Genome–wide patterns of selection in 230 ancient Eurasians

Iain Mathieson1, Josif Lazaridis1,2, Nadin Rohland1,2, Swapan Mallick1,2,3, Nick Patterson2, Songül Alpaslan Roodenberg4, Eadaoïn Harney1,3, Kristin Stewardson1,3, Daniel Fernandes5, Mario Novak5,6, Kendra Sirak5,7, Cristina Gamba5,8, Eppie R. Jones8, Bastien Llamas9, Stanislav Dryomov10,11, Joseph Pickrell1, Juan Luis Arsuaga12,13, José María Bermúdez de Castro1, Eudald Carbonell14,15, Fokke Gerritsen17, Aleksandr Khokhlov18, Pavel Kuznetsov18, Marina Lozano15,16, Harald Meller19, Oleg Mochalov20, Vyacheslav Moiseyev20, Manuel A. Rojo Guerra21, Jacob Roodenberg22, Josep Maria Vergés15,16, Johannes Krause23,24, Alan Cooper4, Curt W. Alt25,26,27,28,29, David Anthony25, Carles Lalueza–Fox28, Wolfgang Haak9,29,30, Ron Pinhasi30,31,32 & David Reich1,2,3

Ancient DNA makes it possible to observe natural selection directly by analysing samples from populations before, during and after adaptation events. Here we report a genome–wide scan for selection using ancient DNA, capitalizing on the largest ancient DNA data set yet assembled: 230 West Eurasians who lived between 6500 and 300 bc, including 163 with newly reported data. The new samples include, to our knowledge, the first genome–wide ancient DNA from Anatolian Neolithic farmers, whose genetic material we obtained by extracting from petrous bones, and who we show were members of the population that was the source of Europe’s first farmers. We also report a transect of the steppe region in Samara between 5600 and 300 bc, which allows us to identify admixture into the steppe from at least two external sources. We detect selection at loci associated with diet, pigmentation and immunity, and two independent episodes of selection on height.

The arrival of farming in Europe around 8,500 years ago necessitated adaptation to new environments, pathogens, diets and social organizations. While indirect evidence of this adaptation can be detected in patterns of genetic variation in present-day people, these patterns are often confounded by neutral processes. Ancient DNA provides a direct way to study these patterns, and should be a transformative technology for studies of selection, just as it has transformed studies of human pre-history. Until now, however, the large sample sizes required to detect selection have meant that studies of ancient DNA have concentrated on characterizing effects at parts of the genome already believed to have been affected by selection.

**Genome–wide ancient DNA from West Eurasia**

We assembled genome-wide data from 230 ancient individuals from West Eurasia dated to between 6500 and 300 bc (Fig. 1a, Extended Data Table 1, Supplementary Data Table 1 and Supplementary Information section 1). To obtain this data set, we combined published data from 67 samples from relevant periods and cultures with 163 samples for which we report new data, of which 83 have, to our knowledge, never previously been analysed (the remaining 80 samples include 67 whose targeted single nucleotide polymorphism (SNP) coverage we tripled from 390,000 (‘390k capture’) to 1,240,000 (‘1240k capture’); and 13 with shotgun data for which we generated new data using our targeted enrichment strategy3,8). The 163 samples for which we report new data are drawn from 270 distinct individuals who we screened for evidence of authentic DNA7. We used in-solution hybridization with synthesized oligonucleotide probes to enrich promising libraries for the targeted SNPs (Methods). The targeted sites include nearly all SNPs on the Affymetrix Human Origins and Illumina 610-Quad arrays, 49,711 SNPs on chromosome X, 32,681 SNPs on chromosome Y, and 47,384 SNPs with evidence of functional importance. We merged libraries from the same individual and filtered out samples with low coverage or evidence of contamination to obtain the final set of individuals. The 1240k capture gives access to genome-wide data from ancient samples with small fractions of human DNA and increases efficiency by targeting sites in the human genome that will actually be analysed. For this reason, the approach can be seen by comparing our results to the largest previously published ancient DNA study, which used a shotgun sequencing strategy5. Our median coverage on analysed SNPs

1Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. 2Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. 3Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts 02115, USA. 4Independent researcher, Santpoort-Noord, The Netherlands. 5School of Archaeology and Earth Institute, Belfield, University College Dublin, Dublin 4, Ireland. 6Institute for Anthropological Research, Zagreb 10000, Croatia. 7Department of Anthropology, Emory University, Atlanta, Georgia 30322, USA. 8Simurfit Institute of Genetics, Trinity College Dublin, Dublin 2, Ireland. 9Australian Centre for Ancient DNA, School of Biological Sciences & Environment Institute, University of Adelaide, Adelaide, South Australia 5005, Australia. 10Laboratory of Human Molecular Genetics, Institute of Molecular and Cellular Biology, Siberian Branch of the Russian Academy of Sciences, Novosibirsk 630090, Russia. 11Department of Paleolithic Archaeology, Institute of Archaeology and Ethnography, Siberian Branch of the Russian Academy of Sciences, Novosibirsk 630090, Russia. 12Centro Mixto UCM-ISCIII de Evolución y Comportamiento Humanos, 28040 Madrid, Spain. 13Departamento de Paleontología, Facultad Ciencias Geológicas, Universidad Complutense de Madrid, 28040 Madrid, Spain. 14Centro Nacional de Investigación sobre Evolución Humana (CENIEH), 40022 Burgos, Spain. 15IPHES, Institut Catalá de Paleoecologia Humana i Evolució Social, Campus Sesclades-URV, 43007 Tarragona, Spain. 16Area de Prehistoria, Universitat Rovira i Virgili (URV), 43002 Tarragona, Spain. 17Netherlands Institute in Turkey, Istiklal Caddesi, Nur-i Ziya Sokak 5, Beyoğlu 34433, Istanbul, Turkey. 18Volga Valladolid, 47002 Valladolid, Spain. 19The Netherlands Institute for the Near East, Leiden RA-2300, the Netherlands. 20Peter the Great Museum of Anthropology and Ethnography (Kunstkamera) RAS, St Petersburg 199034, Russia. 21Department of Prehistory and Archaeology, University of Valladolid, 47002 Valladolid, Spain. 22Department of Prehistory and Archaeology, University of Leiden, 2300 RA Leiden, the Netherlands. 23Max Planck Institute for the Science of Human History, D-07745 Jena, Germany. 24Institute for Archaeological Sciences, University of Basel, CH-4003 Basel, Switzerland. 25Department of Prehistory and Archaeology, Institute of Archaeology and Ethnography, Siberian Branch of the Russian Academy of Sciences, Novosibirsk 630090, Russia. 26Laboratory of Prehistory and Archaeology, Department of Prehistory and Archaeology, University of Helsinki, 00170 Helsinki, Finland. 27Danube Private University, A-3500 Krems, Austria. 28Institute for Prehistory and Archaeology, University of Vienna, 1090 Vienna, Austria. 29Max Planck Institute for the Science of Human History, D-07745 Jena, Germany. 30Australian Centre for Ancient DNA, School of Biological Sciences & Environment Institute, University of Adelaide, Adelaide, South Australia 5005, Australia. 31Institute of Prehistory and Archaeology, University of Valladolid, 47002 Valladolid, Spain. 32Laboratory of Prehistory and Archaeology, Department of Prehistory and Archaeology, University of Helsinki, 00170 Helsinki, Finland. 33Institute for Prehistory and Archaeology, University of Vienna, 1090 Vienna, Austria. 34Max Planck Institute for the Science of Human History, D-07745 Jena, Germany. 35Australian Centre for Ancient DNA, School of Biological Sciences & Environment Institute, University of Adelaide, Adelaide, South Australia 5005, Australia. 36Institute of Prehistory and Archaeology, Department of Prehistory and Archaeology, University of Helsinki, 00170 Helsinki, Finland. 37Institute for Prehistory and Archaeology, University of Vienna, 1090 Vienna, Austria.

© 2015 Macmillan Publishers Limited. All rights reserved
Insight into population transformations

To learn about the genetic affinities of the archaeological cultures for which genome-wide data are reported for the first time here, we studied either 1,055,209 autosomal SNPs when analysing 230 ancient individuals alone, or 592,169 SNPs when co-analysing them with 2,345 present-day individuals. Admixture between populations of Near Eastern ancestry and the EHG began as early as the Eneolithic (~5200–4000 BC), with some individuals resembling EHG and some resembling Yamnaya (Fig. 1b and Extended Data Fig. 2). Chalcolithic Iberians have no evidence of steppe ancestry (Fig. 1b and Extended Data Fig. 2), in contrast to central Europeans of the same period. Thus, the steppe-related ancestry that is ubiquitous across present-day Europe arrived in Iberia later than in central Europe (Supplementary Information section 2).

The Iberian Chalcolithic individuals from El Mirador cave are genetically similar to Middle Neolithic Iberians who preceded them (Fig. 1b and Extended Data Fig. 2), and have more WHG ancestry than their Early Neolithic predecessors (Fig. 1b). They also have more WHG ancestry than their Anatolian relatives (Extended Data Fig. 2), and four of them (and one Poltavka outlier) were genetically similar to the Middle Neolithic Iberians who preceded them (Fig. 1b and Extended Data Fig. 2). In contrast to central Europeans of the same period, the steppe-related ancestry that is ubiquitous across present-day Europe arrived in Iberia later than in central Europe (Supplementary Information section 2).

To understand population transformations in the Eurasian steppe, we analysed a time transect of 37 samples from the Samara region spanning ~5600–1500 BC and including the Eastern hunter-gatherers (EHG), Eneolithic, Yamnaya, Poltavka, Potapovka and Subnubya cultures. Admixture between populations of Near Eastern ancestry and the EHG began as early as the Eneolithic (~5200–4000 BC), with some individuals resembling EHG and some resembling Yamnaya (Fig. 1b and Extended Data Fig. 2). The Yamnaya from Samara and Kalmykia, the Afanasievo people from the Altai (~3300–3000 BC), and the Poltavka Middle Bronze Age (~2900–2200 BC) population that followed the Yamnaya in Samara are all genetically homogeneous, forming a tight ‘Bronze Age steppe’ cluster in PCA (Fig. 1b), sharing predominantly R1b Y chromosomes (Supplementary Information section 2). Previous work documented that such ancestry appeared east of the Urals beginning at least by the time of the Sintashta culture, and suggested that it reflected an eastward migration from the Corded Ware peoples of central Europe. However, the fact that the Subnubya also had such ancestry indicates that the Anatolian Neolithic or EEF ancestry could have come into the steppe from a more eastern source. Further evidence that migrations originating as far west as central Europe may not have had an important impact on the Late Bronze Age steppe comes from the fact that the Subnubya possess exclusively (~6%) R1a Y chromosomes (Supplementary Data Table 1), and four of them (and one Poltavka male) belonged to haplogroup R1a-Z93, which is common in central and eastern Europe. In contrast, very rare in present-day Europeans, and absent in all ancient central European samples.

Twelve signals of selection

To study selection, we created a data set of 1,084,781 autosomal SNPs in 617 samples by merging 213 ancient samples with genome-wide sequencing data from four populations of European ancestry from the 1,000 Genomes Project. Most present-day Europeans can be modelled as a mixture of three ancient populations related to Western hunter-gatherers (WHG), early European farmers (EEF) and steppe pastoralists (Yamnaya), and so to scan for selection, we divided our
samples into three groups based on which of these populations they clustered with most closely (Fig. 1b and Extended Data Table 1). We estimated mixture proportions for the present-day European ancestry populations and tested every SNP to evaluate whether its present-day frequencies were consistent with this model. We corrected for test statistic inflation by applying a genomic control correction analogous to that used to correct for population structure in genome-wide association studies. Of approximately one million non-monomorphic autosomal SNPs, the ~50,000 in the set of potentially functional SNPs were significantly more inconsistent with the model than neutral SNPs (Fig. 2), suggesting pervasive selection on polymorphisms of functional importance. Using a conservative significance threshold of \( P = 5.0 \times 10^{-8} \), and a genomic control correction of 1.38, we identified 12 loci that contained at least three SNPs achieving genome-wide significance within 1 Mb of the most associated SNP (Fig. 2, Extended Data Table 3, Extended Data Fig. 3 and Supplementary Data Table 3).

The strongest signal of selection is at the SNP (rs4988235) responsible for lactase persistence in Europe. Our data (Fig. 3) strengthens previous reports that an appreciable frequency of lactase persistence in Europe only dates to the last 4,000 years. The allele’s earliest appearance in the dataset is in a central European Bell Beaker sample (individual I0112) dated to between 2450 and 2140 BC. Two other independent signals related to diet are located on chromosome 11 near FADS1 and DHCR7. FADS1 and FADS2 are involved in fatty acid metabolism, and variation at this locus is associated with plasma lipid and fatty acid concentration. This locus has experienced independent selection in non-European populations and is likely to be a critical component of adaptation to different diets. Variants at DHCR7 and NADSYN1 are associated with circulating vitamin D levels and the most associated SNP in our analysis, rs7940244, is highly differentiated across closely related northern European populations, suggesting selection related to variation in dietary or environmental sources of vitamin D.

Two signals have a potential link to coeliac disease. One occurs at the ergothioneine transporter SLC22A4 that is hypothesized to have experienced a selective sweep to protect against ergothioneine deficiency in agricultural diets. Common variants at this locus are associated with increased risk for ulcerative colitis, coeliac disease, and irritable bowel disease and may have hitchhiked to high frequency as a result of this sweep. However, the specific variant (rs1050152, L503F) that was thought to be the target did not reach high frequency until relatively recently (Extended Data Fig. 4). The signal at ATXN2/SH2B3—also associated with coeliac disease—shows a similar pattern (Extended Data Fig. 4).

The second strongest signal in our analysis is at the derived allele of rs16891982 in SLC45A2, which contributes to light skin pigmentation and is almost fixed in present-day Europeans but occurred at much lower frequency in ancient populations. In contrast, the derived allele of SLC24A5 that is the other major determinant of light skin pigmentation in modern Europe (and that is not significant in the genome-wide scan for selection) appears fixed in the Anatolian Neolithic, suggesting that its rapid increase in frequency to around 0.9 in Early Neolithic Europe was mostly due to migration (Extended Data Fig. 4). Another pigmentation signal is at GRM5, where SNPs are associated with pigmentation possibly through a regulatory effect on nearby TYR.

We also find evidence of selection for the derived allele of rs12913832 at HERC2/OCA2, which is at 100% frequency in the European hunter-gatherers we analysed, and is the primary determinant of light eye colour in present-day Europeans. In contrast to the other loci, the range of frequencies in modern populations is within that of ancient populations (Fig. 3). The frequency increases with higher latitude, suggesting a complex pattern of environmental selection.

The TLR1–TLR6–TLR10 gene cluster is a known target of selection in Europe, possibly related to resistance to leprosy, tuberculosis or other mycobacteria. There is also a strong signal of selection at the major histocompatibility complex (MHC) on chromosome 6. The strongest signal is at rs2269424 near the genes PPT2 and EGFL8, but there are at least six other apparently independent signals in the MHC (Extended Data Fig. 3); and the entire region is significantly more associated than the genome-wide average (residual inflation of 2.07 in the region on chromosome 6 between 29–34 Mb after genome-wide genomic control correction). This could be the result of multiple sweeps, balancing selection, or increased drift as a result of background selection reducing effective population size in this gene-rich region.

We find a surprising result in six Scandinavian hunter-gatherers (SHG) from Motala in Sweden. In three of six samples, we observe the haplotype carrying the derived allele of rs3827760 in the EDAR...
gene (Extended Data Fig. 5), which affects tooth morphology and hair thickness\(^3\)\(^\text{3},\)\(^\text{34}\), has been the target of a selective sweep in East Asia\(^3\)\(^\text{5}\), and today is at high frequency in East Asians and Native Americans. The \textit{EDAR} derived allele is largely absent in present-day Europe, except in Scandinavia, plausibly owing to Siberian movements into the region millennia after the date of the Motala samples. The SHG have no evidence of East Asian ancestry\(^4\)\(^\text{3},\)\(^\text{3}\), suggesting that the \textit{EDAR} derived allele may not have originated in the main ancestral population of East Asians as previously suggested\(^3\)\(^\text{5}\). A second surprise is that, unlike closely related WHGs, the Motala samples have predominantly derived pigmentation alleles at \textit{SLC45A2} and \textit{SLC24A5}.

**Evidence of selection on height**

We also tested for selection on complex traits. The best-documented example of this process in humans is height, for which the differences between northern and southern Europe have been driven by selection\(^3\)\(^\text{6}\). To test for this signal in our data, we used a statistic that tests whether trait-affecting alleles are both highly correlated and more differentiated, compared to randomly sampled alleles\(^3\)\(^\text{7}\). We predicted genetic heights for each population and applied the test to all populations together, as well as to pairs of populations (Fig. 4). Using 180 height-associated SNPs\(^3\)\(^\text{8}\) (restricted to 169 for which we successfully obtained genotypes from at least two individuals from each population), we detect a significant signal of directional selection on height \((P = 0.002)\). Applying this to pairs of populations allows us to detect two independent signals. First, the Iberian Neolithic and Chalcolithic samples show selection for reduced height relative to both the Anatolian Neolithic \((P = 0.042)\) and the central European Early and Middle Neolithic \((P = 0.003)\). Second, we detect a signal for increased height in the steppe populations \((P = 0.030)\) relative to the central European Early and Middle Neolithic. These results suggest that the modern South–North gradient in height across Europe is due to both increased steppe ancestry in northern populations, and selection for decreased height in Early Neolithic migrants to southern Europe. We did not observe any other significant signals of polygenic selection in five other complex traits we tested: body mass index\(^3\)\(^\text{9}\) \((P = 0.20)\), waist-to-hip ratio\(^4\)\(^\text{0}\) \((P = 0.51)\), type 2 diabetes\(^4\)\(^\text{1}\) \((P = 0.37)\), inflammatory bowel disease\(^2\)\(^\text{6}\) \((P = 0.17)\) and lipid levels\(^1\)\(^\text{8}\) \((P = 0.50)\).

**Future studies of selection with ancient DNA**

Our results, which take advantage of the massive increase in sample size enabled by optimized techniques for sampling from the inner-ear regions of the petrous bone, as well as in-solution enrichment methods for targeted SNPs, show how ancient DNA can be used to perform a genome-wide scan for selection. Our results also directly document selection on loci related to pigmentation, diet and immunity, painting a picture of populations adapting to settled agricultural life at high latitudes. For most of the signals, allele frequencies of modern Europeans are outside the range of any ancient populations, indicating that phenotypically, Europeans of 4,000 years ago were different in important respects from Europeans today, despite having overall similar ancestry. An important direction for future research is to increase the sample size for European selection scans (Extended Data Fig. 6), and to apply this approach to regions beyond Europe and to other species.
Received 12 March; accepted 30 October 2015. Published online 23 November 2015.

Online Content
Methods, along with any additional Extended Data display items and Online Content, along with any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additional Extended Data display items and any additionalExtended
METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Ancient DNA analysis. We screened 433 next-generation sequencing libraries from 270 distinct samples for authentic ancient DNA using previously reported protocols.12 All libraries that we included in nuclear genome analysis were treated with uracil-DNA glycosylase (UDG) to reduce characteristic errors of ancient DNA.42,43

We performed in-solution enrichment for a targeted set of 1,237,207 SNPs using previously reported protocols.42,44 The targeted SNP set merges 394,577 SNPs first reported in ref. 7 (390k capture), and 842,630 SNPs first reported in ref. 44 (840k capture). For 67 samples for which we newly report data in this study, there was pre-existing 390k capture data.7 For these samples, we only performed 840k capture and merged the resulting sequences with previously generated 390k data. For the remaining samples, we pooled the 390k and 840k reagents together to produce a single enrichment reagent, which we call 1240k.

We attempted to sequence each enriched library up to the point where we estimated that it was economically inefficient to sequence further. Specifically, we iteratively sequenced more and more from each sample and only stopped when we estimated that the expected increase in the number of targeted SNPs hit at least once would be less than about one for every 100 new read pairs generated. After sequencing, we filtered out samples with <30,000 targeted SNPs covered at least once, with evidence of contamination based on mitochondrial DNA polymorphism,41 a high rate of heterozygosity on chromosome X despite being male,45 or an atypical ratio of X to Y sequences.

Of the targeted SNPs, 47,384 are ‘potentially functional’ sites chosen as follows (with some overlap): 1,290 SNPs identified as targets of selection in Europeans by the Composite of Multiples Signals (CMS) test;9,12 21,733 SNPs identified as significant hits by genome-wide association studies, or with known phenotypic effect (with some overlap); 1,290 SNPs identified as targets of selection in Europeans by the Human Origins SNP array;46 13,672 expression quantitative trait loci (eQTL).

For a targeted set of 1,237,207 SNPs using previously reported protocols.42,44

We investigated how the genomic control correction responded when we simulated what would happen if the allele had been under selection for 50, 100 or 200 generations, starting at the observed frequency. We then simulated what would happen if the allele had been under selection 0.1.

To detect deviations in allele frequency from expectation, we test the null hypothesis H0: pmod = C ptrue against the alternative H1: pmod unconstrained. We numerically maximize this likelihood in both the constrained and unconstrained model and use the fact that the twice difference in log-likelihood is approximately \( \chi^2 \) distributed to compute a test statistic and \( P \) value.

We defined the ancient source populations by the ‘Selection group 1’ label in Extended Data Table 1 and Supplementary Table 1 and used the 1000 Genomes CEU, GBR, IBS and TSI as the present-day populations. We removed SNPs that were monomorphic in all of these four modern populations as well as in 1000 Genomes Yoruba (TYI). We do not use FIN as one of the modern populations, because they do not fit this three-population model well. We estimated the proportions of (Hg, EAS) to be CEU = (0.196, 0.257, 0.547), GBR = (0.362, 0.229, 0.409), IBS = (0, 0.686, 0.314) and TSI = (0, 0.645, 0.355). In practice, we found that there was substantial inflation in the test statistic, most likely due to unmodelled ancestry or additional drift. To address this, we applied a genomic control correction, dividing all the test statistics by a constant, \( \lambda \), chosen so that the median \( P \) value matched the median of the null \( \chi^2 \) distribution. Excluding sites in the potentially functional set, we estimated \( \lambda = 1.38 \) and used this value as a correction throughout. One limitation of this test is that, although it identifies likely signals of selection, it cannot provide much information about the strength or date of selection.

We then simulated what would happen if the allele had been under selection in all of the modern populations by simulating a Wright–Fisher trajectory with selection for 50, 100 or 200 generations, starting at the observed frequency. We took the final frequency from this simulation, sampled observations to replace the actual observations in that population, and counted the proportion of simulations that gave a genome-wide significant result after GC correction (Extended Data Fig. 6a). We resampled sequence counts for the observed distribution for each population to simulate the impact of increasing sample size, assuming that the coverage and distribution of the sequences remained the same (Extended Data Fig. 6b).

We investigated how the genomic control correction responded when we simulated small amounts of admixture from a highly diverged population (Yoruba;
To estimate population-level genetic height in Fig. 4a, we assumed a uniform prior on [0,1] for the frequency of all height-associated alleles, and then sampled from the posterior joint frequency distribution of the alleles, assuming they were independent, using a Metropolis–Hastings sampler with a \( N(0,0.001) \) proposal density. We then multiplied the sampled allele frequencies by the effect sizes to get a distribution of genetic height.

**Test for polygenic selection.** We implemented the test for polygenic selection described by ref. 37. This evaluates whether trait-associated alleles, weighted by their effect size, are over-dispersed compared to randomly sampled alleles, in the directions associated with the effects measured by genome-wide association studies (GWAS). For each trait, we obtained a list of significant SNP associations and effect estimates from GWAS data, and then applied the test both to all populations combined and to selected pairs of populations. For height, we restricted the list of GWAS associations to 169 SNPs where we observed at least two chromosomes in all tested populations (Selection population 2). We estimated frequencies in each population by computing the maximum likelihood estimate (MLE), using the likelihood described above. For each test we sampled SNPs, frequency-matched in each population by computing the maximum likelihood estimate (MLE), using the likelihood described above. For each test we sampled SNPs, frequency-matched in 20 bins, computed the test statistic \( Q_X \) and for ease of comparison converted these to \( Z \) scores, signed according the direction of the genetic effects. Theoretically \( Q_X \) has a \( \chi^2 \) distribution but in practice, it is over-dispersed. Therefore, we report bootstrap \( P \) values computed by sampling 10,000 sets of frequency-matched SNPs.

To estimate population-level genetic height in Fig. 4a, we assumed a uniform prior on [0,1] for the frequency of all height-associated alleles, and then sampled from the posterior joint frequency distribution of the alleles, assuming they were independent, using a Metropolis–Hastings sampler with a \( N(0,0.001) \) proposal density. We then multiplied the sampled allele frequencies by the effect sizes to get a distribution of genetic height.

**Code availability.** Code implementing the selection analysis is available at https://github.com/mathii/europe_selection.

---

1000 Genomes YRI) into a randomly chosen modern population. The genomic inflation factor increases from around 1.38 to around 1.51 with 10% admixture, but there is little reduction in power (Extended Data Fig. 6c). Finally, we investigated how robust the test was to misspecification of the mixture matrix \( C \). We re-ran the power simulations using a matrix \( C = xC + (1-x)R \) for \( x \in [0,1] \) where \( R \) was a random matrix chosen so that for each modern population the mixture proportions of the three ancient populations were jointly uniformly distributed on [0,1]. Increasing \( x \) increases the genomic inflation factor and reduces power, demonstrating the advantage of explicitly modelling the ancestries of the modern populations (Extended Data Fig. 6d).

**Test for polygenic selection.** We implemented the test for polygenic selection described by ref. 37. This evaluates whether trait-associated alleles, weighted by their effect size, are over-dispersed compared to randomly sampled alleles, in the directions associated with the effects measured by genome-wide association studies (GWAS). For each trait, we obtained a list of significant SNP associations and effect estimates from GWAS data, and then applied the test both to all populations combined and to selected pairs of populations. For height, we restricted the list of GWAS associations to 169 SNPs where we observed at least two chromosomes in all tested populations (Selection population 2). We estimated frequencies in each population by computing the maximum likelihood estimate (MLE), using the likelihood described above. For each test we sampled SNPs, frequency-matched in each population by computing the maximum likelihood estimate (MLE), using the likelihood described above. For each test we sampled SNPs, frequency-matched in 20 bins, computed the test statistic \( Q_X \) and for ease of comparison converted these to \( Z \) scores, signed according the direction of the genetic effects. Theoretically \( Q_X \) has a \( \chi^2 \) distribution but in practice, it is over-dispersed. Therefore, we report bootstrap \( P \) values computed by sampling 10,000 sets of frequency-matched SNPs.

To estimate population-level genetic height in Fig. 4a, we assumed a uniform prior on [0,1] for the frequency of all height-associated alleles, and then sampled from the posterior joint frequency distribution of the alleles, assuming they were independent, using a Metropolis–Hastings sampler with a \( N(0,0.001) \) proposal density. We then multiplied the sampled allele frequencies by the effect sizes to get a distribution of genetic height.

**Code availability.** Code implementing the selection analysis is available at https://github.com/mathii/europe_selection.
Extended Data Figure 1 | Efficiency and cost-effectiveness of 1240k capture. We plot the number of raw sequences against the mean coverage of analysed SNPs after removal of duplicates, comparing the 163 samples for which capture data are newly reported in this study, against the 102 samples analysed by shotgun sequencing in ref. 5. We caution that the true cost is more than that of sequencing alone.
Extended Data Figure 2 | Early isolation and later admixture between farmers and steppe populations. a, Mainland European populations later than 3000 BC are better modelled with steppe ancestry as a third ancestral population, (closer correspondence between empirical and estimated $f_2$-statistics as estimated by resnorm; Methods). b, Later (post-Poltavka) steppe populations are better modelled with Anatolian Neolithic as a third ancestral population. c, Estimated mixture proportions of mainland European populations without steppe ancestry. d, Estimated mixture proportions of Eurasian steppe populations without Anatolian Neolithic ancestry. e, Estimated mixture proportions of later populations with both steppe and Anatolian Neolithic ancestry. f, Admixture plot at $K = 17$ showing population differences over time and space. EN, Early Neolithic; MN, Middle Neolithic; LN, Late Neolithic; BA, Bronze Age; LNBA, Late Neolithic and Bronze Age.
Extended Data Figure 3 | Regional association plots. Locuszoom plots for genome-wide significant signals. Points show the $-\log_{10}$ $P$ value for each SNP, coloured according to their linkage disequilibrium (LD; units of $r^2$) with the most associated SNP. The blue line shows the recombination rate, with scale on right hand axis in centimorgans per megabase (cM/Mb). Genes are shown in the lower panel of each subplot.
Extended Data Figure 4 | PCA of selection populations and derived allele frequencies for genome-wide significant signals. a, Ancient samples projected onto principal components of modern samples, as in Fig. 1, but labelled according to selection populations defined in Extended Data Table 1. b, Allele frequency plots as in Fig. 3. Six signals not included in Fig. 3—for SLC22A4 we show both rs272872, which is our strongest signal, and rs1050152, which was previously hypothesized to be under selection, and we also show SLC24A5, which is not genome-wide significant but is discussed in the main text.
Extended Data Figure 5 | Motala haplotypes carrying the derived, selected \textit{EDAR} allele. This figure compares the genotypes at all sites within 150 kb of rs3827760 (in blue) for the 6 Motala samples and 20 randomly chosen CHB (Chinese from Beijing) and CEU (Utah residents with northern and western European ancestry) samples. Each row is a sample and each column is a SNP. Grey means homozygous for the major (in CEU) allele. Pink denotes heterozygous and red indicates homozygous for the other allele. For the Motala samples, an open circle means that there is only a single sequence, otherwise the circle is coloured according to the number of sequences observed. Three of the Motala samples are heterozygous for rs3827760 and the derived allele lies on the same haplotype background as in present-day East Asians. The only other ancient samples with evidence of the derived \textit{EDAR} allele in this data set are two Afanasievo samples dating to 3300–3000 bc, and one Scythian dating to 400–200 bc (not shown).
Extended Data Figure 6 | Estimated power of the selection scan.

a, Estimated power for different selection coefficients (s) for a SNP that is selected in all populations for either 50, 100 or 200 generations.
b, Effect of increasing sample size, showing estimated power for a SNP selected for 100 generations, with different amounts of data, relative to the main text.
c, Effect of admixture from Yoruba (YRI) into one of the modern populations, showing the effect on the genomic inflation factor (blue, left axis) and the power to detect selection on a SNP selected for 100 generations with a selection coefficient of 0.02.
d, Effect of mis-specification of the mixture proportions. Here 0 on the x axis corresponds to the proportions we used, and 1 corresponds to a random mixture matrix.
**Extended Data Table 1** | 230 ancient individuals analysed in this study

| By population | Population | Date range  | N  | Out | Rel | Eff N Chr | Selection population 1 | Selection population 2 |
|---------------|------------|-------------|----|-----|-----|-----------|-------------------------|-------------------------|
|               | WHG        | 8.2-8.0 kya | 3  | 0   | 0   | 4.66      | HG                      | HG                      |
|               | Motala_HG  | 7.9-7.5 kya | 6  | 0   | 0   | 5.19      | HG                      | HG                      |
|               | Anatolia_Neolithic | 8.4-8.3 kya | 24 | 1   | 1   | 22.49     | EF                      | AF                      |
|               | Hungary_EN | 7.7-7.7 kya | 10 | 0   | 0   | 8.81      | EF                      | CEM                     |
|               | LBK_EN     | 7.5-7.1 kya | 15 | 0   | 0   | 11.15     | EF                      | CEM                     |
|               | Central_MN | 5.9-5.8 kya | 6  | 0   | 0   | 3.66      | EF                      | CEM                     |
|               | Iberia_EN  | 7.3-7.2 kya | 4  | 0   | 1   | 3.54      | EF                      | INC                     |
|               | Iberia_MN  | 5.9-5.6 kya | 4  | 0   | 0   | 3.47      | EF                      | INC                     |
|               | Iberia_Chalcolithic | 4.8-4.2 kya | 12 | 0   | 2   | 5.93      | EF                      | INC                     |
|               | Remedello  | 5.5-5.1 kya | 3  | 0   | 0   | 0.93      | EF                      | -                       |
|               | Iceman     | 5.4-5.1 kya | 1  | 0   | 0   | 1.90      | EF                      | -                       |
|               | Central_LNBA | 4.9-4.6 kya | 35 | 1   | 2   | 17.55     | SA                      | CLB                     |
|               | Yarmaya_Samara | 5.4-4.9 kya | 9  | 0   | 0   | 6.55      | SA                      | STP                     |
|               | Yarmaya_Kalmykia | 5.3-4.7 kya | 6  | 0   | 0   | 3.50      | SA                      | STP                     |
|               | Afanasievo | 5.3-5.0 kya | 5  | 0   | 0   | 3.01      | SA                      | STP                     |
|               | Poltavka   | 4.9-4.7 kya | 4  | 1*  | 0   | 4.28      | SA                      | STP                     |
|               | Sintashta  | 4.3-4.1 kya | 5  | 0   | 0   | 2.35      | SA                      | STP                     |
|               | Potapovka  | 4.2-4.1 kya | 3  | 0   | 0   | 0.66      | SA                      | STP                     |
|               | Snubnaya   | 3.9-3.6 kya | 12 | 1*  | 1   | 7.68      | SA                      | STP                     |
|               | Andronovo  | 3.8-3.6 kya | 3  | 1*  | 0   | 3.87      | SA                      | STP                     |
|               | Russia_EBA | 4.9-4.5 kya | 1  | 0   | 0   | 0.21      | SA                      | -                       |
|               | Northern_LNBA | 4.9-4.5 kya | 10 | 0   | 0   | 3.81      | SA                      | -                       |
|               | Bell_Beaiker_LNBA | 4.5-4.5 kya | 17 | 0  | 1   | 6.64      | SA^                     | CLB                     |
|               | Hungary_BA | 4.2-4.1 kya | 12 | 0  | 0   | 4.18      | SA^                     | CLB                     |
|               | EHG        | 7.7-7.6 kya | 3  | 0   | 0   | 2.15      | -                       | -                       |
|               | Samara_Eneolithic | 7.2-6.0 kya | 3  | 0   | 0   | 1.07      | -                       | -                       |
|               | Scythian_IA | 2.4-2.2 kya | 1  | 0   | 0   | 1.26      | -                       | -                       |

**By selection population**

| Selection population 1 | Date range  | N  | Out | Rel | Eff N Chr | Description |
|-------------------------|-------------|----|-----|-----|-----------|-------------|
| EF                      | 8.4-4.2 kya | 79 | 0   | 0   | 61.88     | Early Farmer |
| HG                      | 8.2-7.5 kya | 9  | 0   | 0   | 9.85      | Hunter-gatherer |
| SA                      | 5.4-3.6 kya | 93 | 3   | 0   | 52.14     | Steppe Ancestry |

| Selection population 2 | Date range  | N  | Out | Rel | Eff N Chr | Description |
|-------------------------|-------------|----|-----|-----|-----------|-------------|
| AN                      | 8.4-8.3 kya | 24 | 0   | 0   | 22.49     | Anatolian Neolithic |
| CEM                     | 7.7-5.8 kya | 31 | 0   | 0   | 23.62     | Central European Early and Middle Neolithic |
| INC                     | 7.3-4.2 kya | 20 | 0   | 0   | 12.95     | Iberian Neolithid and Chalcolithic |
| HG                      | 8.2-7.5 kya | 9  | 0   | 0   | 9.85      | Hunter-gatherer |
| CLB                     | 4.9-4.1 kya | 64 | 0   | 0   | 28.38     | Central European Late Neolithic and Bronze Age |
| STP                     | 5.4-3.6 kya | 47 | 3   | 0   | 30.58     | Steppe |

Population, samples grouped by a combination of date, source, archaeology and genetics; Date range, approximate date range of samples in this group; N, number of individuals sampled; Out, number of PCA outliers (marked with an asterisk if used in selection analysis); Rel, number of related individuals removed; Eff N Chr, average over sites of the effective number of chromosomes when we use genotype likelihoods, computed as 2 per called site for samples with genotype calls, or $2 - 0.5^{c - 1}$ for samples with read depth c; Selection population 1, coarse population labels (marked with a caret if not used in genome-wide scan); Selection population 2, fine population labels. E/M/LN, Early/Middle/Late Neolithic; LBK, Linearbandkeramik; E/S/WHG, Eastern/Scandinavian/Western hunter-gatherer; EBA, Early Bronze Age; IA, Iron Age.
Extended Data Table 2 | Key *f*-statistics used to support claims about population history

| A             | B           | C       | D       | *f*(A, B, C, D) | Z        | Number of SNPs | Interpretation                                                                 |
|---------------|-------------|---------|---------|----------------|----------|----------------|-------------------------------------------------------------------------------|
| Anatolia_Neolithic | LBK_EN      | WHG     | Chimp   | -0.00114       | -6.8     | 1003751        | Early European Farmers had more WHG ancestry than Anatolian Neolithic          |
| Anatolia_Neolithic | Hungary_EN  | WHG     | Chimp   | -0.00212       | -11.9    | 929553         |                                                                                  |
| Anatolia_Neolithic | Iberia_EN   | WHG     | Chimp   | -0.00244       | -9.6     | 904437         |                                                                                  |
| Iberia_EN     | Iberia_Chalcolthic | WHG   | Chimp   | -0.00311       | -10.5    | 802471         | Iberian Chalcolthic had more WHG ancestry than Iberian Early Neolithic         |
| Iberia_MN     | Iberia_Chalcolthic | WHG   | Chimp   | 0.00010        | 0.3      | 779905         | Iberian Chalcolthic did not have more WHG ancestry than Iberian Middle Neolithic |
| EHG           | Samara_Eneolithic | MA1   | Chimp   | 0.00140        | 2.3      | 463388         | First dilution of Ancient North Eurasian ancestry (prior to the Bronze Age Yamnaya culture) |
| EHG           | Ymnaya_Samara | MA1     | Chimp   | 0.00513        | 10.6     | 645211         |                                                                                  |
| Samara_Eneolithic | Ymnaya_Samara | MA1   | Chimp   | 0.00366        | 7.6      | 482492         |                                                                                  |
| EHG           | Ymnaya_Samara | Armenian | Chimp   | -0.00191       | -6.1     | 547370         | Contribution of Near Eastern ancestry to the Bronze Age Yamnaya culture       |
| EHG           | Ymnaya_Kalmykia | Armenian | Chimp   | -0.00180       | -5.4     | 536989         |                                                                                  |
| Samara_Eneolithic | Ymnaya_Samara | Armenian | Chimp   | -0.00100       | -3.3     | 405599         |                                                                                  |
| EHG           | Poltava     | Armenian | Chimp   | -0.00175       | -4.9     | 541983         |                                                                                  |
| Ymnaya_Samara | Ymnaya_Kalmykia | MA1   | Chimp   | -0.00010       | -0.3     | 675630         | Stability of Ancient North Eurasian ancestry between Early Bronze Age Yamnaya from Kalmykia and Samara, and the Middle Bronze Age Poltava            |
| Ymnaya_Samara | Poltava     | MA1     | Chimp   | -0.00014       | -0.4     | 673726         |                                                                                  |
| Ymnaya_Kalmykia | Poltava     | MA1     | Chimp   | 0.00012        | 0.3      | 659346         |                                                                                  |
| Ymnaya_Samara | Srubnaya    | MA1     | Chimp   | 0.00151        | 5.1      | 691149         | Second dilution of Ancient North Eurasian ancestry (prior to the Late Bronze Age Srubnaya culture)                                        |
| Ymnaya_Kalmykia | Srubnaya    | MA1     | Chimp   | 0.00161        | 4.8      | 676735         |                                                                                  |
| Poltava       | Srubnaya    | MA1     | Chimp   | 0.00164        | 4.5      | 674756         |                                                                                  |
| Ymnaya_Samara | Srubnaya    | LBK_EN  | Chimp   | -0.00225       | -11.4    | 974659         | Arrival of Early European Farmer-related ancestry prior to the Late Bronze Age Srubnaya culture. Statistics with Anatolia_Neolithic instead of LBK_EN are similar (Z<6, not shown). |
| Ymnaya_Kalmykia | Srubnaya    | LBK_EN  | Chimp   | -0.00264       | -11.4    | 951827         |                                                                                  |
| Poltava       | Srubnaya    | LBK_EN  | Chimp   | -0.00210       | -9.0     | 948968         |                                                                                  |
| EHG           | Ymnaya_Samara | Armenian | LBK_EN  | -0.00080       | -5.0     | 559478         | Different source of dilution of Ancient North Eurasian ancestry prior to the Yamnaya (Near Eastern) vs. to the Srubnaya (Early European Farmer-related) |
| EHG           | Ymnaya_Kalmykia | Armenian | LBK_EN  | -0.00086       | -5.2     | 548882         |                                                                                  |
| EHG           | Poltava     | Armenian | LBK_EN  | -0.00069       | -4.1     | 553996         |                                                                                  |
| Ymnaya_Samara | Srubnaya    | Armenian | LBK_EN  | 0.00138        | 13.1     | 585240         |                                                                                  |
| Ymnaya_Kalmykia | Srubnaya    | Armenian | LBK_EN  | 0.00142        | 11.3     | 574333         |                                                                                  |
| Poltava       | Srubnaya    | Armenian | LBK_EN  | 0.00134        | 10.7     | 577082         |                                                                                  |

| Ref1 | Ref2 | Test  | *f*(Test; Ref1, Ref2) | Z       | Number of SNPs | Interpretation                                                                 |
|------|------|-------|-----------------------|---------|----------------|-------------------------------------------------------------------------------|
| WHG  | Anatolia_Neolithic | Hungary_EN | -0.00412        | -6.7     | 548445         | Early European farmers were formed by admixture between Anatolia Neolithic and WHG (the non-significant signal in the Iberian_EN may be due to genetic drift specific to this population) |
| WHG  | Anatolia_Neolithic | LBK_EN     | -0.00257        | -4.6     | 654357         |                                                                                  |
| WHG  | Anatolia_Neolithic | Iberia_EN  | 0.00179         | 1.4      | 389101         |                                                                                  |
| EHG  | Armenian | Poltava | -0.00539        | -3.9     | 213055         | Early and Middle Bronze Age steppe pastoralists were formed by admixture between EHG and a population of Near Eastern ancestry |
| EHG  | Armenian | Ymnaya_Kalmykia | -0.00537 | -4.2     | 213996         |                                                                                  |
| EHG  | Armenian | Ymnaya_Samara | -0.00586 | -6.2     | 276568         |                                                                                  |
| LBK_EN | Ymnaya_Samara | Srubnaya | -0.00630        | -11.2    | 584111         | Srubnaya was formed by admixture between populations related to Yamnaya and Early European Farmers |
Extended Data Table 3 | Twelve genome-wide significant signals of selection

| Lead SNP  | Chromosome | Position (hg19) | P Value | Range (Mb) | Genes       | Potential function                                  |
|-----------|------------|----------------|---------|------------|-------------|-----------------------------------------------------|
| rs4988235 | 2          | 136,608,646    | 3.19E-49 | 135.3-137.3| MCM6, LCT   | Lactase persistence*15                              |
| rs16891982| 5          | 33,951,693     | 7.05E-40 | 33.8-34.0  | SLC45A2     | Skin pigmentation*53                                |
| rs2269424 | 6          | 32,132,233     | 5.41E-21 | 29.9-33.1  | MHC region  | Immunity*                                           |
| rs174546  | 11         | 61,569,830     | 8.18E-19 | 61.5-61.6  | FADS1, FADS2| Fatty acid metabolism*18, 54, 55                     |
| rs4833103 | 4          | 38,815,502     | 2.58E-18 | 38.7-38.8  | TLR1, TLR6, TLR10 | Immunity*31, 32                                    |
| rs653178  | 12         | 112,007,756    | 1.96E-13 | 111.9-112.6| ATXN2, SH2B3| Unknown                                             |
| rs7944926 | 11         | 71,165,625     | 2.86E-13 | 71.2-71.2  | DHCR7, NADSYN1 | Vitamin D metabolism*21, 56                        |
| rs7119749 | 11         | 88,515,022     | 2.03E-12 | 88.5-88.9  | GRM5        | Skin pigmentation*57                                |
| rs272872  | 5          | 131,675,864    | 2.56E-12 | 131.4-131.8| SLC22A4     | Ergothioneine transport*57                          |
| rs6903823 | 6          | 28,322,296     | 2.96E-11 | 28.3-28.7  | ZKSCAN3, ZSCAN31 | Autophagy*58, Lung function*59                      |
| rs1979866 | 13         | 38,825,900     | 4.60E-11 | 38.1-38.8  | -           | Unknown                                             |
| rs12913832| 11         | 28,365,618     | 1.50E-10 | 29.9-30.1  | HERC2, OCA2 | Eye color*28, 29                                    |

Chromosome/Position/Range, co-ordinates (hg19) of the SNP with the most significant signal, and the approximate range in which genome-wide significant SNPs are found. Genes, genes in which the top SNP is located, and selected nearby genes. Potential function, function of the gene, or specific trait under selection. Marked with an asterisk if the signal was still genome-wide significant in an analysis that used only the populations that correspond best to the three ancestral populations (WHG, Anatolian Neolithic and Bronze Age steppe), resulting in a less powerful test with the effective number of chromosomes analysed at the average SNP reduced from 125 to 50, a genomic control correction of 1.32, and five genome-wide significant loci that are a subset of the original twelve. Refs 53–59 are cited in this table.