH2B3* allele showed stronger activation of the *NOD2* recognition pathway. This suggests that *SH2B3* also plays a role in protection against bacterial infection<sup>68</sup>.

The second strongest signal was in the gene encoding the toll-like receptor locus *TLR*, which is expressed on the membranes of leukocytes. Variants at this locus have been associated with host immune

398 response against a variety of diseases, including *Helicobacter pylori* infection, leprosy, plague, and  
399 tuberculosis.

400

401 We detect continued change in frequencies of variants at the *SLC45A2* gene, driving the light-  
402 pigmentation allele under selection to near fixation in Historical samples.

403

404 Finally, we found candidate variants on the genes *FADS1* and *FADS2*, which are involved in fatty  
405 acid metabolism. Variation at this locus is associated with plasma lipid and fatty acid concentration. The  
406 most significant SNP (rs174550) at this locus has also been associated with decreased triglyceride levels,  
407 and perhaps selection at this locus could reflect transition to a starch-heavy diet. This locus has also  
408 experienced independent selection in non-European populations and is likely to be a critical component of  
409 adaptation to different diets. In agreement with previous work suggesting that natural selection acted on  
410 these alleles only after the Neolithic transition<sup>7</sup>, in our analysis we see that the major signal of selection at  
411 this locus is focused on the most recent epoch.

412

413 In this period we also find statistically significant enrichment in gene sets involved in a large  
414 variety of traits, from anthropometric traits such as BMI and autoimmune disease like Crohn's and  
415 ulcerative colitis, to hormone-related disorders like hyperthyroidism, blood biomarkers such as serum  
416 metabolite and cholesterol levels, and cardiovascular disease traits ([Supplementary Table 4](#)).

417

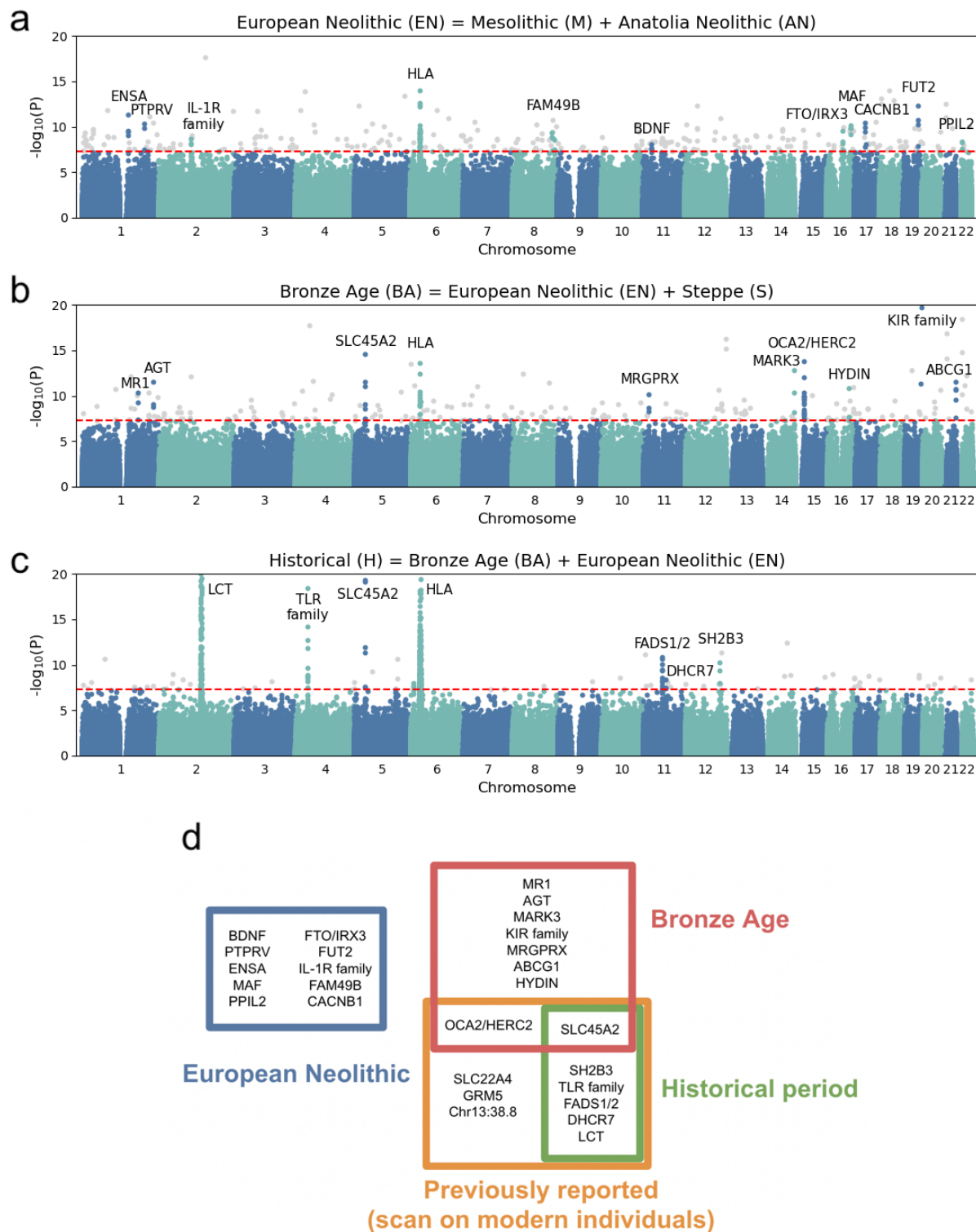
## 418 **Timing of selection of alleles**

419

420 By separating our analysis into different epochs, we were able to examine overlaps in candidate  
421 alleles across epochs as well as a previous scan examining modern Europeans from the 1000 Genomes  
422 Project. Outside of the HLA region, we found no overlaps of any of the loci discovered in the Neolithic  
423 period with any of the other epochs. All of the candidates we discovered in the Historical period had also  
424 been discovered in a scan comparing modern Europeans to ancient Europeans<sup>4</sup>. As expected, this scan  
425 was largely blind to selection during the European Neolithic, showing the value of direct comparison of  
426 groups of ancient DNA samples to study the selection that occurred in this time ([Fig. 3](#)).

427

428



429

430 **Fig. 3: Signals of natural selection in three epochs.** a-c, Manhattan plots of  $P$  values for the likelihood  
 431 ratio test for selection (Methods; Fig. 2) in the Neolithic, Bronze Age and Historical period. The red line  
 432 shows the genome-wide significance threshold ( $5 \times 10^{-8}$ ). d, Venn diagram showing the overlap of  
 433 variants seen in each epoch and the variants that were previously published (Mathieson et al.<sup>4</sup>) from a  
 434 scan on present day humans. HLA, which was previously published and seen in all epochs, is not shown.

435

Epoch	Gene Name	Chr	Pos	Category	Function
Neolithic period	BDNF	11	27679916	Diet	Body weight, Appetite
	PTPRV	1	202143512	Diet	Obesity, Circulating glucose levels
	ENSA	1	150596411	Diet	Insulin secretion
	MAF	16	80036594	Diet	Insulin secretion
	FTO/IRX3	16	54231250	Diet	Obesity
	FUT2	19	49206603	Blood group, Immune	ABO secretor status, Resistance to norovirus infection
	IL-1R family	2	102824201	Immune	Interleukin receptor, Inflammatory response
	FAM49B	8	130981907	Immune	Resistance to Salmonella infection
	CACNB1	17	37355093	Immune	T-cell function and immunity to infection
Bronze Age	PPIL2	22	22027348	Immune	Resistance to HIV infection
	SLC45A2	5	33951693	Pigmentation	Light skin pigmentation
	OCA2/HERC2	15	28386626	Pigmentation	Light eye color
	MR1	1	181018799	Immune	MHC locus, Bacterial sensing
	MRGPRX	11	18167630	Immune	Allergen itch and hypersensitivity reactions
	MARK3	14	103867320	Immune	Host response to SARS-Cov-2, Monocyte count
	KIR family	19	55315616	Immune	Natural killer cell pathogen detection
	AGT	1	230854999	Cardiovascular	Blood pressure regulator
	ABCG1	21	43679554	Cardiovascular	HDL level regulator
Historical Period	HYDIN	16	71096248	Reproductive	Sperm motility, Infertility
	LCT	2	136608646	Vitamin D	Lactase persistence
	DHCR7	11	71153459	Vitamin D	7-dehydrocholesterol for conversion to vitamin D3
	SH2B3	12	112007756	Vitamin D, Immune	Vitamin D-binding, NOD2/bacterial signaling
	FADS1/2	11	61571478	Diet	Lipid metabolism
	SLC45A2	5	33951693	Pigmentation	Light skin pigmentation
	TLR family	4	38776107	Immune	Macrophage pathogen detection

436

437 **Table 1: Summary of genes with evidence of selection during the three epochs in Europe.** The HLA  
438 region, which appears to be under selection in all epochs, is not shown.

## 439 Polygenic selection

440 Evidence from contemporary genomes suggests that in recent human history, monogenic  
441 selective sweeps are rare<sup>69,70</sup>. Further, theoretical arguments and some empirical evidence in the last  
442 decade suggest that polygenic adaptation may be the more frequent mode of selection<sup>52,71–74</sup>. Therefore, to  
443 complement the picture we obtain of selection in the last 10,000 years from the monogenic genome scan,  
444 we sought to test for polygenic selection on complex traits. We did so by integrating the same signal of  
445 deviation from expected admixture proportions with trait-associations from genome-wide association  
446 studies (GWAS).

447

## 448 Mitigating confounders of our tests for polygenic selection

449

450 Despite the theoretical appeal of screens for polygenic selections, clear evidence for polygenic  
451 selection has been elusive due to challenges in application and interpretation<sup>71,72,75–79</sup>. Here, we take an  
452 approach that offers more robustness against a major challenge: the portability of GWAS associations  
453 from contemporary GWAS to ancient samples.



454

455 GWAS-based estimates may often be poor or even biased with respect to individuals in  
456 populations different from the groups in which the GWAS was carried out, due to differences in ancestry,  
457 environment, or other characteristics. This can lead to biased or uninterpretable results<sup>80,81</sup>. This also  
458 applies for porting modern GWAS associations to ancient genomes. There are several reasons for poor  
459 portability. A major problem, which can lead to systematic biases, is uncorrected population stratification  
460 (axes of ancestry that correlate with a trait) in GWAS. Regardless of the cause of the correlation with the  
461 trait, numerous alleles that tag these ancestry axes may still associate with the trait. This problem may be  
462 amplified as GWAS sample sizes increase and many small effects become more statistically  
463 significant<sup>75,76,82</sup>.

464

465 We took two measures that are expected to reduce statistical power but increase robustness to  
466 population stratification. First, for our primary analysis we use GWAS summary statistics (for 38 case-  
467 control and 177 quantitative traits) from the Biobank of Japan (BBJ), rather than summary statistics based  
468 on GWAS in Europeans with higher sample sizes. Since West Eurasians and East Eurasians largely split  
469 from a common ancestral population more than about 40,000 years ago, the population structure present  
470 in the BBJ sample is expected to be uncorrelated with the main axes of population structure among  
471 Europeans. In addition, as suggested by Chen et al.<sup>83</sup>, linkage disequilibrium (LD) and allele frequency  
472 differences between the BBJ sample and the different ancient European target samples are mediated  
473 through a common ancestral population and thus should be approximately equal.

474

475 Following Chen et al.<sup>83</sup>, we evaluated residual stratification by examining the association between  
476 effect sizes estimated from each GWAS panel and PC loadings conducted on a set of diverse West  
477 Eurasian populations that reflects the various ancestry sources in Europe<sup>23</sup>. The first PC separates Steppe  
478 Pastoralists from Western European Hunter Gatherers, and the second PC separates Anatolian and Iranian  
479 Farmers from Steppe Pastoralists and Western European Hunter-Gatherers. To measure the impact of  
480 uncorrected stratification on estimated effect sizes for a set of ascertained trait-associated variants, we  
481 regressed the SNP effect sizes on the PC loadings of each SNP. We carried out this analysis on 38  
482 quantitative traits for which we had GWAS summary statistics that were matched between the European  
483 and Japanese Biobanks. After controlling for multiple hypothesis correction, only a single PC (PC10) in a  
484 single trait, total bilirubin, was significantly associated with effect size using the BBJ GWAS results, but  
485 24 different PCs across 14 different traits were significantly associated with PCs using the UK Biobank  
486 (UKB) GWAS results ([Supplementary Table 5](#), [Supplementary Table 6](#), [Extended Data Fig. 8](#)),  
487 suggesting that residual stratification remains an issue with using GWAS from the UKB but not from  
488 BBJ.

489

490 A second measure that we took to increase the portability of BBJ-based associations to the target  
491 populations at the expense of statistical power was to limit our analysis to significant associations  
492 (GWAS  $P < 1 \times 10^{-6}$ ), as well as to the sign of the effect on the trait alone. Across a large number of  
493 matched traits between BBJ and UKB<sup>84,85</sup>, we found that >95% of all significantly-associated alleles have  
494 the same direction of effect. In contrast, effect sizes between BBJ and the UKB were only correlated at  
495 ~70% ([Supplementary Table 7](#)), and so effect sizes appear to be less portable. Second, evidence from  
496 model organisms, particularly from plants where over 300 studies have been conducted with isogenic  
497 lines grown across different environmental conditions, suggest that across a range of traits, while the

498 effect size of QTLs vary, effect directions are almost entirely conserved (98% consistency in effect  
499 direction across all comparisons tested)<sup>86</sup>.

500

501 In summary, we tested for polygenic selection in a way that reduces statistical power but is more  
502 robust to confounding. We identify a set of variants significantly associated with a trait in a Japanese  
503 sample, whose population structure and thus potential for population stratification is uncorrelated to that  
504 in our ancient European sample. We then carried out a test for selection by comparing the chi-squared  
505 statistic of trait-associated alleles, considered alongside with the direction of change in frequency, to those  
506 of random SNPs ([Methods](#)).

507

## 508 **Signals of polygenic selection that differ across epochs of European history**

509

510 In each of the three epochs, we tested for selection by comparing our selection statistic in trait-  
511 associated SNPs to a distribution of matched controls resampled 10,000 times. The control was composed  
512 of an equal number of SNPs matched for deciles of derived allele frequency, recombination rate<sup>87</sup>, and  
513 intensity of background selection<sup>88</sup> ([Methods](#)). We restricted the 220 total traits in the Biobank of Japan to  
514 only those that had at least 20 SNPs significantly ( $P < 1 \times 10^{-6}$ ) associated with the trait. To assess the  
515 directionality of genetic change, we polarized our selection statistic to the direction of the effect allele in  
516 the GWAS (polarized chi-squared statistic) and asked whether the mean observed polarized statistic for a  
517 trait was below the 2.5 or above the 97.5 percentiles of all the matched null samples. In total, we  
518 identified 39 traits that reach this level of significance across the three epochs ([Fig. 4](#), [Supplementary](#)  
519 [Table 8](#)). In carrying out this analysis, we checked that null distributions for all traits were approximately  
520 normally distributed ([Extended Data Fig. 10](#)), and that we had enough variants to prevent the same variant  
521 being sampled multiple times ([Supplementary Table 9](#)).

522

523 In the hunter-gatherer to farming transition, we detected evidence for selection favoring body  
524 mass-decreasing and cholesterol-increasing alleles. We also found evidence for selection on traits related  
525 to blood cell biomarkers such as platelet and hemoglobin concentration.

526

527 In the Bronze Age, we detected evidence for selection on alleles associated with some disease  
528 endpoints such as hepatitis and ulcerative colitis, as well as several blood and blood-pressure-related  
529 traits. We also observed evidence for selection favoring alleles increasing triglyceride levels.

530

531 The vast majority of polygenic adaptation signals we observe were in the Historical period,  
532 though several of the traits we identify could be genetically correlated. Importantly, we detect selection  
533 favoring alleles that increase rates of heart disease due to myocardial infarction (heart attacks). In  
534 addition, we detected evidence for selection on alleles associated with related phenotypes like angina, as  
535 well as biomarkers for cardiovascular disease and cardiovascular prescriptions such as beta-blockers.  
536 Finally, we observed signals of selection favoring alleles that increase risk for several common auto-  
537 immune diseases.

538

539 To investigate alternative ways of carrying out these analyses, we repeated these studies by  
540 including effect sizes in the polarized chi-squared statistic; we found that the majority of our signals were

541 replicated using the magnitude of effect as well as direction ([Fig. 4, Supplementary Table 10](#)). The cases  
542 of non-replication were largely sub-significant (for example, ~2% vs ~5% for body weight). We also  
543 carried out an additional sensitivity analysis by carrying out the same scan, but this time removing SNPs  
544 that were in the lowest 27.5% of the chi-squared statistic distribution ([Extended Data Fig. 11](#)). These are  
545 positions where the direction of frequency change may have been mis-estimated. Therefore, by restricting  
546 to sites that are deviating more significantly from expectation, we might increase our confidence in our  
547 estimates of the effect direction but reduce power as the number of SNP positions that we could use in the  
548 analysis would decrease. This analysis replicated 90% of the original signals across all epochs  
549 ([Supplementary Table 11](#)). Importantly, we did not detect evidence for natural selection on height,  
550 perhaps because of lack of power or because previous analyses may have been confounded by population  
551 stratification<sup>4,89</sup>. A final sensitivity analysis we carried out was to repeat the polygenic selection analysis  
552 using summary statistics that were identified from a large consortia study of 25 phenotypes carried out  
553 using a within-sibling GWAS design<sup>90</sup>. In theory, family-based GWAS designs can control for  
554 demographic and indirect genetic effects, but even with the relatively large number of samples, only 6 out  
555 the 25 traits had more than 20 SNPs that met our inclusion *P* value threshold. Out of these 6, only 3 traits  
556 overlapped with traits that were also seen in the Biobank of Japan dataset. Largely, the within-sib GWAS  
557 data agreed with our previous analysis, with the exception of selection favoring BMI-decreasing alleles in  
558 the Neolithic. However, the within-sib analysis was considerably underpowered compared with using the  
559 BBJ dataset, as the total number of SNPs were different by an order of magnitude ([Supplementary Table](#)  
560 [13](#)).  
561

Type	Category	Trait ID	Trait	Number of SNPs (Within sib)	Analysis method / Epoch								
					No effect size			Effect size			Within-sib		
					EN	BA	H	EN	BA	H	EN	BA	H
Anthropometric	Biomarker	BMI	Body mass index	306 (30)	0.4	44.9	81.3	2.0	5.3	90.4	11.6	51.8	46.1
		BW	Body weight	517	2.4	47.5	98.4	4.6	17.9	95.4			
Auto-Immune	Medication	L04	Immunosuppressants	65	0.4	79.6	0.0	0.0	92.8	0.3			
		CHB	Chronic Hepatitis B	43	21.5	0.0	0.0	2.2	0.0	0.0			
	Disease	T2D	Type 2 diabetes	244	15.6	2.5	37.9	23.8	0.0	22.1			
		ChS	Chronic sinusitis	68	50.5	50.8	0.1	80.4	31.3	0.3			
		RA	Rheumatoid arthritis	91	18.7	75.0	0.4	2.1	95.6	0.5			
UC	Ulcerative colitis	30	33.4	85.0	99.7	7.2	96.7	99.6					
Blood cell	Biomarker	MCHC	Mean corpuscular hemoglobin conc.	133	1.4	100.0	28.0	1.9	99.0	87.1			
		PLT	Platelet	450	0.5	48.8	0.6	7.1	30.6	4.2			
		MCV	Mean corpuscular volume	426	98.4	75.7	63.4	91.2	92.5	77.5			
		MCH	Mean corpuscular hemoglobin	372	71.8	99.1	54.0	64.3	94.3	92.3			
		BAS	Basophil	99	46.9	98.6	98.2	96.9	99.3	98.1			
EOS	Eosinophil	168	47.0	67.8	98.6	70.6	88.7	92.0					
Cancer	Disease	CeC	Cervical cancer	21	2.3	27.2	99.4	0.3	3.8	99.5			
Cardiovascular	Biomarker	SBP	Systolic blood pressure	111	53.0	94.8	97.6	39.8	98.2	98.1			
		C08	Calcium channel blockers	93	68.1	99.3	100.0	98.5	99.8	100.0			
		B01A	Antithrombotic agents	63	67.7	98.5	100.0	90.0	99.4	100.0			
	Medication	C09	Renin-angiotensin system agents	72	73.1	96.4	100.0	92.4	98.9	99.9			
		C07	Beta blocking agents	84	36.5	84.9	100.0	38.1	92.3	100.0			
		C10AA	HMG CoA reductase inhibitors	161	60.9	35.2	99.3	87.3	35.0	99.9			
	Disease	Ang	Angina pectoris	67	35.4	94.0	99.7	16.5	99.1	100.0			
		SAP	Stable angina pectoris	111	29.6	87.8	100.0	14.2	98.7	99.7			
		UAP	Unstable angina pectoris	31	33.3	9.5	99.1	3.4	1.5	99.4			
		MI	Myocardial infarction	180	75.0	30.9	100.0	89.3	5.1	100.0			
Dermatological	Medication	N02BA	Salicylic acid and derivatives	84	91.6	92.7	100.0	97.2	98.8	100.0			
	Disease	AD	Atopic dermatitis	49	47.5	69.9	97.7	86.1	98.8	99.6			
Kidney	Disease	Uro	Urolithiasis	42	7.9	85.6	99.4	1.0	99.0	99.8			
Liver	Biomarker	AST	Aspartate transaminase	155	71.2	68.4	99.9	91.5	93.0	95.2			
		ALT	Alanine aminotransferase	105	100.0	10.4	79.6	98.7	1.6	94.7			
		TC	Total cholesterol	130	99.4	43.8	97.5	97.7	9.1	97.6			
		LDLC	LDL cholesterol	99 (60)	98.9	11.3	99.5	98.1	3.1	98.2	89.1	4.5	98.8
		UA	Uric acids	236	97.8	44.4	97.8	81.8	10.6	89.5			
		ALP	Alkaline phosphatase	151	28.4	60.6	97.7	37.0	92.9	95.9			
		TP	Total protein	174	40.6	18.5	0.0	26.5	6.8	5.9			
Metabolic	Biomarker	Cir	Cirrhosis	21	99.9	29.0	0.0	100.0	1.4	0.0			
		HbA1c	HbA1c	65 (37)	77.3	2.7	94.0	99.0	0.3	99.6	86.8	0.1	6.8
Metabolic	Medication	Glucose	Glucose	50	81.0	23.6	100.0	99.1	5.2	99.6			
		A10	Drugs used in diabetes	159	42.0	0.1	18.6	97.5	0.0	3.7			

562 **Fig. 4: Signals of polygenic selection.** Traits shown in red in a given epoch are ones for which we find  
563 evidence for post-admixture selection favoring trait-increasing alleles during that epoch. Traits shown in  
564 blue show evidence for trait-decreasing in that epoch. In gray are non-significant results. The within-sib  
565 GWAS results are only available for a small subset of traits that overlap with BBJ and have a greatly  
566 reduced number of SNPs that are genome-wide significant in the GWAS (SNP counts shown in brackets  
567 where available). These rows with unavailability of GWAS results for traits from within-sib GWAS are  
568 left blank.  
569

## 570 Discussion

571  
572 Our analysis highlights the power of ancient DNA time series data to reveal evidence of natural  
573 selection in humans that has later become obscured by subsequent evolution. To evaluate the extent to  
574 which our results replicate previous findings, we compared our candidate targets of selection with two  
575 previous studies. The first, Mathieson et al.<sup>4</sup>, used ancient DNA and a similar approach to ours—detecting  
576 sites with unusual allele frequencies compared to genome-wide admixture proportions—but used modern  
577 samples from the 1000 Genomes Project as a target population<sup>4</sup>. The second, Pybus et al.<sup>91</sup>, used a

578 composite approach integrating multiple classical selection tests on modern European genomes<sup>91</sup>. We  
579 found only one candidate shared between our study and with modern genomes from the ancient DNA  
580 based scan in the EN (the HLA region) and two in the BA period (the HLA region and *SLC45A2*), but  
581 9/12 of Mathieson et al.'s<sup>4</sup> candidates were shared with our Historical epoch candidates ([Fig. 3](#)).  
582 Similarly, Pybus et al.<sup>91</sup>, found none of the candidates we found in the EN epoch, one in the BA epoch,  
583 but 7 that match our candidates from the Historical period. A possible explanation for this is that the  
584 admixture in Europe over the past 10,000 years has obscured signals of selection that occurred before the  
585 immediate past<sup>14</sup>.

586  
587 Our approach looking at time trajectories of alleles over a 10,000-year old period also made it  
588 possible to assess the impact of alleles in the germline over long time scales. As an example of this, we  
589 studied the frequency trajectory of the *CCR5-Δ32* variant, which in homozygous form provides protection  
590 against HIV in European individuals. We studied the frequency of this allele using a proxy SNP  
591 (rs73833033) that is in high LD with it. Across the 3 epochs, we find no evidence for selection of this  
592 allele ( $p=0.55, 0.05, 0.34$  for the EN, BA, and H epochs respectively) in line with the evidence from  
593 modern samples, despite previous reports of selection at this locus<sup>92,93</sup> ([Extended Data Fig. 4](#)).  
594

595 It is important to recognize that the number of candidates we find in each epoch does not reflect  
596 the intensity of natural selection in that time. Rather, many factors feed in to epoch-specific statistical  
597 power. Consider the example of *LCT*: it is possible that 6,000-3,000 years ago, selective pressures on  
598 lactase persistence have been stronger than in the Historical period. Here, selection overcame genetic drift  
599 and drove the very rare allele to a frequency of several percentage points of the population. Yet, the  
600 largest change in allele frequency, from a few percent to the majority allele in northern Europe, only  
601 occurred in the Historical period. These are the changes that we are most powered to detect. Another  
602 important caveat is that the genomic control and null model we rely on may not be equally adequate in all  
603 parts of the genome, particularly in the HLA region where mutation rates, recombination rates, natural  
604 selection, and genetic drift are highly atypical<sup>94</sup>. Nevertheless, our estimation of genome-wide admixture  
605 proportions using *qpAdm* suggests that our expected frequencies broadly capture the allele frequency  
606 shifts associated with admixture.

607  
608 Our results also allow us to interpret our signals of selection in light of archaeological,  
609 evolutionary, and biological evidence. In particular, they allow us to test theories about gene-culture co-  
610 evolution, specifically with regard to hypotheses about how major changes in lifestyle in the last ten  
611 millennia in Europe have or have not resulted in signals of genetic adaptations.

612  
613 One important set of insights relates to the genomic impact of the major transition from hunting  
614 and gathering to farming. A set of alleles that were targets of selection in this period have to do with  
615 decreased body weight/size. Complementarily, the archeological record also shows an overall decrease in  
616 body size during the Neolithic transition<sup>95</sup>. One hypothesis is that a reduction in overall calorie intake, a  
617 trait that would be genetically correlated with reduced body weight, was advantageous in the Neolithic  
618 when famines and resource instability became more frequent<sup>96</sup>. Similarly, selection for more efficient  
619 storage and use of glucose in tissues during periods of famine or pathogen outbreaks might also underlie  
620 several of our selection signals associated with insulin secretion, regulating glucose in the blood stream.  
621



622 Our results also allow us to re-examine the hypothesis for selection on the lactase persistence  
623 allele. A recent study suggested that milk exploitation and consumption started well before the lactase  
624 persistence allele began to be selected, and that the selected allele did not show consistent associations  
625 with improved fitness or health in modern individuals, perhaps suggesting that the ability to digest lactose  
626 into adulthood was only selected for in conditions of food scarcity<sup>97</sup>. While this study was exclusively  
627 restricted to just this allele and phenotype, our genome-wide scan adds additional perspective on the  
628 rationale for selection at this locus by connecting it to selection at other candidates in the same time  
629 period. In the Historical period, along with the *LCT* locus, we detect selection candidates in *SH2B3* and  
630 *DHCR7*—two genes that are directly related to vitamin D binding as well as a candidate in *SLC45A2*, a  
631 major locus of light skin pigmentation in Europeans, a phenotype which also promotes vitamin D  
632 synthesis from sunlight. Taking these results together, our results may suggest an alternative to the caloric  
633 supplementation hypothesis; namely, that selection acted to increase calcium uptake—via improved  
634 vitamin D absorption as well as increased dietary uptake through the consumption of milk—a finding that  
635 has also been discussed in another recent study<sup>13</sup>. Vitamin D is almost entirely absent in a plant-based  
636 diets, and these results might also help explain the differences in both lactase persistence and skin  
637 pigmentation between Northern and Southern Europe, with increased sun exposure in Southern Europe  
638 allowing for sufficient vitamin D synthesis despite a similar dietary transition.

639  
640 A third major set of loci we find as candidates are involved in pathogen response or are expressed  
641 on the cell surface of immune cells. A hypothesis behind selection at these loci could be related to the  
642 potentially increased infectious disease load in the Neolithic brought about from people living in closer  
643 proximity to animals as well as to each other, a set of pressures that would have become dramatically  
644 stronger as population sizes increased exponentially in the Bronze Age and Historical periods. Indeed,  
645 over the past few years, a number of ancient DNA studies have reported pathogen sequences from the  
646 Neolithic period and later<sup>98–102</sup>. These studies revealed past epidemics and found evidence for adaptation  
647 of viruses and bacteria to the human host. Evidence from population history analysis also shows that  
648 Europe in the past 10,000 years has seen large scale migrations into the European subcontinent from  
649 individuals practicing different lifestyles. Signals of selection in these immune loci could be reflect the  
650 arrival of new zoonotic pathogens that arrive with incoming farmers and pastoralists who brought  
651 domesticated animals with them (sheep, cattle, and goats in the case of the first farmers<sup>103</sup>, and horses in  
652 the case of Steppe Pastoralists<sup>104</sup>). Our findings of pervasive upward shifts in the frequencies of alleles  
653 increasing cardiovascular disease and auto-immune disease risk can also be interpreted in this light. All  
654 else being equal, our findings suggest that today, people with hunter-gatherer genomes would have been  
655 at lower risk for cardiovascular and autoimmune disease. The high prevalence of cardiovascular disease  
656 in modern societies could be in part due to past selection for heightened inflammatory response—the  
657 immune system’s primary response to harmful stimuli including pathogens<sup>105</sup>. That is, beginning in the  
658 Neolithic and intensifying in the subsequent periods, humans were subject to a greater infectious load,  
659 and selection for proinflammatory genes and a strong inflammatory function due to the secretion of  
660 adipokines, which underlie cardiometabolic diseases, may have resulted in increased risk for  
661 cardiovascular disease.

662  
663 While we find evidence for two hypotheses concerning gene-culture co-evolution in the last ten  
664 thousand years—with selection for traits related to metabolism as well as immune response—we did not  
665 have power to detect selection for cognitive or neuro-psychiatric disease traits, due to the limited data and

666 relatively small sample size for these traits in the Biobank of Japan data. There is no evidence in the  
667 genetic data for selection on such traits, but future larger studies might provide power to detect such  
668 signals.

669  
670 While our work offers some methodological improvements compared to previous efforts, the  
671 greatest improvement in resolution comes from the quality and quantity of ancient DNA data. More  
672 ancient genomes are becoming available from different geographic regions and time periods. Extending  
673 the type of analysis we report here to these datasets has the potential to enrich our understanding of the  
674 history of natural selection on humans beyond what could be learned through analyses of contemporary  
675 sample alone, where ancient selective events are obscured by admixture and drift, and where their timing  
676 cannot be directly determined.

677

## 678 **Methods**

### 679 **Ancient DNA data curation**

680

681 We obtained ancient DNA sequencing data from the Allen Ancient DNA Resource  
682 (<https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data>, version 51), and subsetted the data to only include samples that were enriched for ~1.2  
683 million nuclear targets with an in-solution hybridization capture reagent.  
684

685

686 To analyze the data, we began with the raw read data for all of these samples and sorted the read  
687 pairs by searching for the expected identification indices and barcodes for each library, allowing up to one  
688 mismatch from the expected sequence in each case. We removed adapters and merged together sequences  
689 requiring a 15 base pair overlap (allowing up to one mismatch), taking the highest quality base in the  
690 merged segment to represent the allele. We mapped the resulting sequences to the hg19 human reference  
691 genome<sup>106</sup> using the same command of BWA<sup>107</sup> (version 0.6.1). We removed duplicate sequences  
692 (mapping to the same position in the genome and having the same barcode pair), and merged libraries  
693 corresponding to the same sample (merging across samples that the genetic data revealed were from the  
694 same individual). For each individual, we restricted to sequences passing filters (not overlapping known  
695 insertion/deletion polymorphisms, and having a minimum mapping quality 10), and trimmed two  
696 nucleotides from the end of each sequence to reduce deamination artifacts. In addition, we also restricted  
697 to sequence data with a minimum base quality of 20.

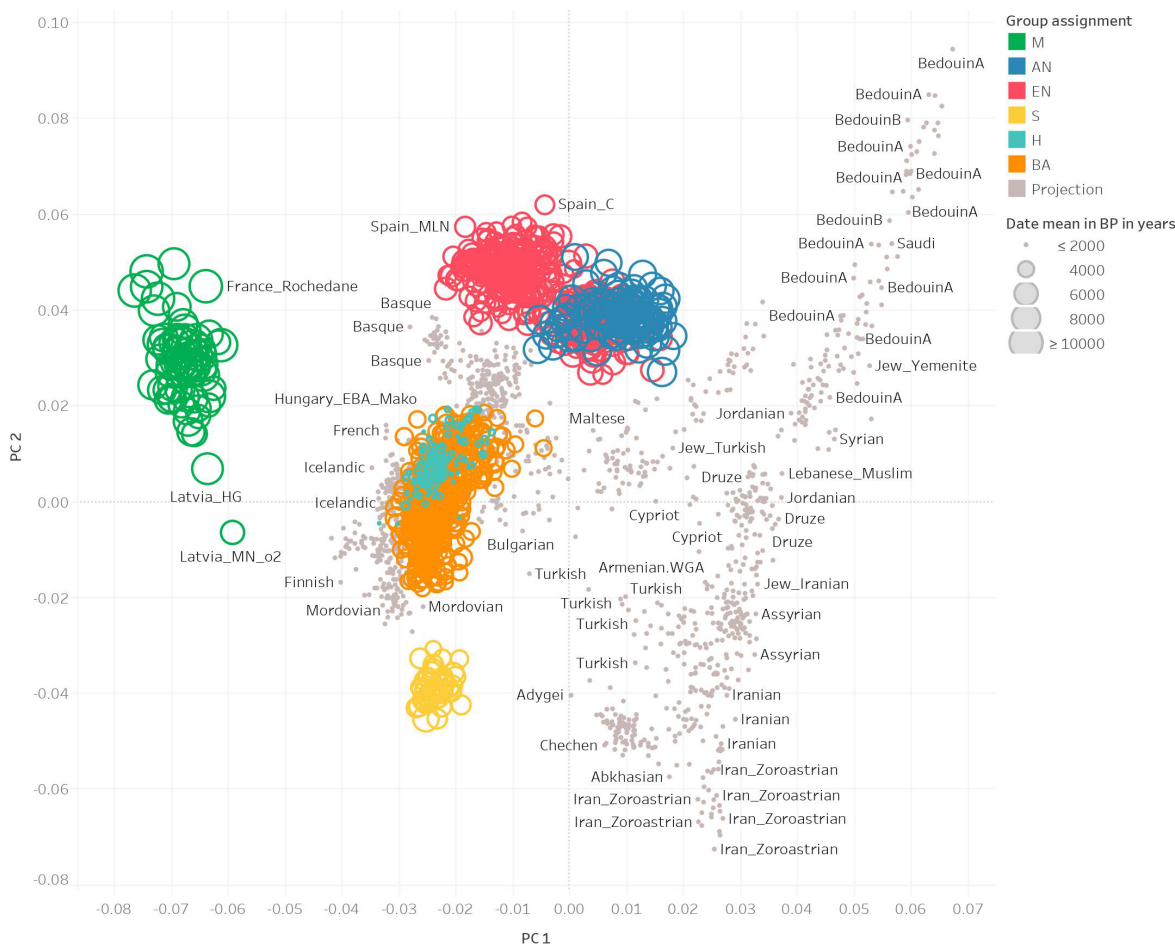
698

699 We assessed evidence for ancient DNA authenticity by measuring the rate of damage in the first  
700 nucleotide, flagging individuals as potentially contaminated if they had less than a 3% cytosine-to-  
701 thymine substitution rate in the first nucleotide for a UDG-treated library and less than a 10% substitution  
702 rate for a non-UDG-treated library. We used contamMix to test for contamination based on  
703 polymorphism in mitochondrial DNA<sup>108</sup> and ANGSD to test for contamination based on polymorphism  
704 on the X chromosome in males<sup>109</sup>, removing individuals with evidence for contamination. For population  
705 genetic analysis to represent each individual at each SNP position, we randomly selected a single  
706 sequence (if at least one was available). For the selection analysis, in order to obtain read count

707 information on a per sample basis, we used BCFtools<sup>110</sup> version 1.3.1 to obtain reference and alternate  
708 counts at each genomic position.

## 709 Principal components analysis

710  
711 We carried out PCA using the smartpca package of EIGENSOFT 7.2.1<sup>111</sup>. We used default  
712 parameters and added two options (lsqproject:YES and numoutlieriter:0) to project the ancient individuals  
713 onto the PCA space. We used 991 present-day West Eurasians<sup>22,112,113</sup> as a basis for projection of the  
714 ancient individuals. We also computed  $F_{ST}$  between groups using the parameters inbreed:YES and  
715 fstonly:YES. We restricted these analyses to the dataset obtained by merging our ancient DNA data with  
716 the modern DNA data on the Human Origins Array and restricted it to 597,573 SNPs. We treated  
717 positions where we did not have sequence data as missing genotypes. [Fig. 1](#) shows the PCA of all ancient  
718 samples. [Extended Data Fig. 1](#) shows underlying modern samples used for the projection, along with the  
719 ancient individuals.



720  
721 **Extended Data Fig. 1.** PCA of ancient samples as well as the basis set of modern samples used in the  
722 projection analysis (in grey).

## 723 **Population grouping and f-statistics**

724 We grouped samples into several broad genetic and cultural categories that represented the major  
725 ancestry groups observed in our European time transect. Our group assignments were:

726 M, Mesolithic - Hunter-Gatherers with no evidence of any admixture from Anatolian Farmers dated to an  
727 average of around 8,600 BP with the majority from Eastern Europe

728 AN, Anatolia Neolithic - Farmer samples dated to an average of around 7,400 BP largely from present  
729 day Turkey, Greece, and the Balkans, with little to no European Hunter-Gatherer admixture

730 EN, Europe Neolithic - Farmer samples dated to an average of around 5,400 BP with the majority of  
731 samples from Central and Western Europe

732 BA, Bronze Age - Samples dated to an average of around 4,000 BP largely from the Bell Beaker cultures  
733 of Czech Republic, Great Britain, Germany, and Slovakia

734 S - Steppe Pastoralists dated to an average of around 4,800 BP with many from Yamnaya and Afanasievo  
735 cultures of the Eastern Steppe

736 H, Historical era - Samples dated to an average of around 2,000 BP with the vast majority from England  
737 and Scotland, as well as a minority of samples from Central Europe

## 738 **Admixture modeling of ancient Europeans**

739 We used *qpAdm* from ADMIXTOOLS to estimate the mixing proportions for the ancestral  
740 populations of each model<sup>113</sup>. *qpAdm* estimates the mixing proportions using the expected values of  $f_4$ -  
741 statistics, where  $f_4(A, B; C, D)$  represents the correlation in allele frequency differences between the  
742 groups  $(A, B)$  and  $(C, D)$ <sup>114</sup>. We used seven outgroups for the computation of the  $f_4$ -statistics:  
743 Ethiopia\_4500BP, Russia\_Ust\_Ishim\_HG\_published.DG, Russia\_MAI\_HG.SG, Israel\_Natufian,  
744 Italy\_North\_Villabruna, Iran\_Ganj\_Dareh\_N, and Russia\_Boisman\_MN.

745  
746 We leveraged previous work that provides a demographic model for major ancestry transitions in  
747 Europe<sup>4,22,23,27</sup>. We modeled European Farmers (EN) as a 84% mixture of early farmers from Anatolia  
748 (AN) and 16% mixture of European Hunter-Gatherers (M). We modeled European Bronze Age samples  
749 as a 48% mixture of European Farmers (EN) and 52% mixture of Steppe Pastoralists (S). Finally, we  
750 modeled Historical era samples from Europe (H) as a 85% mixture of Bronze Age samples and a 15%  
751 mixture of European Neolithic samples, reflecting the additional ancestry changes at that time.

## 752 **Genome-wide scan for natural selection**

753  
754 To estimate the population allele frequencies at each site, we obtained the maximum likelihood  
755 estimate from the likelihood of a given frequency  $p$  using an approach first described in Mathieson et al.  
756 2015. Let  $p$  be a reference allele frequency,  $R_i$  be the number of reads with the reference allele,  $T_i$  be the  
757 total number of sequences,  $N$  be the number of samples for the population, and  $\epsilon$  be a probability of error.  
758 Let the binomial probability mass function be denoted as  $B(x, p, n) = \binom{n}{k} p^x (1 - p)^{n-x}$ . Then the  
759 likelihood of a frequency  $p$  given the read data is  
760

761 
$$L(p; N, R_i, T_i) = \prod_i^N (p^2 B(R_i, T_i, 1 - \epsilon) + 2p(1 - p)B(R_i, T_i, 0.5) + (1 - p)^2 B(R_i, T_i, \epsilon))$$

762

763 Minimizing the negative log-likelihood function produced the allele frequency estimate for each  
764 population at every site. Samples with 0 reference and alternate reads at a site were excluded from the  
765 calculation of the maximum likelihood estimate. We used the SLSQP solver from SciPy<sup>115</sup> to minimize  
766 the negative log-likelihood function, setting the bounds for the allele frequency at 0.01 and 0.99. We also  
767 removed all positions where all reads were missing in any of the populations used in the scan.

768

769 The expected frequency of the target population was also obtained given the mixing proportions  
770 and estimated frequencies of the ancestral populations. For instance, suppose we have the admixture  
771 model  $C = \alpha A + (1 - \alpha)B$ . Then under neutrality, the expected allele frequency for population  $C$  would  
772 be  $p_E = \alpha p_A + (1 - \alpha)p_B$ , where  $p_A$  and  $p_B$  are the observed allele frequencies of populations  $A$  and  $B$ ,  
773 respectively.

774 Let  $p_E$  be the expected frequency of the target population computed as the sum of the products of  
775 the allele frequencies of the ancestral populations and their mixing proportions, and let  $p_O$  be the observed  
776 frequency of the target population. We tested when the observed allele frequency deviated from  
777 expectation using the likelihood ratio test.

778

779 
$$\text{statistic} = -2 \log \left( \frac{L(p_E)}{L(p_O)} \right)$$

780

781 This statistic was used to compute a  $P$  value from the  $\chi_1^2$  distribution. To address genomic  
782 inflation, a control factor was applied to the statistics such that  $\left( \frac{\text{median statistic}}{0.675} \right)^2 = 0.45494$  after  
783 removing 49,000 SNPs of functional importance<sup>4,31</sup>. We also removed genomic positions that were  
784 covered by >15,000 reads (coverage >10x mean coverage) across our dataset, due to potential mis-  
785 mapping artifacts.

786

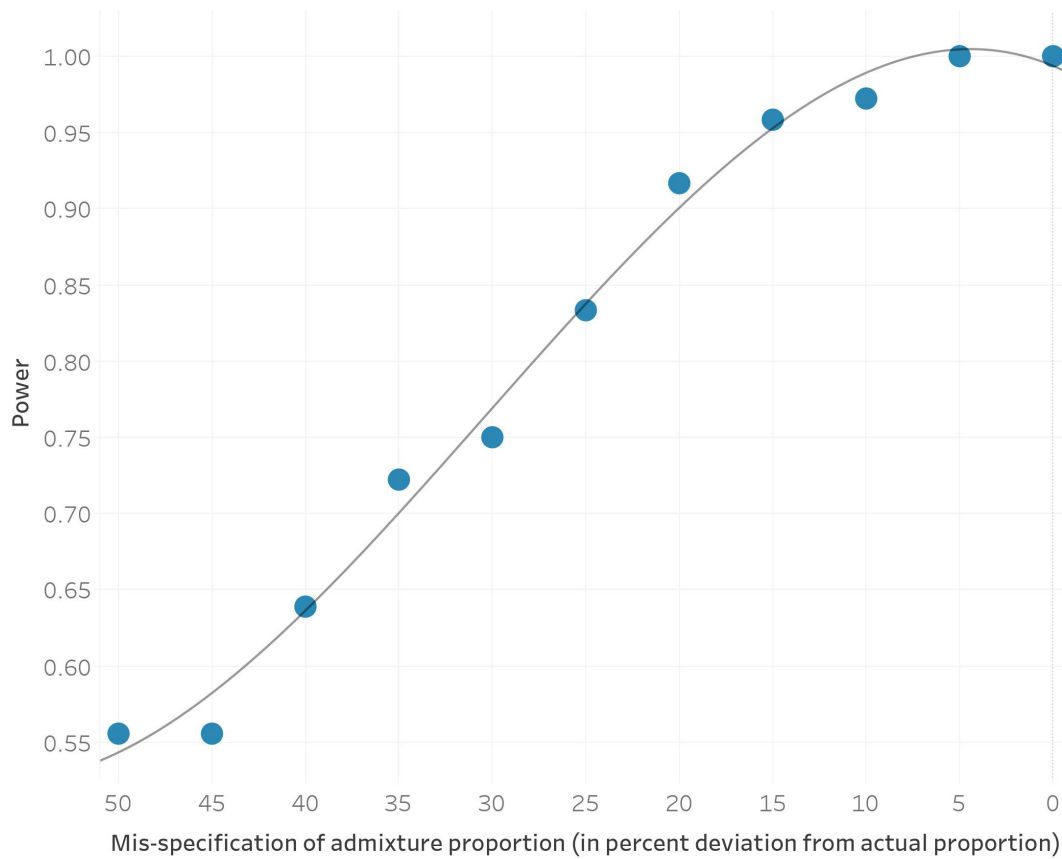
787 Previous work has examined the robustness of this particular approach to mis-estimation of allele  
788 frequencies as well as sample sizes, but we added additional power calculations to our particular scenario.  
789 First, we carried out an analysis where we modified the ancestry proportions in 5% increments from the  
790 actual proportion and again examined the number of 1Mb regions that remained significant according to  
791 our criterion after genomic inflation correction. Our results suggest that we are well-powered to detect the  
792 majority of our signals even with mis-specification of the admixture proportion by over 30% ([Extended  
793 Data Fig. 2](#)).

794

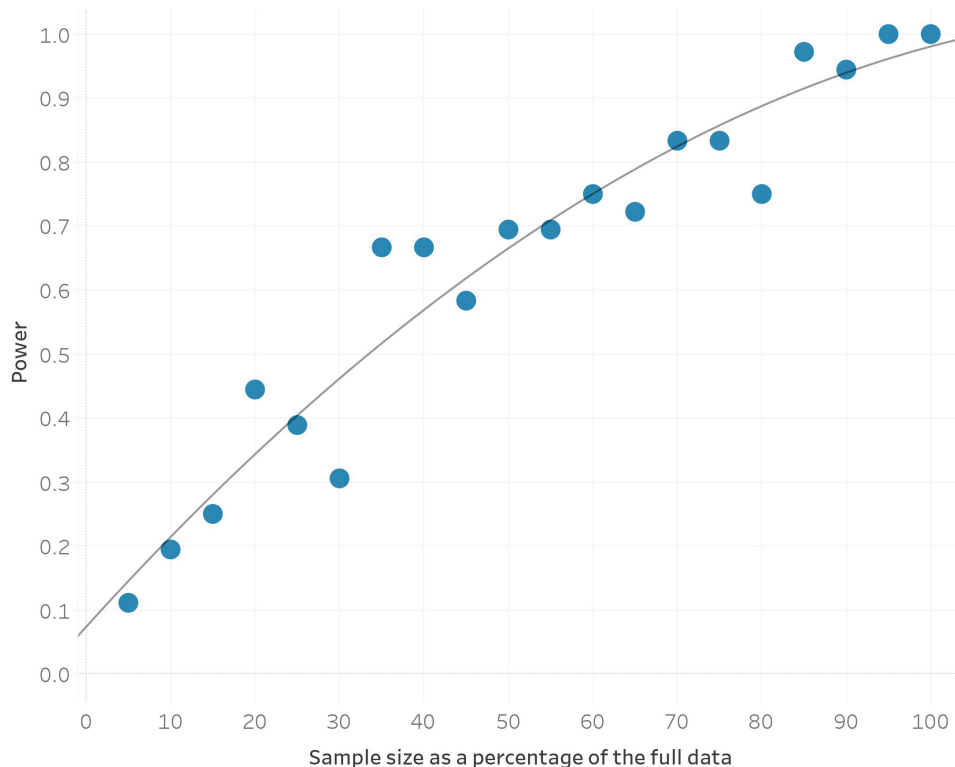
795 Second, we carried out a sub-sampling analysis where we down sampled the overall dataset in 5%  
796 increments (that is, reducing the sample sizes of both the two source populations and the target population  
797 across all 3 epochs in steps of 5%), and then examined the number of 1 Mb regions that remained  
798 significant according to our criterion after genomic inflation correction. We see that with 90% of the data,  
799 we are essentially recapturing most of our signals, though the lack of a clear plateau in our analysis  
800 suggests that increasing sample sizes further is likely to continue to improve the power to discover new  
801 loci ([Extended Data Fig. 3](#)). Small increases in power are seen even at slightly larger sample sizes, as our



802 sampling process is carried out at the level of individuals. This is as expected, as coverage varies greatly  
803 by sample and across genomic position, but the overall upward trajectory of increased power with  
804 increased sample size is clear.  
805

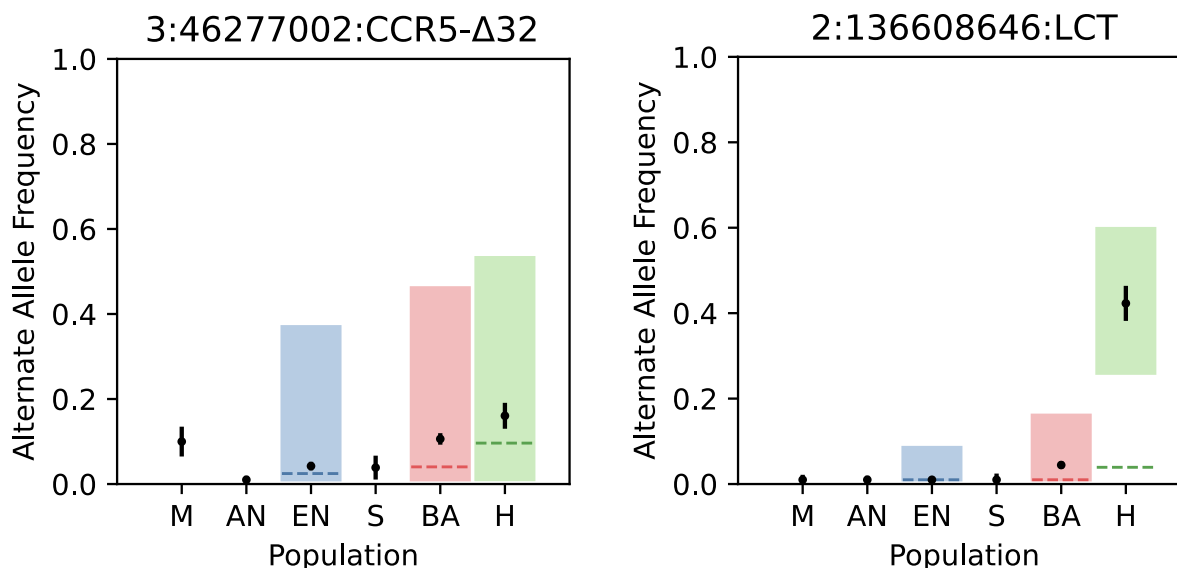


806  
807 **Extended Data Fig. 2.** The power to discover significant genomic regions as a function of admixture  
808 proportion mis-specification shown in percent deviation from the actual mixture proportion. Grey line  
809 shows the quadratic fitted estimate.  
810  
811



812  
813 **Extended Data Fig. 3.** The power to discover significant genomic regions as a function of sample size  
814 (here reported as a percentage of the full dataset of 1,291 samples). Grey line shows the quadratic fitted  
815 estimate.

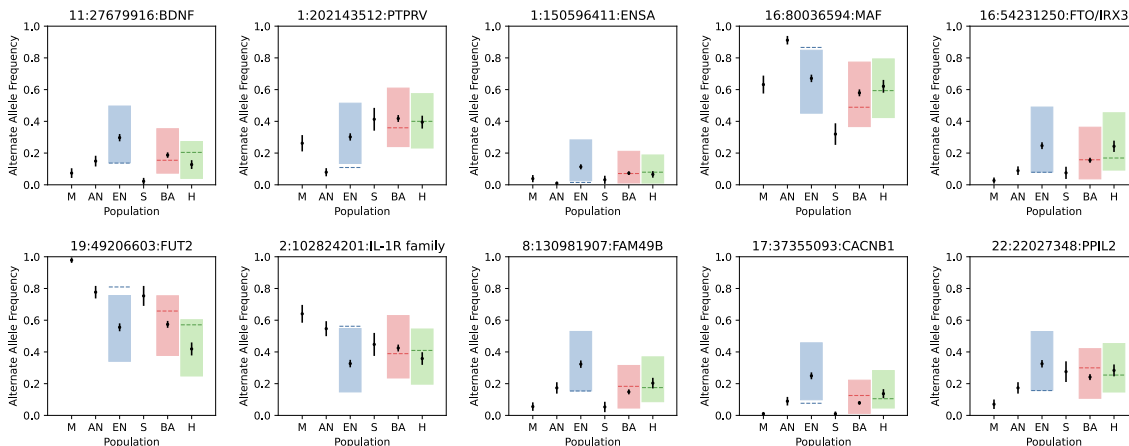
816



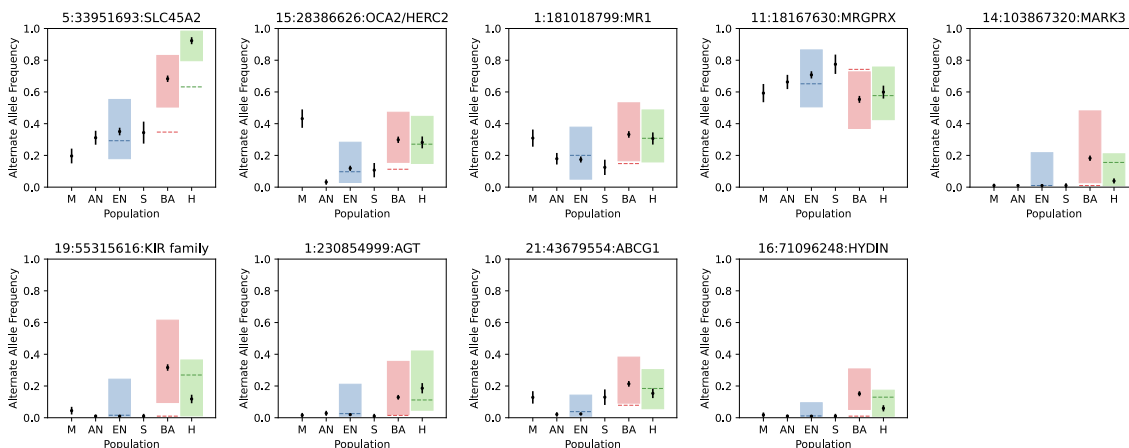
817  
818 **Extended Data Fig. 4.** Point estimates and standard errors of alternate allele frequencies in each  
819 population. The dashed lines (blue for EN, red for BA, and green for the H epoch) are the expected allele  
820 frequencies of the alternate allele based on genome-wide expectations of admixture proportions. An  
821 expected allele frequency that falls outside of the shaded regions would result in a significant  $P$  value

822 from the likelihood ratio test after correction for genomic inflation. The *CCR5-Δ32* allele does not appear  
 823 to be under selection in any of the epochs, but the *LCT* allele shows major changes in frequency in the  
 824 Historical period.

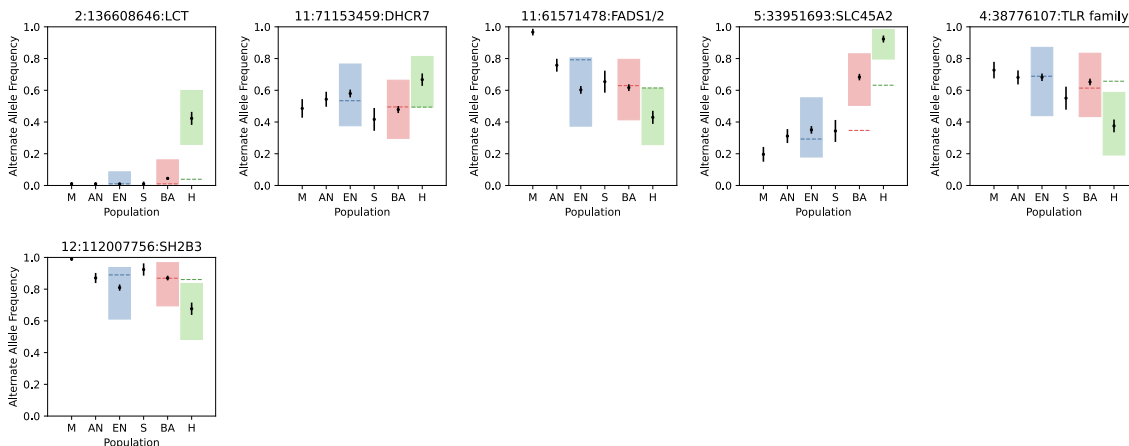
### European Neolithic (EN)



### Bronze Age (BA)



### Historical (H)



826 **Extended Data Fig. 5.** Allele frequency and 95% confidence intervals for selected variants across the 3  
827 epochs. The dashed lines (blue for EN, red for BA, and green for the H epoch) are the expected allele  
828 frequencies of the reference allele based on genome-wide expectations of admixture proportions. An  
829 expected allele frequency that falls outside of the shaded regions would result in a significant  $P$  value  
830 from the likelihood ratio test after correction for genomic inflation.

831

## 832 **Correlation of ancient selective events to those seen in modern Europeans**

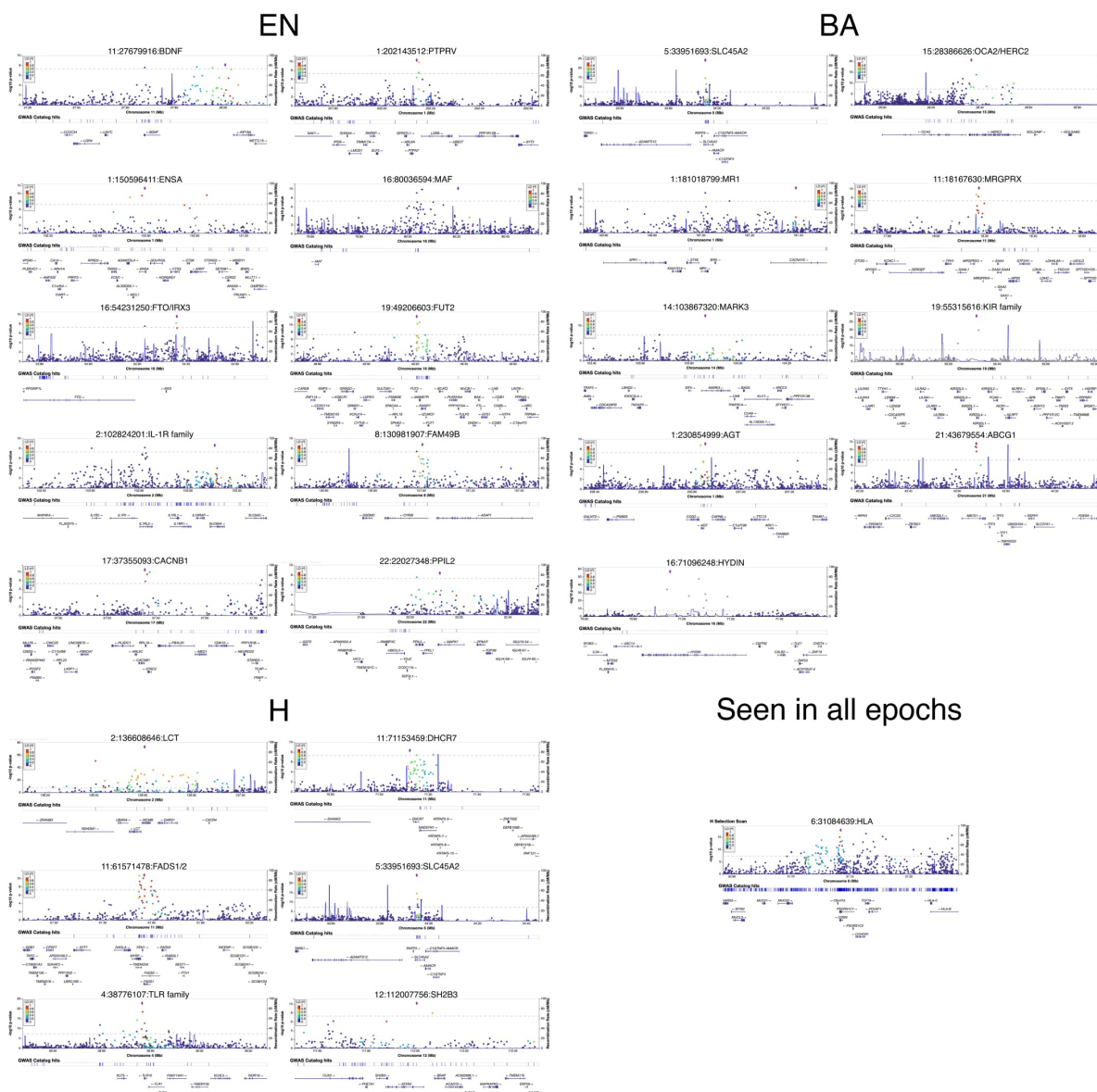
833

834 Selective events that occurred further back in time might be obscured by drift, admixture, or  
835 fluctuations in selective direction over the generations. We wanted to examine if the signals we uncovered  
836 through our ancient selective scan might also have been seen in selection scans of modern Europeans. In  
837 order to do this, we compared the chi-squared statistic we obtained at every locus across all epochs to a  
838 machine learning-based ensemble classifier that integrates several different classical selection tests into a  
839 single predictor<sup>91</sup> in two ways. First, we computed simple overlaps of regions under selection using the  
840 XGBoost algorithm and regions found to be under selection in our analysis. Second, we examined the  
841 number of loci that overlapped with a previous ancient DNA based scan for natural selection<sup>4</sup>.

## 842 **Variant effect predictor**

843 For each population, we filtered the SNPs to a list of variants that had corrected  $P$  values above a  
844 genome-wide significance level of  $5 \times 10^{-8}$  and at least two other SNPs above the significance cutoff  
845 within 1 Mb. We then used the Ensembl Variant Effect Predictor<sup>116</sup> to obtain a list of the nearest genes  
846 associated to each variant and filtered to retain only protein coding genes. All annotations, frequencies  
847 and enrichment analysis were performed on the human reference genome build GRCh37<sup>116</sup>. We include  
848 significant loci in our selection scan and genomics annotations in a 1 Mb neighborhood using  
849 LocusZoom<sup>117</sup> ([Extended Data Fig. 6](#)).

850



851  
852 **Extended Data Fig. 6.** LocusZoom plots of all selected variants and gene annotations in a 1 Mb region  
853 around them.

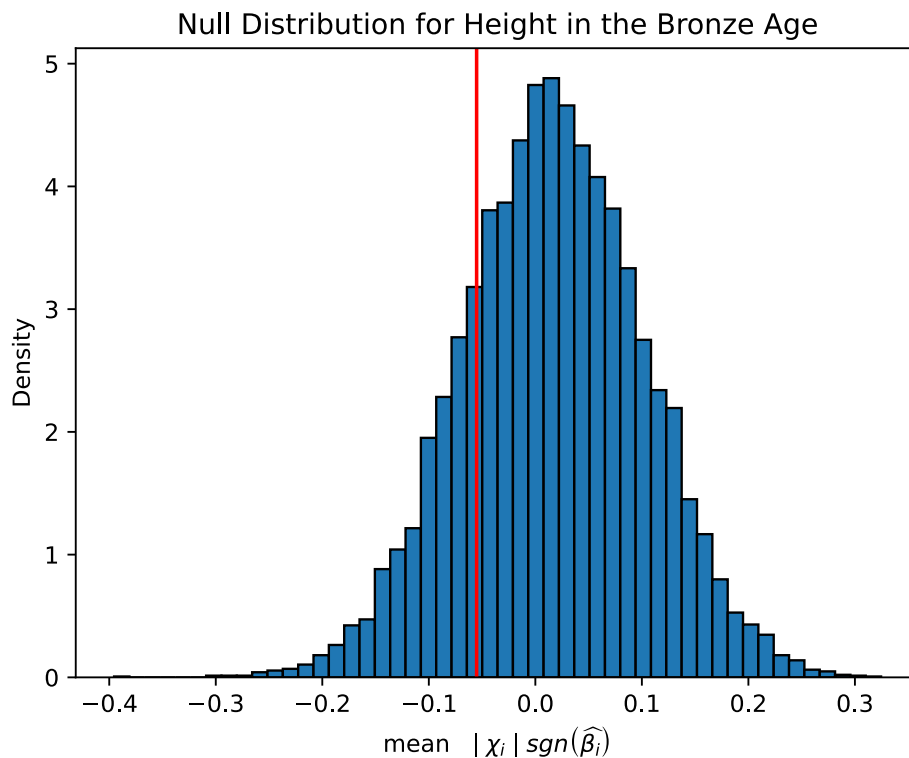
854 **Enrichment analysis**

855 We used the Functional Mapping and Annotation of Genome-Wide Association Studies tool to  
856 obtain significant gene sets for each epoch. The gene sets were produced by comparing the genes of  
857 interest against sets of genes from MsigDB using hypergeometric tests. We performed this analysis for  
858 gene sets from the GWAS and GO functional categories<sup>34</sup>.  
859



## 860 Polygenic selection

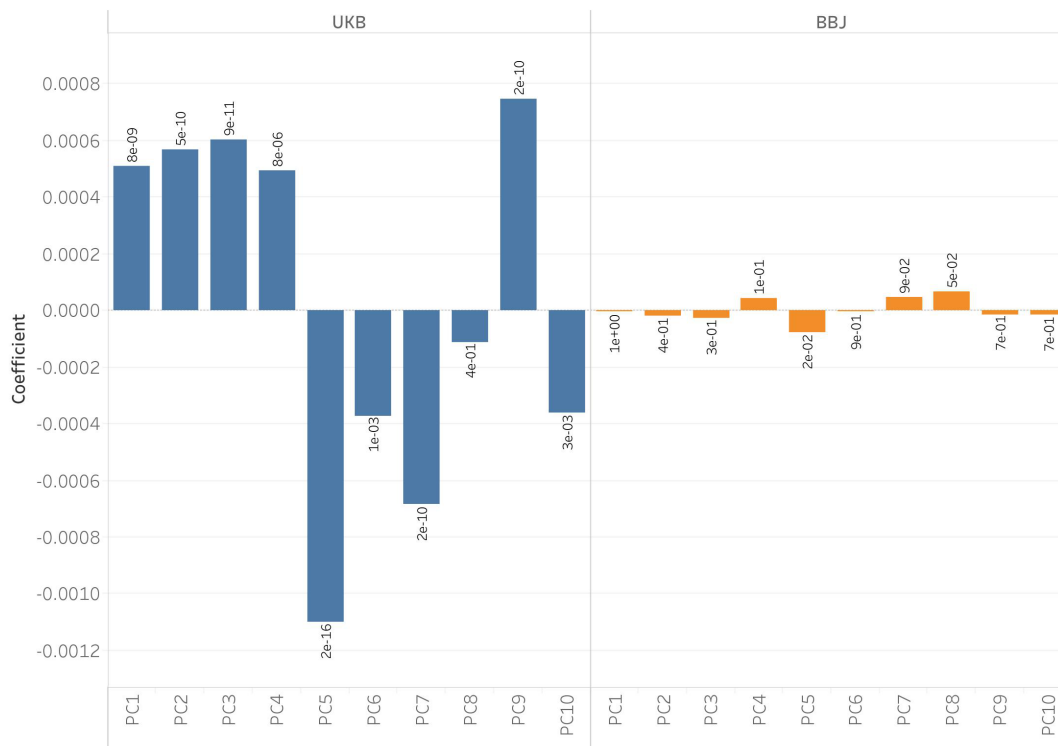
861 We used a modified version of the test from Choin et al.<sup>118</sup> to test for evidence of polygenic  
862 selection. We used GWAS summary statistics from the UK Biobank to test for selection in a control set of  
863 traits, which included skin color, hair color, and triglycerides<sup>119</sup>. We used summary statistics from the  
864 Biobank of Japan to test for selection in 220 different traits<sup>120</sup>. For each trait, we classified each allele as  
865 trait-increasing or trait-decreasing using the effect direction. We then polarized our admixture scan  
866 selection statistic such that a positive sign indicated directional selection of the trait-increasing allele. In  
867 other words, for a given loci  $i$ , our polarized statistic was computed as  $|\chi_i| \text{sgn}(\hat{\beta}_i)$ , where  $|\chi_i|$  is the  
868 magnitude of the chi-squared statistic from the monogenic selection scan, and  $\text{sgn}(\hat{\beta}_i)$  is the sign of the  
869 effect size for the allele that increased or decreased from expectation. For each trait, we compared the  
870 mean polarized statistic of the GWAS significant SNPs (at significance level  $P < 1 \times 10^{-6}$ ) to the distribution  
871 of the mean polarized statistics of randomly sampled SNPs ([Extended Data Fig. 7](#)). The rationale for this  
872 test is that trait-associated SNPs would be more likely than random to undergo short-term directional  
873 selection.  
874



875 **Extended Data Fig. 7.** Null distribution of random subsamples of matched controls (in blue bars) and the  
876 observed statistic (the red line) for a single trait, height, in the Bronze Age. Observed statistics below the  
877 2.5th percentile or above the 97.5th percentile of the null were considered significant.  
878

879 To investigate the impact of population stratification on the GWAS effect sizes, we regressed  
880 GWAS effect sizes on PC loadings on both the UK Biobank (UKB) and Biobank of Japan (BBJ) datasets.  
881 In [Extended Data Fig. 8](#), we show the results of this regression on the best studied and most heritable of  
882

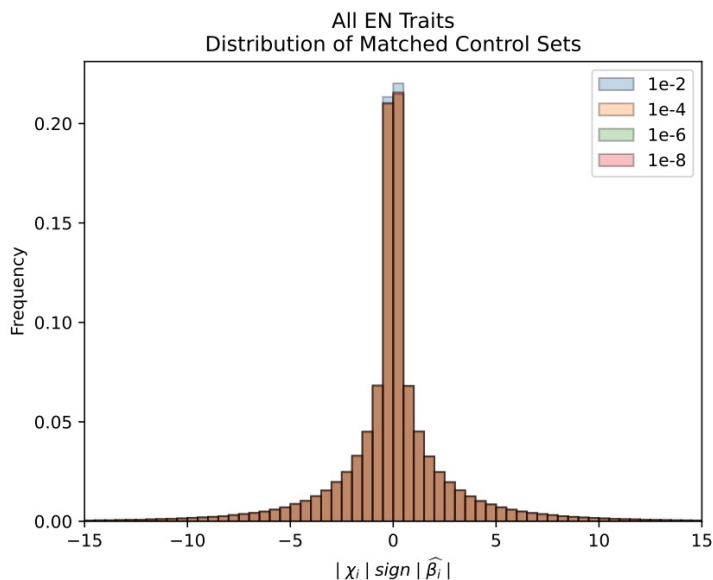
883 the traits we examined: height. Our results show that effect sizes from the UKB are significantly  
884 associated with PC loadings, but effect sizes from BBJ are relatively uncorrelated with PC loadings,  
885 which is in agreement with work from Chen et al<sup>83</sup>. We also observed these results on a set of 38 other  
886 matched quantitative traits and found that only 1 trait has a single PC (PC100) that had PC loadings  
887 significantly associated with effect size using the BBJ dataset, but using the European GWAS 24, PC  
888 loadings across 14 traits were significantly associated with effect size ([Supplementary Table 5](#),  
889 [Supplementary Table 6](#)).  
890  
891



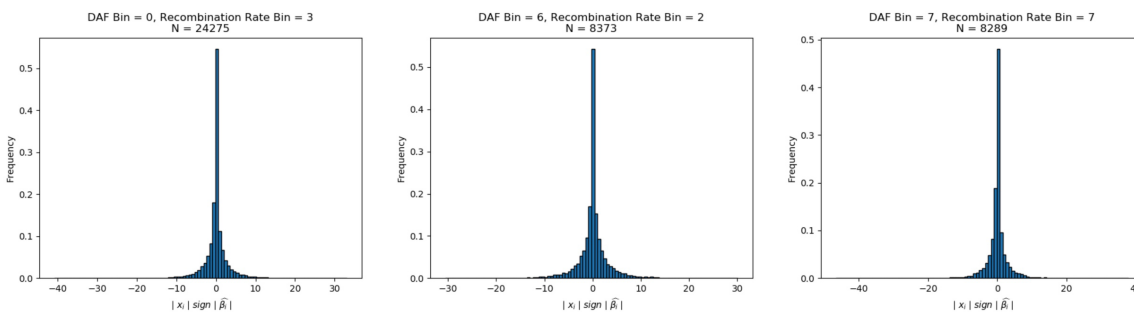
892  
893 **Extended Data Fig. 8.** Regression coefficients of the GWAS effect sizes for height from the UKB and  
894 BBJ on PC loadings from the West Eurasian basis space generated from diverse samples of modern  
895 European genomes. There are no PCs that are significantly associated with effect sizes using BBJ, but this  
896 is not the case for UKB where several PCs are significantly associated with the trait.

897  
898 For each trait, we took 100 kb windows and chose only a single SNP with the lowest  $P$  value to  
899 represent blocks of independent associations with the trait, and we computed the mean polarized statistic  
900 of the set of SNPs that were also below a genome-wide significance threshold of  $1 \times 10^{-6}$ . To match these  
901 observed variants to controls, we binned the other variants in the genome based on derived allele  
902 frequency, B statistic, and recombination rate. The derived allele frequency bins were separated into 8  
903 equally sized bins that ranged from 0 to 1. The B statistic bins were divided by deciles<sup>121</sup>. The  
904 recombination rate bin thresholds were computed such that the recombination rates of the SNPs used in  
905 the admixture scans were evenly distributed across 8 bins. In carrying out this empirical sampling  
906 procedure, we ensured that matched control distributions were normally distributed ([Extended Data Fig.](#)  
907 [9](#)), and that this was the case across different bins ([Extended Data Fig. 10](#)). To ensure that we very rarely

908 resampled the same alleles each time in our random distributions, we ensured that there were a minimum  
909 of 100 variants in each bin ([Supplementary Table 9](#)).  
910



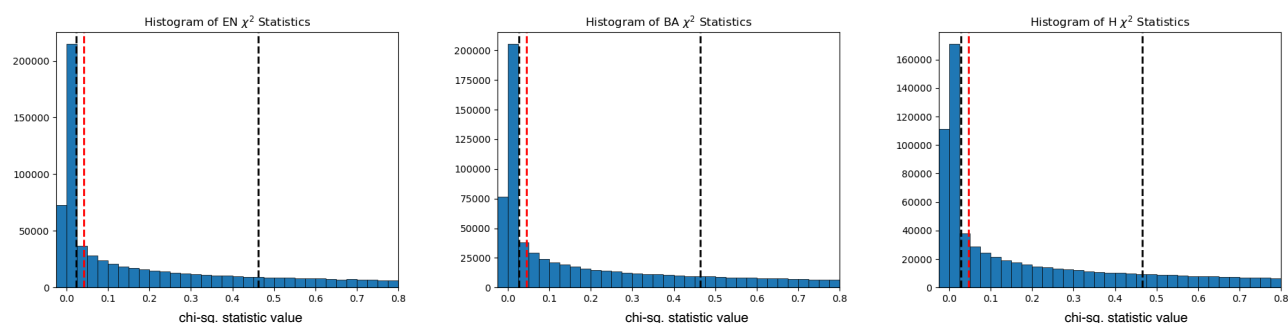
911  
912 **Extended Data Fig. 9.** Null distributions across all the traits we tested, across different  $P$  value  
913 thresholds of the GWAS, showed normality in the null distributions and were centered around 0.  
914



915  
916 **Extended Data Fig. 10.** Distribution of chi-squared statistics across different bins also show null  
917 distributions centered around 0.

918  
919 We then randomly sampled variants that were not associated with the trait 10,000 times to match  
920 the profiles of the variants with the lowest association  $P$  values and computed the mean polarized  
921 selection statistic for the sampled variants. We considered there to be directional selection in the trait-  
922 increasing direction if less than 2.5% of the sampling trials had a mean polarized selection statistic higher  
923 than that of the variants with the lowest  $P$  values. Similarly, we considered there to be directional  
924 selection in the trait-decreasing direction if less than 2.5% of the sampling trials had a mean polarized  
925 selection statistic lower than that of the variants with the lowest  $P$  values. We only report significant  
926 association on traits that had at least 20 SNPs that were significant at a GWAS  $P$  value threshold of less  
927 than  $1 \times 10^{-6}$ ).  
928

929 We also repeated this analysis by multiplying each variant's chi-squared statistic by its effect size,  
930 thereby computing a score that also includes the magnitude of its effect on the trait beyond just looking at  
931 its direction. We report all results for this analysis in [Supplementary Table 10](#). Finally, we carried out an  
932 analysis reducing the SNPs used by removing sites where the chi-squared value was in the bottom 27.5%  
933 of each epoch. We chose to use 27.5% as it was just past the modal value of the distribution at 25%. The  
934 positions that were removed by this process are ones where we have a higher likelihood of mis-estimating  
935 the direction of frequency change. This analysis replicated the majority (90%) of the original signals.  
936 Traits that were not replicated were due to reduction in the overall number of SNPs being reduced to  
937 below 20, a condition we required for the sampling process to be reasonable. Only 4 new traits that were  
938 not significant previously were seen to be significant, but these were sub-significant (5%) in the original  
939 analysis ([Supplementary Table 11](#)).  
940



941  
942 **Extended Data Fig. 11.** Distribution of the frequency (y-axis) of chi-squared statistic values (x-axis)  
943 across epochs. Black lines show the 25th and 50th percentiles and the red line is the 27.5th percentile.

944 In addition to carrying out analysis with the Biobank of Japan dataset, we repeated the analysis  
945 with the UK Biobank dataset [Supplementary Table 12](#). We caution that the results of this analysis are not  
946 readily interpretable or comparable with those from the Biobank of Japan in light of the issues we discuss  
947 with the applications of GWAS with known population stratification artifacts to detect polygenic  
948 selection, but we provide these results for completeness. Finally, we reran the polygenic selection analysis  
949 using data from the within-sibling GWAS consortium. The GWAS estimates from within-sibling studies  
950 are, in theory, better controlled for issues associated with stratification, but the total number of genome-  
951 wide significant loci that met our threshold limited this approach to only a handful of traits. We report the  
952 results of this in [Supplementary Table 13](#).  
953

## 954 Code Availability

955 The code used to run the selection scans for individual alleles and polygenic traits is available at  
956 <https://github.com/Narasimhan-Lab/1000-genomes-natural-selection>.

## 957 References

- 958 1. Richerson, P. J., Boyd, R. & Henrich, J. Gene-culture coevolution in the age of genomics.  
959 *Proc Natl Acad Sci U S A* **107**, 8985–8992 (2010).

- 960 2. Richerson, P. J. & Boyd, R. *Not by Genes Alone : How Culture Transformed Human*  
961 *Evolution*. (University of Chicago Press, 2008).
- 962 3. Armelagos, G. & Cohen, M. *Paleopathology at the Origins of Agriculture*. (1984).
- 963 4. Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature*  
964 *2015 528:7583* **528**, 499–503 (2015).
- 965 5. Margaryan, A. *et al.* Population genomics of the Viking world. *Nature* *2020 585:7825* **585**,  
966 390–396 (2020).
- 967 6. Marnetto, D. *et al.* Ancestral genomic contributions to complex traits in contemporary  
968 Europeans. *Current Biology* **32**, 1412–1419.e3 (2022).
- 969 7. Mathieson, S. & Mathieson, I. FADS1 and the Timing of Human Adaptation to  
970 Agriculture. *Molecular Biology and Evolution* **35**, 2957–2970 (2018).
- 971 8. Segurel, L. *et al.* Why and when was lactase persistence selected for? Insights from Central  
972 Asian herders and ancient DNA. *PLOS Biology* **18**, e3000742 (2020).
- 973 9. Ju, D. & Mathieson, I. The evolution of skin pigmentation-associated variation in West  
974 Eurasia. *Proc Natl Acad Sci U S A* **118**, (2020).
- 975 10. Domínguez-Andrés, J. *et al.* Evolution of cytokine production capacity in ancient and  
976 modern european populations. *Elife* **10**, (2021).
- 977 11. Allentoft, M. E. *et al.* Population Genomics of Stone Age Eurasia. *bioRxiv*  
978 *2022.05.04.490594* (2022) doi:10.1101/2022.05.04.490594.
- 979 12. Kerner, G. *et al.* Genetic adaptation to pathogens and increased risk of inflammatory  
980 disorders in post-Neolithic Europe. *bioRxiv* *2022.07.02.498543* (2022)  
981 doi:10.1101/2022.07.02.498543.
- 982 13. Mathieson, I. & Terhorst, J. Direct detection of natural selection in Bronze Age Britain.  
983 *bioRxiv* *2022.03.14.484330* (2022) doi:10.1101/2022.03.14.484330.
- 984 14. Souilmi, Y. *et al.* Admixture has obscured signals of historical hard sweeps in humans.  
985 *bioRxiv* *2020.04.01.021006* (2021) doi:10.1101/2020.04.01.021006.
- 986 15. Childebayeva, A. *et al.* Population Genetics and Signatures of Selection in Early Neolithic  
987 European Farmers. *Molecular Biology and Evolution* **39**, (2022).
- 988 16. Mühlemann, B. *et al.* Diverse variola virus (smallpox) strains were widespread in northern  
989 Europe in the Viking Age. *Science* **369**, (2020).
- 990 17. Spyrou, M. A. *et al.* Phylogeography of the second plague pandemic revealed through  
991 analysis of historical *Yersinia pestis* genomes. *Nature Communications* *2019 10:1* **10**, 1–13  
992 (2019).
- 993 18. Bos, K. I. *et al.* Eighteenth century *Yersinia pestis* genomes reveal the long-term  
994 persistence of an historical plague focus. *Elife* **5**, (2016).
- 995 19. Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal ancestor.  
996 *Nature* **524**, 216 (2015).
- 997 20. Rohland, N. *et al.* Three Reagents for in-Solution Enrichment of Ancient Human DNA at  
998 More than a Million SNPs. *bioRxiv* *2022.01.13.476259* (2022)  
999 doi:10.1101/2022.01.13.476259.
- 1000 21. Rubinacci, S., Ribeiro, D. M., Hofmeister, R. J. & Delaneau, O. Efficient phasing and  
1001 imputation of low-coverage sequencing data using large reference panels. *Nature Genetics*  
1002 *2021 53:1* **53**, 120–126 (2021).



- 1003 22. Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European  
1004 languages in Europe. *Nature* 2015 522:7555 **522**, 207–211 (2015).
- 1005 23. Lazaridis, I. *et al.* Genomic insights into the origin of farming in the ancient Near East.  
1006 *Nature* 2016 536:7617 **536**, 419–424 (2016).
- 1007 24. Allentoft, M. E. *et al.* Population genomics of Bronze Age Eurasia. *Nature* 2015 522:7555  
1008 **522**, 167–172 (2015).
- 1009 25. Papac, L. *et al.* Dynamic changes in genomic and social structures in third millennium  
1010 BCE central Europe. *Science Advances* **7**, 6941–6966 (2021).
- 1011 26. Lipson, M. *et al.* Parallel palaeogenomic transects reveal complex genetic history of early  
1012 European farmers. *Nature* **551**, 368–372 (2017).
- 1013 27. Narasimhan, V. M. *et al.* The Formation of Human Populations in South and Central Asia.  
1014 *Science* **365**, (2019).
- 1015 28. Olalde, I. *et al.* The Beaker phenomenon and the genomic transformation of northwest  
1016 Europe. *Nature* 2018 555:7695 **555**, 190–196 (2018).
- 1017 29. Olalde, I. *et al.* The genomic history of the Iberian Peninsula over the past 8000 years.  
1018 *Science* (1979) **363**, 1230–1234 (2019).
- 1019 30. Risch, N. & Merikangas, K. The future of genetic studies of complex human diseases.  
1020 *Science* **273**, 1516–1517 (1996).
- 1021 31. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004  
1022 (1999).
- 1023 32. Cuadros-Espinoza, S., Laval, G., Quintana-Murci, L. & Patin, E. The genomic signatures  
1024 of natural selection in admixed human populations. *The American Journal of Human*  
1025 *Genetics* **109**, 710–726 (2022).
- 1026 33. Fauman, E. B. & Hyde, C. An optimal variant to gene distance window derived from an  
1027 empirical definition of cis and trans protein QTLs. *BMC Bioinformatics* **23**, 169 (2022).
- 1028 34. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and  
1029 annotation of genetic associations with FUMA. *Nature Communications* 2017 8:1 **8**, 1–11  
1030 (2017).
- 1031 35. Claussnitzer, M. *et al.* FTO Obesity Variant Circuitry and Adipocyte Browning in Humans  
1032 . *New England Journal of Medicine* **373**, 895–907 (2015).
- 1033 36. Smemo, S. *et al.* Obesity-associated variants within FTO form long-range functional  
1034 connections with IRX3. *Nature* **507**, 371 (2014).
- 1035 37. Aguet, F. *et al.* The GTEx Consortium atlas of genetic regulatory effects across human  
1036 tissues. *Science* (1979) **369**, 1318–1330 (2020).
- 1037 38. Lee, N. K. *et al.* Endocrine Regulation of Energy Metabolism by the Skeleton. *Cell* **130**,  
1038 456–469 (2007).
- 1039 39. Martin Carli, J. F., LeDuc, C. A., Zhang, Y., Stratigopoulos, G. & Leibel, R. L. FTO  
1040 mediates cell-autonomous effects on adipogenesis and adipocyte lipid content by regulating  
1041 gene expression via 6mA DNA modifications. *Journal of Lipid Research* **59**, 1446–1460  
1042 (2018).
- 1043 40. Matsuoka, T. *et al.* Members of the large Maf transcription family regulate insulin gene  
1044 transcription in islet beta cells. *Mol Cell Biol* **23**, 6049–6062 (2003).

- 1045 41. Daily, J. W. & Park, S. Interaction of BDNF rs6265 variants and energy and protein intake  
1046 in the risk for glucose intolerance and type 2 diabetes in middle-aged adults. *Nutrition* **33**,  
1047 187–194 (2017).
- 1048 42. Bathina, S. & Das, U. N. Brain-derived neurotrophic factor and its clinical implications.  
1049 *Archives of Medical Science : AMS* **11**, 1164 (2015).
- 1050 43. Lindesmith, L. C. *et al.* Mechanisms of GII.4 Norovirus Persistence in Human Populations.  
1051 *PLoS Medicine* **5**, 0269–0290 (2008).
- 1052 44. Hazra, A. *et al.* Common variants of FUT2 are associated with plasma vitamin B12 levels.  
1053 *Nat Genet* **40**, 1160 (2008).
- 1054 45. Peters, V. A., Joesting, J. J. & Freund, G. G. IL-1 receptor 2 (IL-1R2) and its role in  
1055 immune regulation. *Brain Behav Immun* **32**, 1 (2013).
- 1056 46. Thali, M. *et al.* Functional association of cyclophilin A with HIV-1 virions. *Nature* **372**,  
1057 363–365 (1994).
- 1058 47. Erdogmus, S. *et al.* Cavβ1 regulates T cell expansion and apoptosis independently of  
1059 voltage-gated Ca<sup>2+</sup> channel function. *Nature Communications* 2022 13:1 **13**, 1–19 (2022).
- 1060 48. Yuki, K. E. *et al.* CYRI/FAM49B negatively regulates RAC1-driven cytoskeletal  
1061 remodelling and protects against bacterial infection. *Nat Microbiol* **4**, 1516–1531 (2019).
- 1062 49. Shang, W. *et al.* Genome-wide CRISPR screen identifies FAM49B as a key regulator of  
1063 actin dynamics and T cell activation. *Proc Natl Acad Sci U S A* **115**, E4051–E4060 (2018).
- 1064 50. Key, F. M. *et al.* Emergence of human-adapted *Salmonella enterica* is linked to the  
1065 Neolithization process. *Nature Ecology & Evolution* 2020 4:3 **4**, 324–333 (2020).
- 1066 51. Ju, D. & Mathieson, I. The evolution of skin pigmentation-associated variation in West  
1067 Eurasia. *Proc Natl Acad Sci U S A* **118**, (2020).
- 1068 52. Field, Y. *et al.* Detection of human adaptation during the past 2000 years. *Science* **354**, 760  
1069 (2016).
- 1070 53. Harriff, M. J. *et al.* MR1 displays the microbial metabolome driving selective MR1-  
1071 restricted T cell receptor usage. *Science Immunology* **3**, 2556 (2018).
- 1072 54. Augusto, D. G., Norman, P. J., Dandekar, R. & Hollenbach, J. A. Fluctuating and  
1073 geographically specific selection characterize rapid evolution of the human Kir region.  
1074 *Frontiers in Immunology* **10**, 989 (2019).
- 1075 55. Cao, C. *et al.* Structure, function and pharmacology of human itch GPCRs. *Nature* 2021  
1076 600:7887 **600**, 170–175 (2021).
- 1077 56. McNeil, B. D. *et al.* Identification of a mast-cell-specific receptor crucial for pseudo-  
1078 allergic drug reactions. *Nature* **519**, 237–241 (2015).
- 1079 57. Wierbowski, S. D. *et al.* A 3D structural SARS-CoV-2–human interactome to explore  
1080 genetic and drug perturbations. *Nature Methods* 2021 18:12 **18**, 1477–1488 (2021).
- 1081 58. Pietzner, M. *et al.* Genetic architecture of host proteins involved in SARS-CoV-2 infection.  
1082 *Nature Communications* 2020 11:1 **11**, 1–14 (2020).
- 1083 59. Merad, M. & Martin, J. C. Author Correction: Pathological inflammation in patients with  
1084 COVID-19: a key role for monocytes and macrophages (Nature Reviews Immunology,  
1085 (2020), 20, 6, (355-362), 10.1038/s41577-020-0331-4). *Nature Reviews Immunology* **20**,  
1086 448 (2020).
- 1087 60. Rascovan, N. *et al.* Emergence and Spread of Basal Lineages of *Yersinia pestis* during the  
1088 Neolithic Decline. *Cell* **176**, 295-305.e10 (2019).

- 1089 61. Andrades Valtueña, A. *et al.* Stone Age *Yersinia pestis* genomes shed light on the early  
1090 evolution, diversity, and ecology of plague. *Proceedings of the National Academy of*  
1091 *Sciences* **119**, (2022).
- 1092 62. Lu, H., Cassis, L. A., Kooi, C. W. V. & Daugherty, A. Structure and functions of  
1093 angiotensinogen. *Hypertension Research* 2016 39:7 **39**, 492–500 (2016).
- 1094 63. Kennedy, M. A. *et al.* ABCG1 has a critical role in mediating cholesterol efflux to HDL  
1095 and preventing cellular lipid accumulation. *Cell Metab* **1**, 121–131 (2005).
- 1096 64. Olbrich, H. *et al.* Recessive HYDIN Mutations Cause Primary Ciliary Dyskinesia without  
1097 Randomization of Left-Right Body Asymmetry. *The American Journal of Human Genetics*  
1098 **91**, 672–684 (2012).
- 1099 65. Patterson, N. *et al.* Large-scale migration into Britain during the Middle to Late Bronze  
1100 Age. *Nature* 2021 601:7894 **601**, 588–594 (2021).
- 1101 66. Jablonski, N. G. & Chaplin, G. Human skin pigmentation as an adaptation to UV radiation.  
1102 *Proc Natl Acad Sci U S A* **107**, 8962–8968 (2010).
- 1103 67. Albiñana, C. *et al.* Genetic correlates of vitamin D-binding protein and 25 hydroxyvitamin  
1104 D in neonatal dried blood spots. *medRxiv* 2022.06.08.22276164 (2022)  
1105 doi:10.1101/2022.06.08.22276164.
- 1106 68. Zhernakova, A. *et al.* Evolutionary and functional analysis of celiac risk loci reveals  
1107 SH2B3 as a protective factor against bacterial infection. *Am J Hum Genet* **86**, 970–977  
1108 (2010).
- 1109 69. Hernandez, R. D. *et al.* Classic selective sweeps were rare in recent human evolution.  
1110 *Science (1979)* **331**, 920–924 (2011).
- 1111 70. Murphy, D., Elyashiv, E., Amster, G. & Sella, G. Broad-scale variation in human genetic  
1112 diversity levels is predicted by purifying selection on coding and non-coding elements.  
1113 *bioRxiv* 2021.07.02.450762 (2021) doi:10.1101/2021.07.02.450762.
- 1114 71. Sella, G. & Barton, N. H. Thinking About the Evolution of Complex Traits in the Era of  
1115 Genome-Wide Association Studies. [https://doi.org/10.1146/annurev-genom-083115-](https://doi.org/10.1146/annurev-genom-083115-022316)  
1116 *022316* **20**, 461–493 (2019).
- 1117 72. Berg, J. J. & Coop, G. A Population Genetic Signal of Polygenic Adaptation. *PLOS*  
1118 *Genetics* **10**, e1004412 (2014).
- 1119 73. Höllinger, I., Pennings, P. S. & Hermisson, J. Polygenic adaptation: From sweeps to subtle  
1120 frequency shifts. *PLOS Genetics* **15**, e1008035 (2019).
- 1121 74. Speidel, L., Forest, M., Shi, S. & Myers, S. R. A method for genome-wide genealogy  
1122 estimation for thousands of samples. *Nature Genetics* 2019 51:9 **51**, 1321–1329 (2019).
- 1123 75. Berg, J. J. *et al.* Reduced signal for polygenic adaptation of height in UK biobank. *Elife* **8**,  
1124 (2019).
- 1125 76. Sohail, M. *et al.* Polygenic adaptation on height is overestimated due to uncorrected  
1126 stratification in genome-wide association studies. *Elife* **8**, (2019).
- 1127 77. Rosenberg, N. A., Edge, M. D., Pritchard, J. K. & Feldman, M. W. Interpreting polygenic  
1128 scores, polygenic adaptation, and human phenotypic differences. *Evolution, Medicine, and*  
1129 *Public Health* **2019**, 26–34 (2019).
- 1130 78. Harpak, A. & Przeworski, M. The evolution of group differences in changing  
1131 environments. *PLOS Biology* **19**, e3001072 (2021).

- 1132 79. Mills, M. C. & Mathieson, I. The challenge of detecting recent natural selection in human  
1133 populations. *Proceedings of the National Academy of Sciences* **119**, (2022).
- 1134 80. Young, A. I., Benonisdottir, S., Przeworski, M. & Kong, A. Deconstructing the sources of  
1135 genotype-phenotype associations in humans. *Science (1979)* **365**, 1396–1400 (2019).
- 1136 81. Mostafavi, H. *et al.* Variable prediction accuracy of polygenic scores within an ancestry  
1137 group. *Elife* **9**, (2020).
- 1138 82. Barton, N., Hermisson, J. & Nordborg, M. Population Genetics: Why structure matters.  
1139 *Elife* **8**, (2019).
- 1140 83. Chen, M. *et al.* Evidence of Polygenic Adaptation in Sardinia at Height-Associated Loci  
1141 Ascertained from the Biobank Japan. *American Journal of Human Genetics* **107**, 60–71  
1142 (2020).
- 1143 84. Zhou, W. *et al.* Global Biobank Meta-analysis Initiative: powering genetic discovery across  
1144 human diseases. *Unnur Thorsteinsdottir* **27**,.
- 1145 85. Kanai, M. *et al.* Insights from complex trait fine-mapping across diverse populations.  
1146 *medRxiv* 2021.09.03.21262975 (2021) doi:10.1101/2021.09.03.21262975.
- 1147 86. des Marais, D. L., Hernandez, K. M. & Juenger, T. E. Genotype-by-environment  
1148 interaction and plasticity: Exploring genomic responses of plants to the abiotic  
1149 environment. *Annual Review of Ecology, Evolution, and Systematics* **44**, 5–29 (2013).
- 1150 87. Halldorsson, B. v. *et al.* Characterizing mutagenic effects of recombination through a  
1151 sequence-level genetic map. *Science* **363**, (2019).
- 1152 88. McVicker, G., Gordon, D., Davis, C. & Green, P. Widespread Genomic Signatures of  
1153 Natural Selection in Hominid Evolution. *PLOS Genetics* **5**, e1000471 (2009).
- 1154 89. Cox, S. L., Ruff, C. B., Maier, R. M. & Mathieson, I. Genetic contributions to variation in  
1155 human stature in prehistoric Europe. *Proc Natl Acad Sci U S A* **116**, 21484–21492 (2019).
- 1156 90. Howe, L. J. *et al.* Within-sibship genome-wide association analyses decrease bias in  
1157 estimates of direct genetic effects. *Nature Genetics* 2022 54:5 **54**, 581–592 (2022).
- 1158 91. Pybus, M. *et al.* Hierarchical boosting: a machine-learning framework to detect and  
1159 classify hard selective sweeps in human populations. *Bioinformatics* **31**, 3946–3952  
1160 (2015).
- 1161 92. Maier, R. *et al.* No statistical evidence for an effect of CCR5-Δ32 on lifespan in the UK  
1162 Biobank cohort. *Nature Medicine* 2019 26:2 **26**, 178–180 (2019).
- 1163 93. Wei, X. & Nielsen, R. Retraction Note: CCR5-Δ32 is deleterious in the homozygous state  
1164 in humans. *Nature Medicine* 2019 25:11 **25**, 1796–1796 (2019).
- 1165 94. Fortier, A. L. & Pritchard, J. K. Ancient Trans-Species Polymorphism at the Major  
1166 Histocompatibility Complex in Primates. *bioRxiv* 2022.06.28.497781 (2022)  
1167 doi:10.1101/2022.06.28.497781.
- 1168 95. Marciniak, S. *et al.* An integrative skeletal and paleogenomic analysis of prehistoric stature  
1169 variation suggests relatively reduced health for early European farmers. *bioRxiv* **12**,  
1170 2021.03.31.437881 (2021).
- 1171 96. Shennan, S. *et al.* Regional population collapse followed initial agriculture booms in mid-  
1172 Holocene Europe. *Nature Communications* 2013 4:1 **4**, 1–8 (2013).
- 1173 97. Evershed, R. P. *et al.* Dairying, diseases and the evolution of lactase persistence in Europe.  
1174 *Nature* 2022 1–10 (2022) doi:10.1038/s41586-022-05010-7.

- 1175 98. de Barros Damgaard, P. *et al.* 137 ancient human genomes from across the Eurasian  
1176 steppes. *Nature* **557**, 369–374 (2018).
- 1177 99. Mühlemann, B. *et al.* Ancient hepatitis B viruses from the Bronze Age to the Medieval  
1178 period. *Nature* **2018 557:7705** **557**, 418–423 (2018).
- 1179 100. Guellil, M. *et al.* Ancient herpes simplex 1 genomes reveal recent viral structure in Eurasia.  
1180 *Science Advances* **8**, 4435 (2022).
- 1181 101. Key, F. M. *et al.* Emergence of human-adapted *Salmonella enterica* is linked to the  
1182 Neolithization process. *Nature Ecology & Evolution* **2020 4:3** **4**, 324–333 (2020).
- 1183 102. Kocher, A. *et al.* Ten millennia of hepatitis B virus evolution. *Science (1979)* **374**, (2021).
- 1184 103. Verdugo, M. P. *et al.* Ancient cattle genomics, origins, and rapid turnover in the Fertile  
1185 Crescent. *Science (1979)* **365**, 173–176 (2019).
- 1186 104. Librado, P. *et al.* The origins and spread of domestic horses from the Western Eurasian  
1187 steppes. *Nature* **2021 598:7882** **598**, 634–640 (2021).
- 1188 105. Chen, L. *et al.* Inflammatory responses and inflammation-associated diseases in organs.  
1189 *Oncotarget* **9**, 7204 (2018).
- 1190 106. Church, D. M. *et al.* Modernizing reference genome assemblies. *PLoS Biol* **9**, (2011).
- 1191 107. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler  
1192 transform. *Bioinformatics* **25**, 1754–1760 (2009).
- 1193 108. Fu, Q. *et al.* A Revised Timescale for Human Evolution Based on Ancient Mitochondrial  
1194 Genomes. *Current Biology* **23**, 553–559 (2013).
- 1195 109. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation  
1196 Sequencing Data. *BMC Bioinformatics* **15**, 1–13 (2014).
- 1197 110. Narasimhan, V. *et al.* BCFtools/RoH: a hidden Markov model approach for detecting  
1198 autozygosity from next-generation sequencing data. *Bioinformatics* **32**, 1749 (2016).
- 1199 111. Patterson, N., Price, A. L. & Reich, D. Population Structure and Eigenanalysis. *PLOS*  
1200 *Genetics* **2**, e190 (2006).
- 1201 112. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-  
1202 day Europeans. *Nature* **2014 513:7518** **513**, 409–413 (2014).
- 1203 113. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
- 1204 114. Reich, D., Thangaraj, K., Patterson, N., Price, A. L. & Singh, L. Reconstructing Indian  
1205 population history. *Nature* **2009 461:7263** **461**, 489–494 (2009).
- 1206 115. scipy.optimize.minimize — SciPy v1.8.1 Manual.  
1207 [https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.minimize.html#scipy.o](https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.minimize.html#scipy.optimize.minimize)  
1208 [ptimize.minimize](https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.minimize.html#scipy.optimize.minimize).
- 1209 116. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biology* **17**, 1–14  
1210 (2016).
- 1211 117. Boughton, A. P. *et al.* LocusZoom.js: Interactive and embeddable visualization of genetic  
1212 association study results. *Bioinformatics* **37**, 3017–3018 (2021).
- 1213 118. Choin, J. *et al.* Genomic insights into population history and biological adaptation in  
1214 Oceania. *Nature* **2021 592:7855** **592**, 583–589 (2021).
- 1215 119. UK Biobank — Neale lab. <http://www.nealelab.is/uk-biobank/>.
- 1216 120. Sakaue, S. *et al.* A global atlas of genetic associations of 220 deep phenotypes. *medRxiv*  
1217 **46**, 2020.10.23.20213652 (2021).



- 1218 121. McVicker, G., Gordon, D., Davis, C. & Green, P. Widespread Genomic Signatures of  
1219 Natural Selection in Hominid Evolution. *PLOS Genetics* **5**, e1000471 (2009).  
1220  
1221

## 1222 **Acknowledgements**

1223 We would like to thank Robert Maier, Alex Okomoto, Iain Mathieson, and Dan Lieberman for useful  
1224 discussions. Megan Le was supported by an Undergraduate Research Fellowship Award, an Advanced  
1225 Summer Research Fellowship Award, and Endowed Presidential Scholarship Awards. Olivia Smith was  
1226 supported by a National Science Foundation Graduate Research Fellowship. This work was supported by  
1227 the National Institutes of Health (GM100233 and HG012287), the John Templeton Foundation (grant  
1228 61220), and by the Allen Discovery Center program, a Paul G. Allen Frontiers Group advised program of  
1229 the Paul G. Allen Family Foundation and the Howard Hughes Medical Institute.  
1230

## 1231 **Author Information**

### 1232 **Authors and Affiliations**

1233 **Oden Institute for Computational Engineering and Sciences, The University of Texas at Austin,**  
1234 **Austin, 78712, Texas, USA**

1235 Megan K. Le

1236

1237 **Department of Computer Science, The University of Texas at Austin, Austin, 78712, Texas, USA**

1238 Megan K. Le

1239

1240 **Department of Integrative Biology, The University of Texas at Austin, Austin, 78712, Texas, USA**

1241 Olivia S. Smith, Arbel Harpak, & Vagheesh M. Narasimhan

1242

1243 **Department of Genetics, Harvard Medical School, Boston, 02115, Massachusetts, USA**

1244 Ali Akbari & David Reich

1245

1246 **Department of Human Evolutionary Biology, Harvard University, Cambridge, 02138,**

1247 **Massachusetts, USA**

1248 Ali Akbari & David Reich

1249

1250 **Department of Population Health, Dell Medical School, Austin, 78712, Texas, USA**

1251 Arbel Harpak

1252

1253 **Howard Hughes Medical Institute, Harvard Medical School, Boston, 02115, Massachusetts, USA**

1254 David Reich

1255

1256 **Broad Institute of MIT and Harvard, Cambridge, 02142, Massachusetts, USA**

1257 David Reich

1258

1259 **Department of Statistics and Data Science, The University of Texas at Austin, Austin, 78712, Texas,**

1260 **USA**

1261 Vagheesh M. Narasimhan

## 1262 **Contributions**

1263 A.H., D.R. and V.M.N. supervised the study. M.K.L, O.S. and A.A. analyzed genetic data. M.K.L., A.H.,

1264 D.R. and V.M.N. wrote the manuscript with input from all co-authors.

1265

## 1266 **Corresponding authors**

1267 Correspondence to Vagheesh M. Narasimhan ([vagheesh@utexas.edu](mailto:vagheesh@utexas.edu)), David Reich

1268 ([reich@genetics.med.harvard.edu](mailto:reich@genetics.med.harvard.edu)) and Arbel Harpak ([arbelharpak@utexas.edu](mailto:arbelharpak@utexas.edu)).

1269

## 1270 **Ethics declarations**

### 1271 **Competing interests**

1272 The authors declare no competing financial interests.

1273

## 1274 **Supplementary Information**

### 1275 **Supplementary Table 1**

1276 This file contains metadata information for all samples used in the analysis.

1277

### 1278 **Supplementary Table 2**

1279 This file contains the significant gene sets from enrichment analysis for the Neolithic period.

1280

### 1281 **Supplementary Table 3**

1282 This file contains the significant gene sets from enrichment analysis for the Bronze Age period.

1283

### 1284 **Supplementary Table 4**

1285 This file contains the significant gene sets from enrichment analysis for the Historical period.

1286

### 1287 **Supplementary Table 5**

1288 This file contains the results of regressing the effect sizes of traits from the Biobank of Japan on PC  
1289 loadings.

1290

1291 **Supplementary Table 6**

1292 This file contains the results of regressing the effect sizes of traits from the UK Biobank on PC loadings.

1293

1294 **Supplementary Table 7**

1295 This file contains the results of the analysis on the correlations and shared effect directions for effect sizes  
1296 of matched traits in the Biobank of Japan and the UK Biobank.

1297

1298 **Supplementary Table 8**

1299 This file contains the results for the polygenic selection scan using the directions of effect sizes for traits  
1300 from the Biobank of Japan.

1301

1302 **Supplementary Table 9**

1303 This file contains the resampling bin distributions of null variants for the polygenic selection scan on a  
1304 single trait.

1305

1306 **Supplementary Table 10**

1307 This file contains the results for the polygenic selection scan using both the magnitudes and directions of  
1308 effect sizes for traits from the Biobank of Japan.

1309

1310 **Supplementary Table 11**

1311 This file contains the results for the polygenic selection scan using the directions of effect sizes for traits  
1312 from the Biobank of Japan and removing SNPs that were in the lowest 27.5% of the chi-squared statistic  
1313 distribution.

1314

1315 **Supplementary Table 12**

1316 This file contains the results for the polygenic selection scan using the directions of effect sizes for traits  
1317 from the UK Biobank.

1318

1319 **Supplementary Table 13**

1320 This file contains the results for the polygenic selection scan using the directions of effect sizes for traits  
1321 from the within-sibling GWAS consortium.

1322

1323 **Supplementary Data 1**

1324 This file contains genome-wide selection scan results and allele frequencies for the Neolithic period.

1325

1326 **Supplementary Data 2**

1327 This file contains genome-wide selection scan results and allele frequencies for the Bronze Age period.

1328

1329 **Supplementary Data 3**

1330 This file contains genome-wide selection scan results and allele frequencies for the Historical period.