Genetic contributions to variation in human stature in prehistoric Europe

Samantha L. Cox[a], Christopher B. Ruff[b], Robert M. Maier[c,d,e] & Iain Mathieson[a,1]

[a] Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia PA.
[b] Center for Functional Anatomy and Evolution, Johns Hopkins University School of Medicine, Baltimore MD.
[c] Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[d] Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[e] Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA 02114, USA
[1] Correspondence to mathi@pennmedicine.upenn.edu

Abstract

The relative contributions of genetics and environment to temporal and geographic variation in human height remain largely unknown. Ancient DNA has identified changes in genetic ancestry over time, but it is not clear whether those changes in ancestry are associated with changes in height. Here, we directly test whether changes over the past 38,000 years in European height predicted using DNA from 1071 ancient individuals are consistent with changes observed in 1159 skeletal remains from comparable populations. We show that the observed decrease in height between the Early Upper Paleolithic and the Mesolithic is qualitatively predicted by genetics. Similarly, both skeletal and genetic height remained constant between the Mesolithic and Neolithic and increased between the Neolithic and Bronze Age. Sitting height changes much less than standing height–consistent with genetic predictions–although genetics predicts a small Bronze Age increase that is not observed in skeletal remains. Geographic variation in stature is also qualitatively consistent with genetic predictions, particularly with respect to latitude. We find that the changes in genetic height between the Neolithic and Bronze Age may be driven by polygenic adaptation. Finally, we hypothesize that an observed decrease in genetic heel bone mineral density in the Neolithic reflects adaptation to the decreased mobility indicated by decreased femoral bending strength. This study provides a model for interpreting phenotypic changes predicted from ancient DNA and demonstrates how they can be combined with phenotypic measurements to understand the relative contribution of genetic and developmentally plastic responses to environmental change.

Introduction

Stature, or standing height, is one of the most heavily studied human phenotypes. It is easy to measure in living individuals and relatively straightforward to estimate from skeletal remains. As a consequence, geographic variation and temporal changes in stature are well documented (1-3), particularly in western Europe, where there is a comprehensive record of prehistoric changes (4). The earliest anatomically modern humans in Europe, present by 42-45,000 BP (5, 6), were relatively tall (mean adult male height in the Early Upper Paleolithic was ~174 cm). Mean male stature then declined from the Paleolithic to the Mesolithic (~164 cm) before increasing to ~167 cm by the Bronze Age (4, 7). Subsequent changes, including the 20th century secular trend increased height to ~170-180 cm (1, 4). It is broadly agreed that these changes are likely to have been driven by a combination of environmental (e.g. climate or diet) and genetic factors (4, 7-9), although the effects of these two variables cannot be separated based on skeletal data alone. In this study, by combining the results of genome-wide association studies (GWAS) with ancient DNA, we directly estimate the genetic component of stature and test whether population-level skeletal changes between ~35,000 and 1,000 BP are consistent with those predicted by genetics.

Height is highly heritable (10-14), and therefore amenable to genetic analysis by genome-wide association studies (GWAS). With sample sizes of hundreds of thousands of individuals, GWAS have identified thousands of genomic variants that are significantly associated with the phenotype (15-17). Though the individual effect of each of these variants is tiny (on the order of +/- 1-2mm per variant (18)), their combination can be highly predictive. Polygenic risk scores (PRS) constructed by summing together the effects of all height-associated variants carried by an individual can now explain upwards of 30% of the phenotypic variance in populations of European ancestry (16). In effect, the PRS can be thought of as an estimate of “genetic height” that predicts phenotypic height, at least in populations closely related to those in which the GWAS was performed. One major caveat is that the predictive power of PRS is much lower in other populations (19). The extent to which differences in PRS between populations are predictive of population-level differences in phenotype is currently unclear (20). Recent studies

60 have demonstrated that such differences may partly be artifacts of correlation between
61 environmental and genetic structure in the original GWAS (21, 22). These studies also
62 suggested best practices for PRS comparisons, including the use of GWAS summary statistics
63 from large homogenous studies (instead of meta-analyses), and replication of results using
64 summary statistics derived from within-family analyses that are robust to population
65 stratification.
66
67 Bearing these caveats in mind, PRS can be applied to ancient populations thanks to recent
68 technological developments that have dramatically increased ancient DNA (aDNA) sample sizes.
69 These have provided remarkable insights into the demographic and evolutionary history of both
70 modern and archaic humans across the world (23-25), particularly in Europe, and allow us to
71 track the evolution of variants underlying phenotypes ranging from pigmentation to diet (26-
72 29). In principle, PRS applied to ancient populations could similarly allow us to make inference
73 about the evolution of complex traits. A few studies have used PRS to make predictions about
74 the relative statures of ancient populations (29-31) but looked at only a few hundred samples in
75 total and did not compare their predictions with stature measured from skeletons. Here, we
76 compare measured skeletal data to genetic predictions and directly investigate the genetic
77 contribution to height independent of environmental effects acting during development.

78 Results

79 PRS and skeletal measurements

80 We collected published aDNA data from 1071 ancient individuals from Western Eurasia (west of
81 50° E), dated to between 38,000 and 1100 years before present (BP) (27, 29, 30, 32-57). Using
82 GWAS summary statistics for height from the UK Biobank (generated and made available by the
83 Neale lab: http://nealelab.is/), we computed height PRS for each individual, using a P-value
84 cutoff of 10^{-6}, clumping variants in 250kb windows, and replacing missing genotypes with the
85 mean across individuals (Methods). We refer to this as PRS(GWAS). Because of concerns about
86 GWAS effect sizes being inflated by residual population stratification, we also computed a PRS
87 where we used GWAS P-values to select SNPs, but computed the PRS using effect sizes

88 estimated from a within-family test for ~17,000 sibling pairs from UK Biobank (Methods) which
89 we refer to as PRS(GWAS/Sibs), and which should be unaffected by stratification. We also
90 obtained stature estimates from 1159 individuals dating to between 33,700 and 1100 BP taken
91 from a larger dataset of 2177 individuals with stature and body proportion estimates from
92 substantially complete skeletons (4, 58). There is limited overlap in these datasets (12
93 individuals), but they cover the same time periods and broadly the same geographic locations
94 (Supplementary Fig. 1), although the genetic data contain more individuals from further east
95 (30-50° E) compared to the skeletal data. We divided these individuals into five groups based
96 on date: Early Upper Paleolithic (>25,000 BP; EUP), Late Upper Paleolithic (25,000-11,000 BP;
97 LUP), Mesolithic (11,000-5500 BP), Neolithic (8500-3900 BP) and post-Neolithic (5000-1100 BP,
98 including the Copper and Bronze Ages, plus later periods). These groups broadly correspond to
99 transitions in both archaeological culture and genetic ancestry (33, 38, 59), and we resolved
100 individuals in the overlapping periods using either archaeological or genetic context (Methods).
101

102 ## Trends in PRS for height are largely consistent with trends in skeletal stature

103 We found a significant effect of group (time period) on mean PRS(GWAS) (ANOVA P= 1.9×10^{-9}),
104 PRS(GWAS/Sibs) (P=0.045) and skeletal stature (P=2.8×10^{-11}). There was no evidence of
105 difference between LUP, Mesolithic and Neolithic groups (Supplementary Fig. 2a-b), so we
106 merged these three groups (we refer to the merged group as LUP-Neolithic). We find that
107 PRS(GWAS) in the LUP-Neolithic period is 0.47 standard deviations (SD) lower than in the EUP
108 (P=0.002), and 0.40 SD lower (P= 8.7×10^{-11}) than in the post-Neolithic period (Fig. 1a).
109 PRS(GWAS/Sib) shows a very similar pattern (Fig. 1b), demonstrating that this is not a result of
110 differential relatedness of the ancient individuals to the structured present-day GWAS
111 populations. Skeletal stature shows a qualitatively similar pattern to the genetic predictions,
112 with a 1.5 SD (9.6cm; P=2.9×10^{-7}) difference between EUP and LUP-Neolithic and a 0.27 SD
113 (1.8cm; P=$3.6\text{x}10^{-5}$) difference between LUP-Neolithic and post-Neolithic. Broad patterns of
114 change in stature over time are therefore consistent with genetic predictions.
115

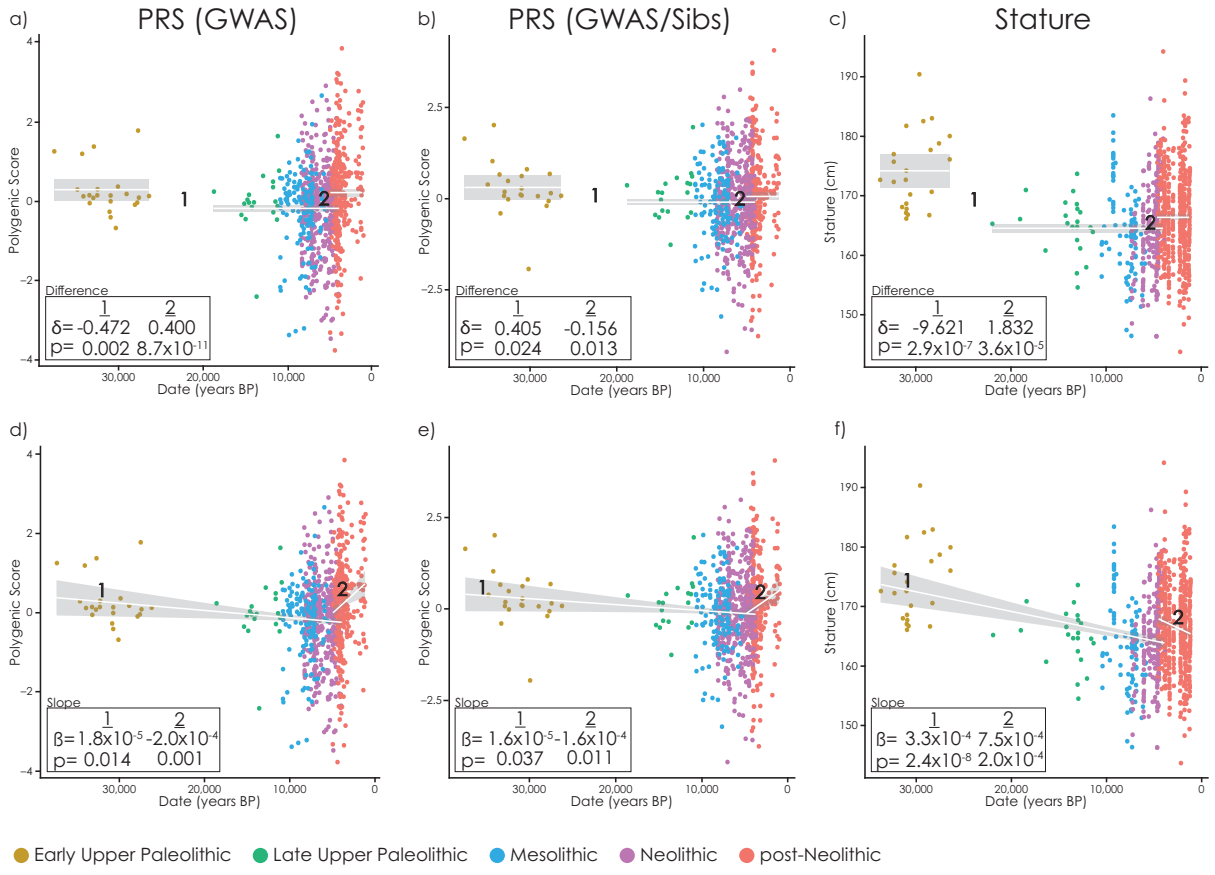


Figure 1: Changes in standing height PRS and stature though time. Each point is an ancient individual, white lines show fitted values, grey area is the 95% confidence interval, and boxes show parameter estimates and p-values for difference in means (δ) and slopes (β). **a-c**) PRS(GWAS) (a), PRS(GWAS/Sibs) (b) and skeletal stature (c) with constant values in the EUP, LUP-Neolithic and post-Neolithic. **d-e**) PRS(GWAS) (d), PRS(GWAS/Sibs) (e) and skeletal stature (f) showing a linear trend between EUP and Neolithic and a different trend in the post-Neolithic.

Additionally, we fit a piecewise linear model allowing PRS to decrease from the EUP to the Neolithic and then increase and change slope in the post-Neolithic (Fig. 1d-f). In this model, PRS(GWAS) decreases by about 1.8×10^{-5} SD/year (P=0.014) from EUP to Neolithic, and increases by 2.0×10^{-4} SD/year (P=0.001) post-Neolithic (Fig. 1d). PRS(GWAS/sib) decreases by about 1.6×10^{-5} SD/year (P=0.037) from EUP to Neolithic, then increases by 1.6×10^{-4} SD/year throughout the period (P=0.011; Fig. 1e). Again, these changes are qualitatively consistent with changes in stature (Fig. 1f), with a 4.7×10^{-5} SD/year (3.3×10^{-4} cm/year; $P=2.4\times10^{-8}$) decrease from EUP to Mesolithic, and an increase of ~0.5 SD into the Neolithic. However, in this model stature, unlike PRS, actually decreases during the post-Neolithic period (7.5×10^{-4} cm/year; $P=2.0\times10^{-4}$).

To further explore these trends, we fitted a broader range of piecewise linear models to both datasets (Methods; Supplementary Table 1; Supplementary Fig. 3-5). In the most general model we allowed both the mean and the slope of PRS or stature with respect to time to vary between groups. More constrained models fix some of these parameters to zero–eliminating change over time–or merging two adjacent groups. We compared the fit of these nested models using Akaike's Information Criterion (AIC, Supplementary Table 1). The linear model in Fig. 1d-f is one of the best models in this analysis. In general, all the best-fitting models support the pattern–for both PRS and measured stature–of a decrease between the EUP and Mesolithic and an increase between the Neolithic and post-Neolithic (Supplementary Fig. 3-5). Some models suggest that the increase in stature–but not PRS–may have started during the Neolithic (Supplementary Figure 5a-c). Finally, we confirmed that these results were robust to different constructions of the PRS–using 100kb and 500kb clustering windows rather than 250kb (Supplementary Fig. 6-7).

Sitting height PRS is partially consistent with trends in body proportions

Standing height is made up of two components: leg length and sitting height (made up of the length of the trunk, neck and head), with a partially overlapping genetic basis (60). Throughout European prehistory, changes in leg length tended to be larger than changes in sitting height

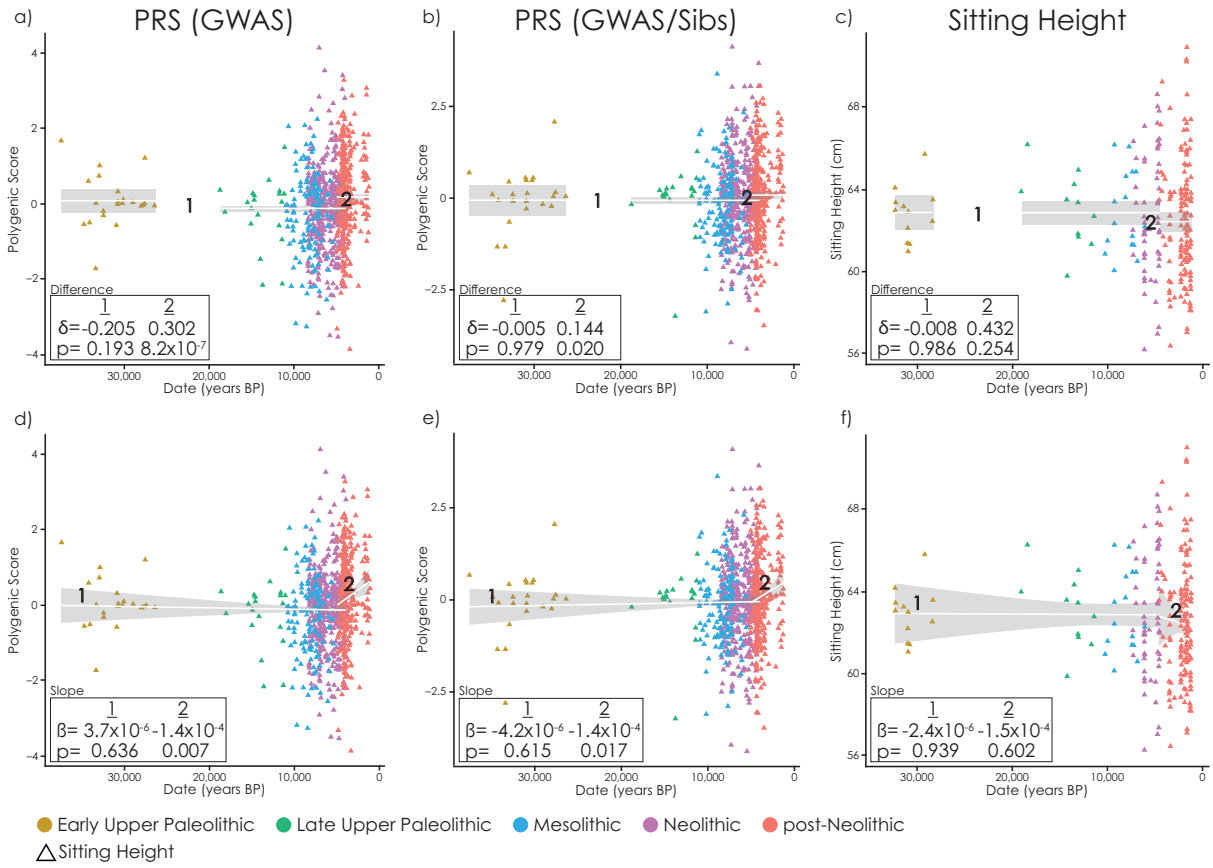


Figure 2: Changes in sitting height PRS and sitting height though time. Each point is an ancient individual, lines show fitted values, grey area is the 95% confidence interval, and boxes show parameter estimates and p-values for difference in means (δ) and slopes (β). **a-c**) PRS(GWAS) (a), PRS(GWAS/Sibs) (b) and skeletal sitting height, with constant values in the EUP, LUP-Neolithic and post-Neolithic. **d-e**) PRS(GWAS) (d), PRS(GWAS/Sibs) (e) and skeletal sitting height showing a linear trend between EUP and Neolithic and a different trend in the post-Neolithic.

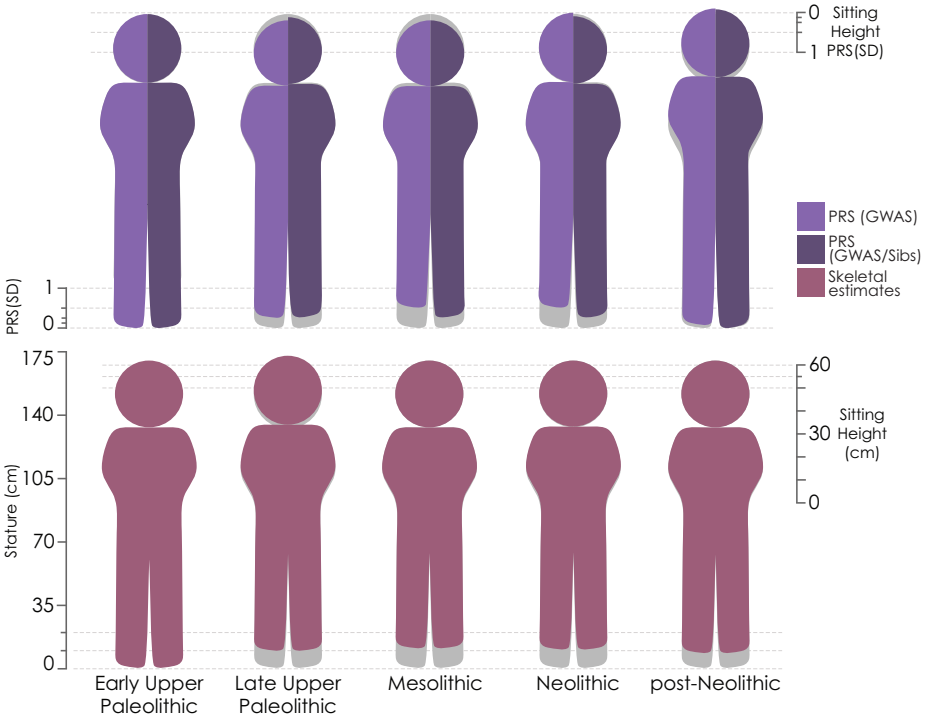


Figure 3: Changes in standing height and sitting height predicted using PRS (relative changes; upper row) and measured from skeletons (absolute values; lower row). Y-axes on the left represent changes in standing height and on the right represent changes in sitting height. Note that changes in PRS are exaggerated compared to the total height of the figures.

145 (4). We constructed PRS(GWAS) and PRS(GWAS/Sibs) for sitting height and analyzed them in
146 the same way as standing height (Fig. 2). In contrast to standing height, we find no evidence of
147 change between the EUP and Neolithic. Both PRS(GWAS) and PRS(GWAS/Sibs) do increase,
148 either between the Neolithic and post-Neolithic, or during the post-Neolithic period (Fig. 2a,b,d
149 & e). On the other hand, using only skeletons with complete torsos to estimate sitting height,
150 we find no evidence of change in any period. Thus, the skeletal data are consistent with the
151 genetic data for the EUP-Neolithic period, but inconsistent in the post-Neolithic period, where
152 PRS predicts an increase that is not reflected in the skeletons. This could be because of more
153 limited skeletal measurements (only 236 out of 1159 skeletons are sufficiently complete to
154 estimate sitting height directly), because the change in PRS is artefactual, it is being buffered by
155 non-genetic effects, or by opposing genetic effects which we do not capture. Overall, we find
156 mixed consistency between PRS and skeletal measurements (Fig. 3). The decrease in standing
157 but not sitting height between the EUP and Neolithic is consistent in both, as is the increase in
158 standing height between the Neolithic and post-Neolithic. However, PRS predicts a continued
159 increase in stature through the post-Neolithic period that is not seen in skeletal remains.
160

161 ## Geographic variation in standing height

162 As well as varying through time, human stature is stratified by geography, with trends related
163 to both longitude and latitude (61). North-South trends following Allen's (62) and Bergmann's
164 (63) rules are most often interpreted as environmental adaptations to the polar-equatorial
165 climate gradient. Today, Northern Europeans are generally taller than Southern Europeans (1),
166 a pattern which emerged between the Mesolithic and post-Neolithic (4, 7). Longitudinal
167 variation within Europe is present during the Mesolithic (64), though these trends are difficult
168 to interpret due to sampling bias across the time period (4). We therefore tested whether
169 geographic variation in PRS could explain these geographic trends, as it partially explains
170 temporal trends.
171

172 We regressed the residuals from our fitted linear height model (the model shown in Fig. 1d-f)
173 on longitude and latitude. Stature increases significantly with latitude ($P=1.2\times10^{-10}$) in the post-

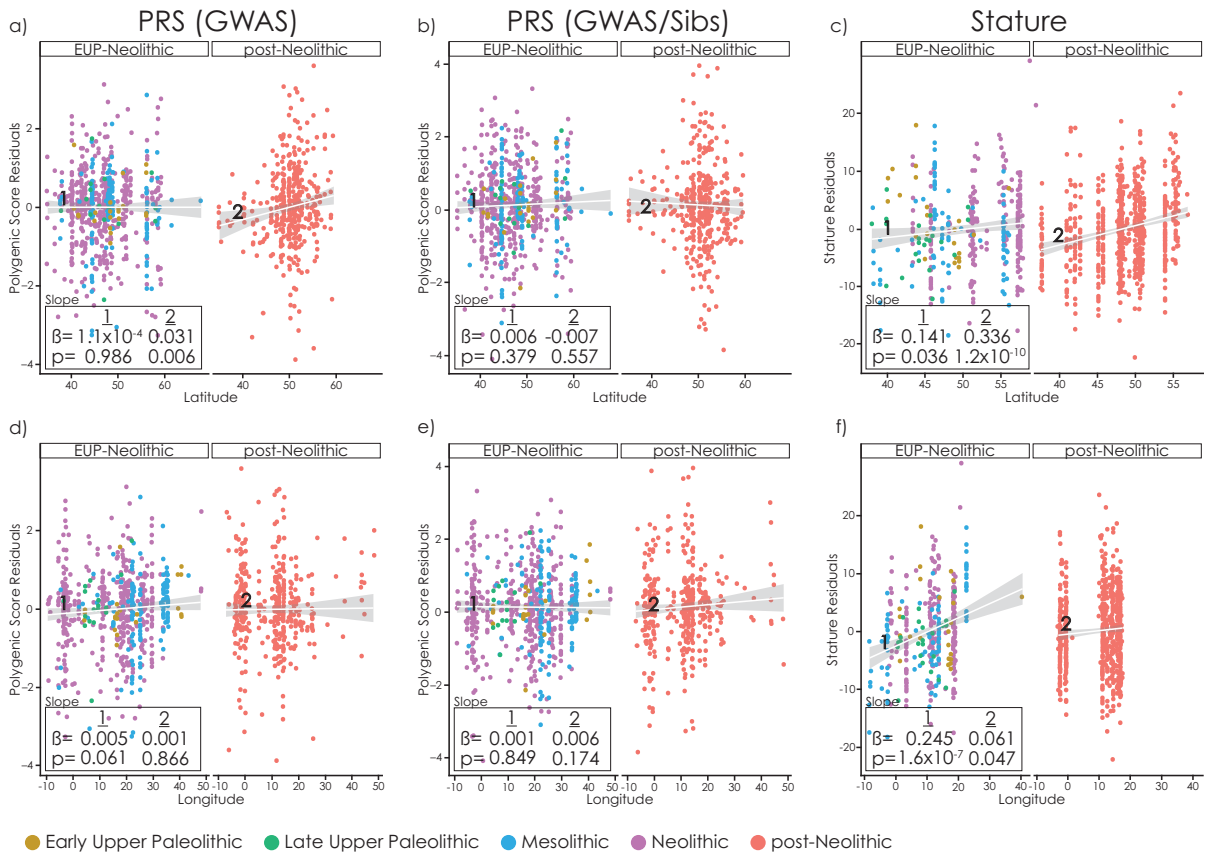


Figure 4: Geographic variation in PRS and skeletal standing height. Residuals for the linear height model (Fig. 1 d-f) against **a-c**) latitude and **d-f**) longitude. Each point is an ancient individual, lines show fitted values, grey area is the 95% confidence interval, and boxes show parameter estimates (β) and p-values for slopes.

174 Neolithic period. PRS(GWAS) increases in the post-Neolithic (P=0.006) although this is not
175 replicated by PRS(GWAS/Sibs) (P=0.557). PRS does not increase significantly with latitude in the
176 EUP-Neolithic period. There is some evidence of a modest trend in stature in the EUP-Neolithic
177 period (Fig. 4c). However, there is only evidence for this in the Neolithic, not in the EUP-
178 Mesolithic (Supplementary Fig. 8a). Further, because time and geography are correlated in our
179 Neolithic sample, this can also be explained by a temporal increase during the Neolithic, in
180 which case there is no geographic trend (Supplementary Fig. 8b).
181
182 In contrast to latitude, there is a significant increasing trend of stature with longitude before
183 but not during the Neolithic (0.36 cm/degree $P=1.6\times10^{-7}$; Fig. 4, Supplementary Fig. 8c). This
184 may be partly driven by a small number of samples from a single site, but still persists if these
185 samples are removed (0.20 standardized residuals per degree, P=0.004; Supplementary Fig. 8d).
186 There is little or no trend (0.06 cm/degree; P=0.047) in the post-Neolithic period (Figure 4f). We
187 find no evidence for longitudinal clines in PRS. In summary, we find that stature increases with
188 latitude in the post-Neolithic, possibly in the Neolithic, but not before. This cline may have a
189 genetic basis. Stature also increases with longitude, particularly in the Mesolithic, but this cline
190 is not predicted by genetics.
191

192 ## Correlated changes in bone density PRS and femoral bending strength

193 Beyond stature, we wanted to investigate the utility of using PRS to interpret other measurable
194 phenotypes in ancient individuals. Decreased mobility though time, associated with large-scale
195 lifestyle transitions between hunting-gathering, agriculture, and ultimately modern
196 industrialism, is well documented through declines in lower limb bone diaphyseal strength and
197 trabecular density (4, 65, 66). Today, heel bone mineral density (hBMD) is often used as an
198 indicator of general activity levels in younger people (67) and of osteoporosis in older
199 individuals (68, 69); UK Biobank has GWAS data for this trait, indirectly estimated by
200 ultrasound. However, evaluating differences in BMD in archaeological and paleontological
201 specimens can be problematic. In the short term soil leaches bone minerals, while later the
202 bone begins to fossilize, leading to unpredictable patterns of density in ancient remains (70)

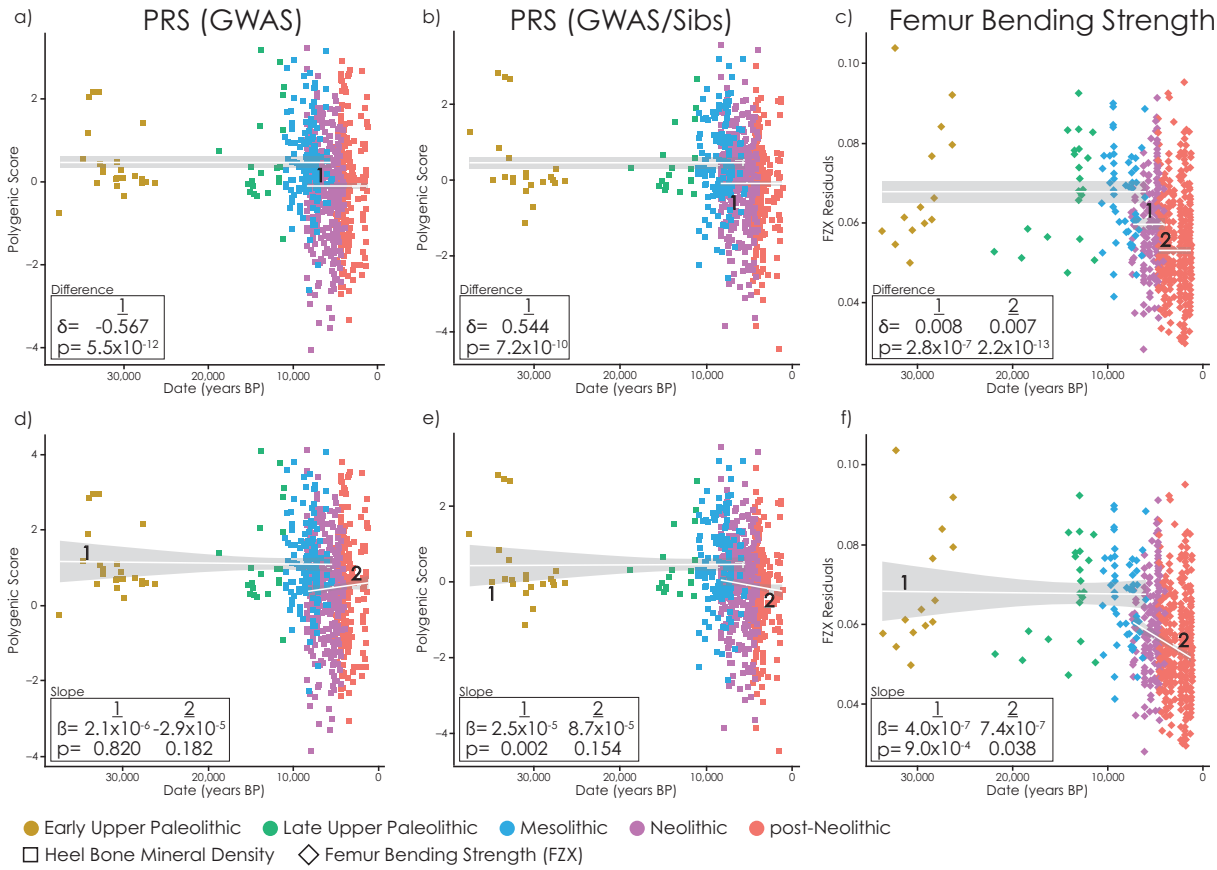


Figure 5: Changes in heel bone mineral density (hBMD) PRS and femur bending strength (FZx) though time. Each point is an ancient individual, lines show fitted values, grey area is the 95% confidence interval, and boxes show parameter estimates and p-values for difference in means (δ) and slopes (β). **a-b**) PRS(GWAS) (a) and PRS(GWAS/Sibs) (b) for hBMD, with constant values in the EUP-Mesolithic and Neolithic-post-Neolithic. **c**) FZx constant in the EUP-Mesolithic, Neolithic and post-Neolithic. **d-e**) PRS(GWAS) (d) and PRS(GWAS/Sibs) (e) for hBMD showing a linear trend between EUP and Mesolithic and a different trend in the Neolithic-post-Neolithic. **f**) FZx with a linear trend between EUP and Mesolithic and a different trend in the Neolithic-post-Neolithic.

203 and requiring special processing methods (65) that are difficult to apply to large samples.
204 However, femoral diaphyseal bending strength can be calculated from bone cross-sectional
205 geometric measurements that are not as affected by bone preservation (71). Here we focus on
206 anteroposterior bending strength (section modulus) of the midshaft femur (FZx), which has
207 been linked specifically to mobility (72). Since both trabecular density and diaphyseal strength
208 should respond to mobility and activity levels, we reasoned that they would be likely to show
209 correlated patterns of temporal change. Following established protocols (71), we standardized
210 FZx first by sex, then the product of estimated body mass and femoral length (4).
211 Qualitatively, PRS(GWAS) and FZx show similar patterns, decreasing through time (Fig. 5,
212 Supplementary Figure 1g-i). There is a significant drop in FZx (Figure 5c) from the Mesolithic to
213 Neolithic (P= 1.2×10^{-8}) and again from the Neolithic to post-Neolithic (P=1.5×10^{-13}). PRS(GWAS)
214 for hBMD decreases significantly from the Mesolithic to Neolithic (Figure 5a; P=5.5×10^{-12}),
215 which is replicated in PRS(GWAS/Sibs) (P=$7.2\text{x}10^{-10}$; Figure 5b); neither PRS shows evidence of
216 decrease between the Neolithic and post-Neolithic. We hypothesize that both FZx and hBMD
217 responded to the reduction in mobility that accompanied the adoption of agriculture (72). In
218 particular, the lower genetic hBMD and skeletal FZx of Neolithic compared to Mesolithic
219 populations may represent adaptation to the same change in environment although we do not
220 know the extent to which the change in FZx was driven by genetic or plastic developmental
221 response to environmental change. On the other hand, FZx continues to decrease between the
222 Neolithic and post-Neolithc (Fig. 5c,f)–which is not reflected in the hBMD PRS (Fig.5 a-b,d-e).
223 One possibility is that the two phenotypes responded differently to the post-Neolithic
224 intensification of agriculture. Another is that the non-genetic component of hBMD, which we
225 do not capture here, also continued to decrease.
226

227 ## Are changes in PRS driven by selection or genetic drift?

228 We tested whether there was evidence for selection on any of these traits, by computing the Q_x
229 statistic (73) for increasing numbers of SNPs from each PRS, with effect sizes taken from either
230 PRS(GWAS) (Fig. 6a-c) or PRS(GWAS/Sibs) (Fig. 6d-f). We computed the statistic between each
231 pair of adjacent time periods, and over all time periods. We estimated empirical P-values by

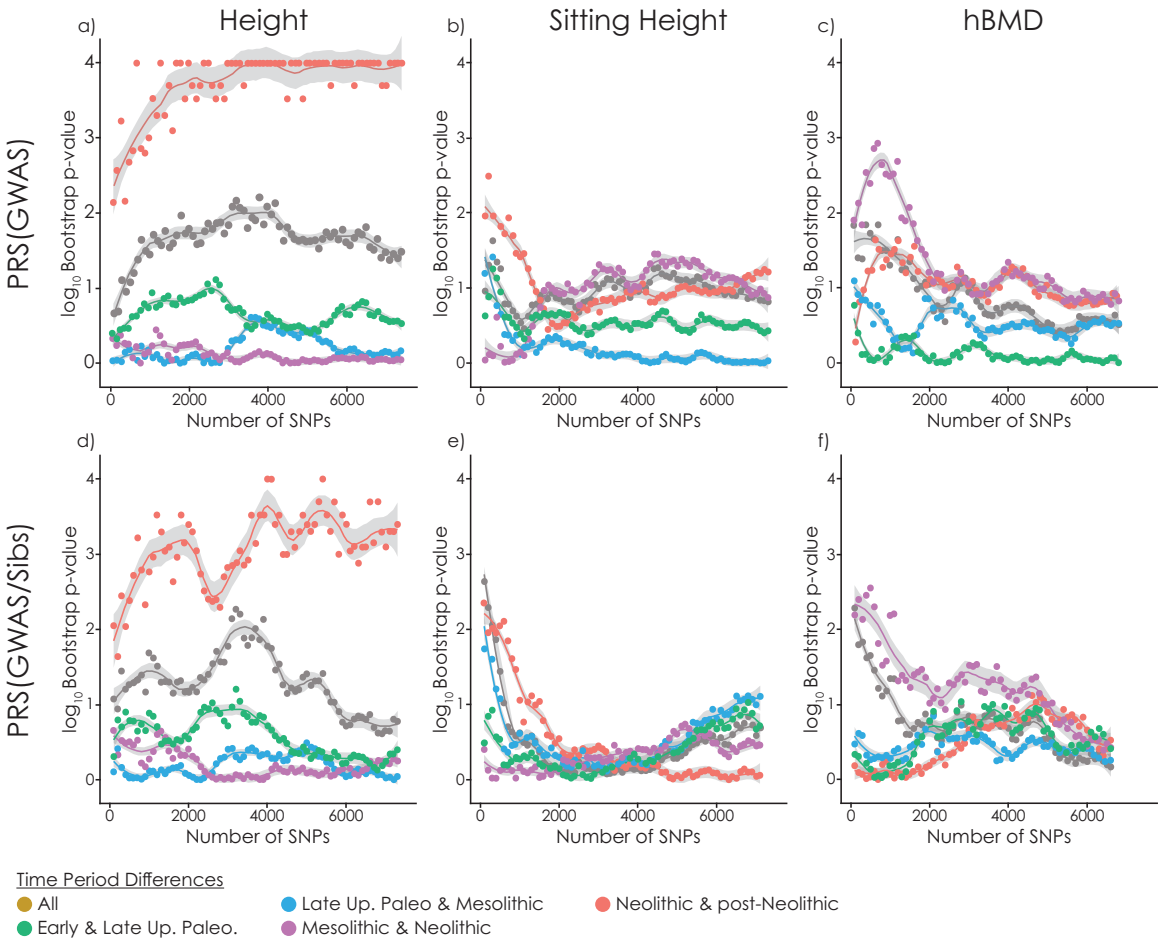


Figure 6: Signals of selection on standing height, sitting height and bone mineral density. We plot the $\log_{10}$ bootstrap P-values for the Q_x statistics (y-axis, capped at 4) for GWAS signals (top row) and GWAS/Sibs (bottom row). We tested each pair of adjacent populations, and the combination of all of them ("All"). We ordered PRS SNPs by increasing P-value and tested the significance of Q_x for increasing numbers of SNPs (x-axis).

232 sampling random frequency-matched SNPs from across the genome. Using GWAS effect sizes,
233 we find selection between the Neolithic and Post-Neolithic for stature ($P<1 \times 10^{-4}$; Fig. 6a),
234 which replicates using effect sizes estimated within siblings ($10^{-4}<P<10^{-2}$; Fig. 6d). The fact that
235 the signal is less strong for GWAS/Sibs than for GWAS could either indicate that some of the
236 signal is driven by stratification (21, 22), or that the power to detect selection for smaller effect
237 sizes is lower when using the nosier sibling effect sizes. We tested this by generating GWAS
238 results on a subsample of individuals, chosen so that the standard error of the effect sizes was
239 equal to those of the within-sibling effects. This produced similar results to the analysis using
240 the within-sibling effects (Supplementary Fig. 9), suggesting that the main reason for the
241 weaker signal is the reduction in sample size of the within-sibling analysis.
242
243 For sitting height, we find little evidence of selection in any time period ($P<10^{-2}$) We conclude
244 that there was most likely selection for increased standing but not sitting height in the Steppe
245 ancestors of Bronze Age European populations, as previously proposed (29). One potential
246 caveat is that, although we re-estimated effect sizes within siblings, we still used the GWAS
247 results to identify SNPs to include. This may introduce some subtle confounding, which remains
248 a question for future investigation. Finally, using GWAS effect sizes, we identify some evidence
249 of selection on heel BMD between when comparing Mesolithic and Neolithic populations (10^{-}
250 $^{3}<P<10^{-2}$; Fig. 6c). However, this signal is relatively weak when using within-sibling effect sizes,
251 and disappears when we include more than about 2000 SNPs.

252 Discussion

253 We showed that the well-documented temporal and geographic trends in stature in Europe
254 between the Early Upper Paleolithic and the post-Neolithic period are broadly consistent with
255 those that would be predicted by polygenic risk scores (PRS) computed using present-day
256 GWAS results combined with ancient DNA. However, because of the limited predictive power of
257 current PRS, we cannot provide a quantitative estimate of how much of the variation in
258 phenotype between populations might be explained by variation in PRS. Similarly, we cannot
259 say whether the changes were continuous, reflecting evolution through time, or discrete,

260 reflecting changes associated with known episodes of replacement or admixture of populations
261 that have diverged genetically over time. Finally, we find cases where predicted genetic
262 changes are discordant with observed phenotypic changes–emphasizing the role of
263 developmental plasticity in response to environmental change and the difficulty in interpreting
264 differences in PRS in the absence of phenotypic data.
265
266 Our results indicate two major episodes of genetic change. First, there was a reduction in
267 stature PRS–but not sitting height PRS–between the Early Upper Paleolithic and Neolithic.
268 These genetic changes are consistent with the decrease in stature–driven by leg length–
269 observed in skeletons during this time period (4, 64, 74, 75). This evolutionary change could
270 have been adaptive, driven by changes in resource availability (76) or to a colder climate (61).
271 Early Upper Paleolithic populations in Europe would have migrated relatively recently from
272 more southern latitudes and had body proportions that are typical of present-day tropical
273 populations (75). It is therefore plausible that they adapted to the colder climate of northern
274 latitudes throughout the Upper Paleolithic. Comparison between patterns of phenotypic and
275 genetic variation suggest that, on a broad scale, variation in body proportions among present-
276 day people reflects adaptation to environment largely along latitudinal gradients (77, 78). On
277 the other hand, we do not find genetic evidence for selection on stature during this time
278 period–although with a small sample size we likely have very low power to detect it. Further,
279 the populations of Early Upper Paleolithic, Late Upper Paleolithic, Mesolithic and Neolithic
280 Europe are substantially discontinuous and deeply diverged genetically (33, 59). For example
281 the ancestors of Mesolithic and Neolithic Europeans are estimated to have diverged ~46,000 BP
282 (40). Therefore, if these genetic changes do reflect adaptation to climate, this adaptation must
283 have occurred at least partly independently in the ancestors of these populations.
284
285 The second episode of genetic change is either between the Neolithic and post-Neolithic, or
286 during the post-Neolithic period. In genome-wide ancestry, this transition is characterized by
287 the eastward movement of substantial amounts of “Steppe ancestry” into Central and Western
288 Europe (27, 30, 38, 50). Our results are thus consistent with previous results that Bronze Age

289 populations of the Eurasian steppe had been selected for increased height and that migration
290 and admixture of these populations with Neolithic European populations increased genetic
291 height in Europe (29, 30). There is no obvious climatic driver for this adaptation but one
292 possibility is that it represents adaptation to a change in social environment. Y chromosome
293 phylogenies suggests an increase in male reproductive variance at this time (29, 48, 50, 79, 80).
294 Culturally, the Bronze Age is characterized by increased social stratification (81) and the
295 introduction of patriarchal Indo-European culture (82). Perhaps these social changes implied
296 increased competition for resources and consequent selection for greater body size. The
297 geographic gradient of increasing skeletal stature is unclear in the Paleolithic, largely West-East
298 in the Mesolithic (7, 64) and largely South-North by the Bronze Age (4, 7, 9). Latitudinal, but not
299 longitudinal, patterns are qualitatively consistent with geographic patterns in PRS suggesting
300 that, like temporal variation, both genetics and environment contribute to geographic variation.
301
302 There is a major confounding factor in analysis of temporal and geographic variation in PRS,
303 particularly in the Bronze Age. Genetic population structure in present-day Europe is correlated
304 with geography (83) and largely driven by variation in proportions of Steppe ancestry, with
305 more Steppe ancestry in Northern Europe and less in Southern Europe (38). Suppose that
306 environmental variation in stature is also correlated with geography, and that Northern
307 Europeans are taller than Southern Europeans for entirely non-genetic reasons. Then, GWAS
308 that do not completely correct for stratification will find that genetic variants that are more
309 common in Steppe populations than Neolithic populations are associated with increased height.
310 When these GWAS results are then used to compute PRS for ancient populations, they will
311 predict that Steppe ancestry populations were genetically taller simply because they are more
312 closely related to present-day Northern Europeans (21, 22). In this study, we attempted to
313 avoid this confounding in two ways: first, by computing PRS using GWAS effect sizes from the
314 UK Biobank–a fairly homogenous dataset that should be well-controlled for population
315 stratification, and second, by replicating our results after re-estimating the effect sizes within
316 siblings, which should be robust to population stratification. The tradeoff between these two
317 methods is that the small sibling sample size means that effect size estimates are noisy, even

318 though they should be unbiased, and our results using sibling-estimated effects may miss subtle
319 trends. However, we cannot exclude the possibility that some confounding remains, for
320 example because although we re-estimated effect sizes using the within-siblings design, we still
321 ascertained loci using the GWAS results. Residual confounding would also tend to create
322 spurious signals of polygenic adaption (21, 22).
323
324 As well as genetic contributions to phenotype, our results shed light on possible environmental
325 contributions. In some cases, we can make hypotheses about the relationship between
326 environmental or lifestyle changes, and genetic change. For example, if we interpret change in
327 femur bending strength as reflecting a decrease in mobility, the coincident Mesolithic/Neolithic
328 change in heel bone mineral density PRS can be seen as a genetic response to this change.
329 However, in the Neolithic/post-Neolithic periods, the two observations are decoupled. This
330 emphasizes the role of developmental plasticity in response to changes in environment, and of
331 joint interpretation of phenotypic and genetic variables. Even when looking at the same
332 phenotype, we find cases where genetic predictions and phenotypic data are discordant–for
333 example in post-Neolithic sitting height. We must therefore be cautious in the interpretation of
334 predicted genetic patterns where phenotypes cannot be directly measured, even if it is possible
335 to control stratification. Predicted genetic changes should be used as a baseline, against which
336 non-genetic effects can be measured and tested.

Methods

Ancient DNA and polygenic risk score construction

We collected published ancient DNA data from 1122 ancient individuals, taken from 29 publications. The majority of these individuals had been genotyped using an in-solution capture reagent ("1240k") that targets 1.24 million single nucleotide polymorphisms (SNPs) across the genome. Because of the low coverage of most of these samples, the genotype data are pseudo-haploid. That is, there is only a single allele present for each individual at each site, but alleles at adjacent sites may come from either of the two chromosomes of the individual. For individuals with shotgun sequence data, we selected a single read at each 1240k site. We obtained the date of each individual from the original publication. Most of the samples have been directly radiocarbon dated, or else are securely dated by context.

We obtained GWAS results from the Neale lab UK Biobank page (http://www.nealelab.is/uk-biobank/; Round 1, accessed February and April 2018). To compute PRS, we first took the intersection of the 1240k sites and the association summary statistics. We then selected a list of SNPs to use in the PRS by selecting the SNP with the lowest P-value, removing all SNPs within 250kb, and repeating until there were no SNPs remaining with P-value less than 10^{-6}. We then computed PRS for each individual by taking the sum of genotype multiplied by effect size for all included SNPs. Where an individual was missing data at a particular SNP, we replaced the SNP with the average frequency of the SNP across the whole dataset. This has the effect of shrinking the PRS towards the mean and should be conservative for the identification of differences in PRS. We confirmed that there was no correlation between missingness and PRS, to make sure that missing data did not bias the results (correlation between missingness and PRS ρ=0.02; P=0.44, Supplementary Fig. 10). Finally, we normalized the PRS across individuals to have mean 0 and standard deviation 1.

We estimated within-family effect sizes from 17,358 sibling pairs in the UK Biobank to obtain effect estimates that are unaffected by stratification. Pairs of individuals were identified as siblings if estimates of IBS0 were greater than 0.0018 and kinship coefficients were greater than

366 0.185. Of those pairs, we only retained those where both siblings were classified by UK Biobank
367 as "white British", and randomly picked two individuals from families with more than two
368 siblings. We used Hail (84) to estimate within-sibling pair effect sizes for 1,284,881 SNPs by
369 regressing pairwise phenotypic differences between siblings against the difference in genotype.
370 We included pairwise differences of sex (coded as 0/1) and age as covariates, and inverse-rank-
371 normalized the phenotype before taking the differences between siblings. To combine the
372 GWAS and sibling results, we first restricted the GWAS results to sites where we had estimated
373 a sibling effect size and replaced the GWAS effect sizes by the sibling effects. We then restricted
374 to 1240k sites and constructed PRS in the same way as for the GWAS results.
375
376 To test whether the differences in the GWAS and GWAS/Sibs PRS results can be explained by
377 differences in power, we created subsampled GWAS estimates which matched the sibling in the
378 expected standard errors, by determining the equivalent sample size necessary and randomly
379 sampling N_{sub} individuals. $N_{sub} = \frac{N_{sib}}{2\,var(\delta_{sib})}$ where δ_{sib} is the difference in normalized
380 phenotype between siblings after accounting for the covariates age and sex.
381

382 ## Stature data

383 We obtained stature data from Ruff (2018) (4) (data file and notes available at
384 http://www.hopkinsmedicine.org/fae/CBR.html), which also includes estimated body mass,
385 femoral midshaft anteroposterior strength (FZx), and other osteometric dimensions. Statures
386 and body masses were calculated from linear skeletal measurements using anatomical
387 reconstruction or sample-specific regression formulae (4, 58). We calculated sitting height as
388 basion-bregma (cranial) height (BBH) plus vertebral column length (VCL). We restricted analysis
389 to 1159 individuals dated earlier than 1165 BP (651 males and 508 females), of which 1130 had
390 estimates for stature, 1014 for FZx and 236 for sitting height. Sitting and standing height were
391 standardized for sex by adding the mean difference between male and female estimates to all
392 the female values. Sex differences in stature remain relatively constant over time (4), making it
393 reasonable to adjust all female heights by the same mean value. For FZx we first standardized

394 for sex as we did for stature then divided each by estimated body mass multiplied by
395 biomechanical femur length (4).
396

397 ## Grouping

398 We grouped individuals into broad categories based on date and, in some cases, archeological
399 and genetic context. All individuals were assigned to one time period group, based on median
400 age estimates of the sample obtained from the original publications. Date ranges for each time
401 period are based on a combination of historical, climatic, and archaeological factors. The Early
402 Upper Paleolithic comprises all samples older than 25,000 BP, which roughly coincides with the
403 end of the last glacial maximum (LGM). The Late Upper Paleolithic begins when the European
404 glaciers are beginning to recede (25,000 BP) and extends until 11,000 BP and a shift in lithic
405 technology that is traditionally used to delineate the beginning of the Mesolithic period.
406 Transitions between the Mesolithic, Neolithic, and Bronze Age are staggered throughout
407 Europe, so creating universally applicable date ranges is not possible. We instead defined
408 overlapping transition periods between the Mesolithic and Neolithic periods (8500-5500 BP)
409 and between the Neolithic and post-Neolithic (5000-3900 BP). For the genetic data, samples in
410 the overlapping periods were assigned based on genetic population affiliation, inferred using
411 supervised ADMIXTURE (48, 85, 86) which, in most of Western Europe, corresponds closely to
412 archaeological context (38, 48). In particular, the Mesolithic/Neolithic overlap was resolved
413 based on whether each individual had more (Neolithic) or less (Mesolithic) than 50% ancestry
414 related to northwest Anatolian Neolithic Farmers. The Neolithic/post-Neolithic overlap was
415 resolved based on whether individuals had more than 25% ancestry related to Bronze Age
416 Steppe populations ("Steppe ancestry"; See Ref. (86) for more details). For the skeletal data,
417 group assignment in the overlapping periods was determined by the archaeology of each site.
418 Broadly, sites belonging to the Neolithic have transitioned to agricultural subsistence. Similarly,
419 post-Neolithic populations are broadly defined by evidence of metal working (Copper, Bronze
420 and Iron Ages, and later periods). In particular, we included Late Eneolithic (Copper Age) sites
421 associated with Corded Ware and Bell Beaker material culture in the post-Neolithic category
422 but for consistency with the genetic classifications, we included 8 Early Eneolithic (before 4500

BP) individuals in the Neolithic category, since this precedes the appearance of Steppe ancestry in Western Europe. We excluded samples more recent than 1165 BP.

Linear models

We fitted a series of linear models to changes in both PRS and stature data with time. In the most general model, we allow both the intercept and slope to vary between groups. We then either force some of the slopes to be zero, or some of the adjacent groups to have identical parameters. We describe the models using underscores to indicate changes in parameters, lowercase to indicate slopes (change with respect to time) fixed to zero, and upper case to indicate free slopes (i.e. linear trends with time). For example, "E_L_M_N_B" is the most general model, "elmnb" indicates that all groups have the same mean and there is no change with time, and "ELMN_B" indicates that the first four groups share the same parameters, and the post-Neolithic has different parameters. The models shown in Figures 1 and 2 are "e_lmn_b" (panels a-b), "e_lm_nb" (panel c), "ELMN_B" (panels d-e) and "ELM_NB" (panel f). To analyze geographic variation, we used the residuals of the "ELMN_B" model for the PRS and "ELM_NB" for skeletal stature, and fitted regressions against latitude and longitude.

Polygenic selection test

We computed bootstap P-values for the Q_x statistic (73) by sampling random sets of SNPs in matched 5% frequency bins, and re-computing the statistic. Unlike for the PRS calculations, we ignored missing data, since the Q_x statistic uses only the population-level estimated allele frequencies and not individual-level data. We tested a series of nested sets of SNPs (x-axis in Fig. 6), adding SNPs in 100 SNP batches, ordered by increasing P-value, down to a P-value of 0.1.

Acknowledgments

I.M. was supported by a Research Fellowship from the Alfred P. Sloan foundation, and a New Investigator Research Grant from the Charles E. Kaufman Foundation. Skeletal data were collected in collaboration with Brigitte Holt, Markku Niskanen, Vladimir Sladék, and Margit Bernor, with the support of the National Science Foundation (BCS-0642297 and BSC-0642710)

and the Grant Agency of the Czech Republic and the Academy of Finland and Finnish Cultural Foundation. We thank Jeremy Berg and Eva Rosenstock for helpful comments on an earlier version of the manuscript. This project was initially conceived during discussions at the workshop “Human stature in the Near East and Europe in a long-term perspective" at the Freie Universität Berlin 25-27 April 2018, organized as part of the Emmy-Noether-Projekt "LiVES" funded by the German Research Foundation, Grant Nr. RO4148/1 (PI Eva Rosenstock).

References

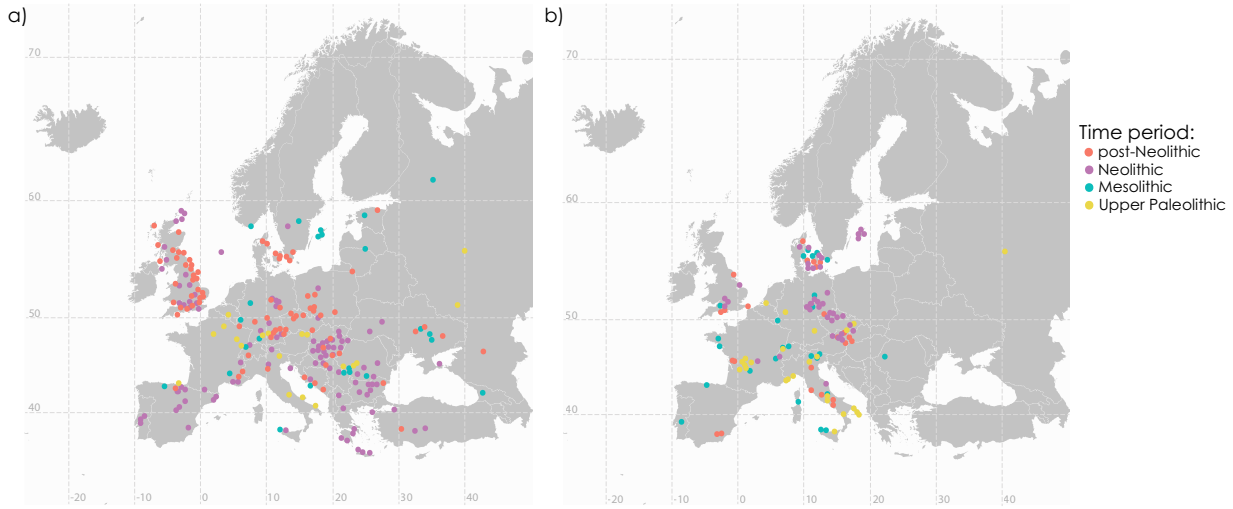
1. NCD Risk Factor Collaboration (2016) A century of trends in adult human height. *Elife* 5.
2. Hatton TJ & Bray BE (2010) Long run trends in the heights of European men, 19th-20th centuries. *Econ Hum Biol* 8(3):405-413.
3. Arcaleni E (2006) Secular trend and regional differences in the stature of Italians, 1854–1980. *Economics & Human Biology* 4(1):24-38.
4. Ruff CB ed (2018) *Skeletal Variation and Adaptation in Europeans: Upper Paleolithic to the Twentieth Century* (John Wiley and Sons, Hoboken).
5. Benazzi S*, et al.* (2011) Early dispersal of modern humans in Europe and implications for Neanderthal behaviour. *Nature* 479(7374):525-528.
6. Higham T*, et al.* (2011) The earliest evidence for anatomically modern humans in northwestern Europe. *Nature* 479(7374):521-524.
7. Rosenstock E*, et al.* (2019) Human Stature in the Near East and Europe ca. 10 000 – 1000 BC: its spatio-temporal development in a Bayesian errors-in-variables model. *Anthropological and Archaeological Sciences* In press.
8. Perkins JM, Subramanian SV, Davey Smith G, & Ozaltin E (2016) Adult height, nutrition, and population health. *Nutr Rev* 74(3):149-165.
9. Rosenstock E, Groß M, Hujić A, & Scheibner A (2015) Back to good shape: biological standard of living in the Copper and Bronze ages and the possible role of food. *The third food revolution? Setting the bronze age table: common trends in economic and subsistence strategies in bronze age europe*, eds Kneisel J, Dal Corso M, Kirleis W, Scholz H, Taylor N, & Tiedtke V (Dr. Rudolph Habelt GmbH, Bonn), pp 121-152.
10. Galton F (1886) Regression Towards Mediocrity in Hereditary Stature. *The Journal of the Anthropological Institute of Great Britain and Ireland* 15:246-263.
11. Lettre G (2011) Recent progress in the study of the genetics of height. *Hum Genet* 129(5):465-472.
12. Fisher RA (1918) The correlation between relatives on the supposition of Mendelian inheritance. *Trans Roy Soc Edin.* 52:399-433.

489 13. Silventoinen K*, et al.* (2003) Heritability of adult body height: a comparative study of
490 twin cohorts in eight countries. *Twin Res* 6(5):399-408.
491 14. Visscher PM, Hill WG, & Wray NR (2008) Heritability in the genomics era--concepts and
492 misconceptions. *Nat Rev Genet* 9(4):255-266.
493 15. Wood AR*, et al.* (2014) Defining the role of common variation in the genomic and
494 biological architecture of adult human height. *Nat Genet* 46(11):1173-1186.
495 16. Yengo L*, et al.* (2018) Meta-analysis of genome-wide association studies for height and
496 body mass index in approximately 700000 individuals of European ancestry. *Hum Mol*
497 *Genet* 27(20):3641-3649.
498 17. Lango Allen H*, et al.* (2010) Hundreds of variants clustered in genomic loci and biological
499 pathways affect human height. *Nature* 467(7317):832-838.
500 18. Marouli E*, et al.* (2017) Rare and low-frequency coding variants alter human adult
501 height. *Nature* 542(7640):186-190.
502 19. Martin AR*, et al.* (2019) Clinical use of current polygenic risk scores may exacerbate
503 health disparities. *Nat Genet* 51(4):584-591.
504 20. Martin AR*, et al.* (2017) Human Demographic History Impacts Genetic Risk Prediction
505 across Diverse Populations. *Am J Hum Genet* 100(4):635-649.
506 21. Berg JJ*, et al.* (2018) Reduced signal for polygenic adaptation of height in UK Biobank.
507 *bioRxiv* https://doi.org/10.1101/354951.
508 22. Sohail M*, et al.* (2018) Signals of polygenic adaptation on height have been
509 overestimated due to uncorrected population structure in genome-wide association
510 studies. *bioRxiv* https://doi.org/10.1101/355057.
511 23. Skoglund P & Mathieson I (2018) Ancient Genomics of Modern Humans: The First
512 Decade. *Annu Rev Genomics Hum Genet* 19:381-404.
513 24. Marciniak S & Perry GH (2017) Harnessing ancient genomes to study the history of
514 human adaptation. *Nat Rev Genet* 18(11):659-674.
515 25. Slatkin M & Racimo F (2016) Ancient DNA and human history. *Proc Natl Acad Sci U S A*
516 113(23):6380-6387.
517 26. Wilde S*, et al.* (2014) Direct evidence for positive selection of skin, hair, and eye
518 pigmentation in Europeans during the last 5,000 y. *Proceedings of the National Academy*
519 *of Sciences* 111(13):4832-4837.
520 27. Allentoft ME*, et al.* (2015) Population genomics of Bronze Age Eurasia. *Nature*
521 522(7555):167-172.
522 28. Burger J, Kirchner M, Bramanti B, Haak W, & Thomas MG (2007) Absence of the lactase-
523 persistence-associated allele in early Neolithic Europeans. *Proc Natl Acad Sci U S A*
524 104(10):3736-3741.
525 29. Mathieson I*, et al.* (2015) Genome-wide patterns of selection in 230 ancient Eurasians.
526 *Nature* 528(7583):499-503.
527 30. Martiniano R*, et al.* (2017) The population genomics of archaeological transition in west
528 Iberia: Investigation of ancient substructure using imputation and haplotype-based
529 methods. *PLoS Genet* 13(7):e1006852.
530 31. Berg JJ, Zhang X, & Coop G (2017) Polygenic Adaptation has Impacted Multiple
531 Anthropometric Traits. *biorXiv*:https://doi.org/10.1101/167551.

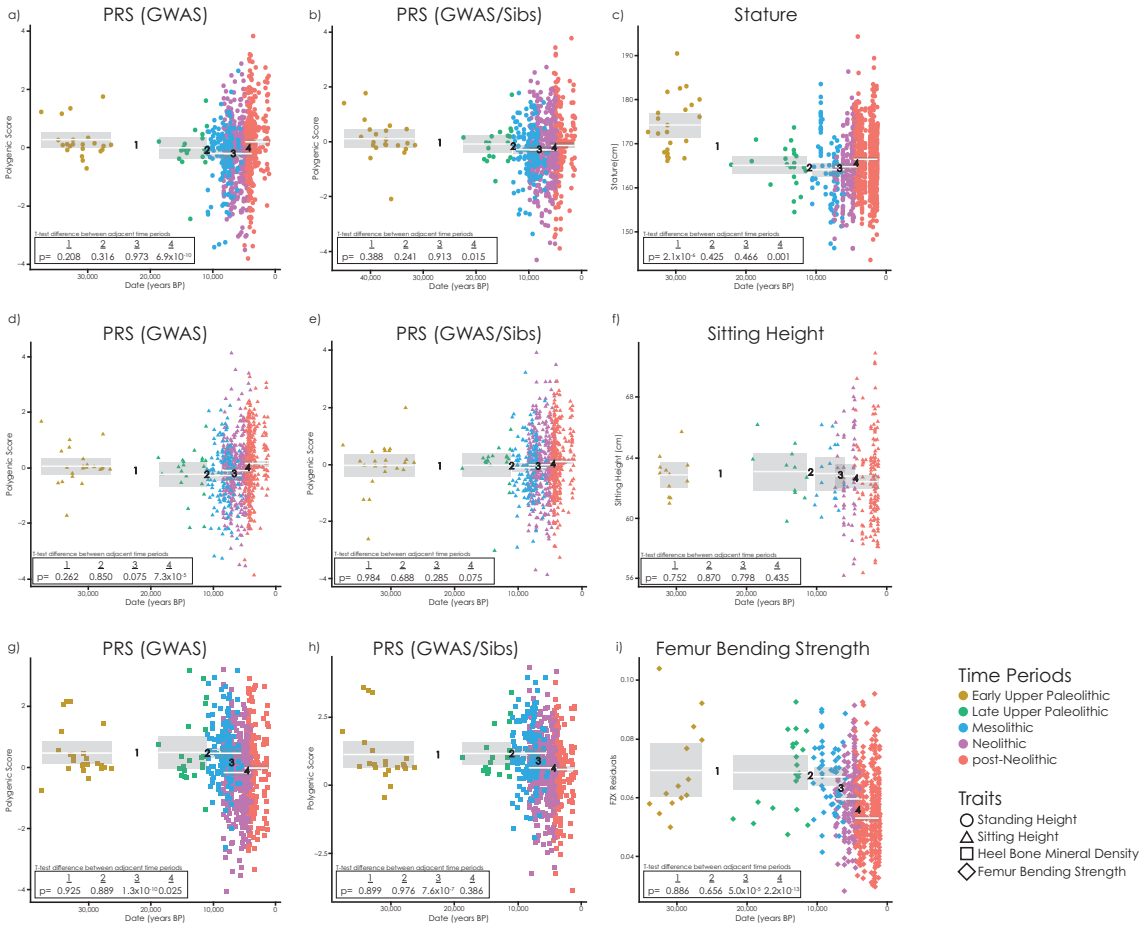
532 32. Cassidy LM, *et al.* (2016) Neolithic and Bronze Age migration to Ireland and
533 establishment of the insular Atlantic genome. *Proceedings of the National Academy of*
534 *Sciences* 113(2):368-373.
535 33. Fu Q, *et al.* (2016) The genetic history of Ice Age Europe. *Nature* 534(7606):200-205.
536 34. Gamba C, *et al.* (2014) Genome flux and stasis in a five millennium transect of European
537 prehistory. *Nat Commun* 5.
538 35. González-Fortes G, *et al.* (2017) Paleogenomic Evidence for Multi-generational Mixing
539 between Neolithic Farmers and Mesolithic Hunter-Gatherers in the Lower Danube
540 Basin. *Current Biology*.
541 36. Gunther T, *et al.* (2018) Population genomics of Mesolithic Scandinavia: Investigating
542 early postglacial migration routes and high-latitude adaptation. *PLoS Biol*
543 16(1):e2003703.
544 37. Gunther T, *et al.* (2015) Ancient genomes link early farmers from Atapuerca in Spain to
545 modern-day Basques. *Proc Natl Acad Sci U S A* 112(38):11917-11922.
546 38. Haak W, *et al.* (2015) Massive migration from the steppe is a source for Indo-European
547 languages in Europe. *Nature* 522(7555):207-211.
548 39. Hofmanová Z, *et al.* (2016) Early farmers from across Europe directly descended from
549 Neolithic Aegeans. *Proceedings of the National Academy of Sciences* 113(25):6886-6891.
550 40. Jones ER, *et al.* (2015) Upper Palaeolithic genomes reveal deep roots of modern
551 Eurasians. *Nature communications* 6.
552 41. Jones ER, *et al.* (2017) The Neolithic Transition in the Baltic Was Not Driven by
553 Admixture with Early European Farmers. *Curr Biol* 27(4):576-582.
554 42. Keller A, *et al.* (2012) New insights into the Tyrolean Iceman's origin and phenotype as
555 inferred by whole-genome sequencing. *Nat Commun* 3:698.
556 43. Kılınç Gülşah M, *et al.* (2016) The Demographic Development of the First Farmers in
557 Anatolia. *Current Biology* 26(19):2659-2666.
558 44. Lazaridis I, *et al.* (2017) Genetic origins of the Minoans and Mycenaeans. *Nature*
559 548(7666):214-218.
560 45. Lazaridis I, *et al.* (2014) Ancient human genomes suggest three ancestral populations for
561 present-day Europeans. *Nature* 513(7518):409-413.
562 46. Lipson M, *et al.* (2017) Parallel ancient genomic transects reveal complex population
563 history of early European farmers. *bioRxiv*:114488.
564 47. Martiniano R, *et al.* (2016) Genomic signals of migration and continuity in Britain before
565 the Anglo-Saxons. *Nature communications* 7:10326.
566 48. Mathieson I, *et al.* (2018) The genomic history of southeastern Europe. *Nature*
567 555(7695):197-203.
568 49. Olalde I, *et al.* (2014) Derived immune and ancestral pigmentation alleles in a 7,000-
569 year-old Mesolithic European. *Nature* advance online publication.
570 50. Olalde I, *et al.* (2018) The Beaker phenomenon and the genomic transformation of
571 northwest Europe. *Nature* 555(7695):190-196.
572 51. Olalde I, *et al.* (2015) A Common Genetic Origin for Early Farmers from Mediterranean
573 Cardial and Central European LBK Cultures. *Molecular Biology and Evolution*
574 32(12):3132-3142.

575 52. Omrak A*, et al.* (2016) Genomic Evidence Establishes Anatolia as the Source of the
576 European Neolithic Gene Pool. *Curr Biol* 26(2):270-275.
577 53. Saag L*, et al.* (2017) Extensive Farming in Estonia Started through a Sex-Biased Migration
578 from the Steppe. *Curr Biol* 27(14):2185-2193 e2186.
579 54. Schiffels S*, et al.* (2016) Iron age and Anglo-Saxon genomes from East England reveal
580 British migration history. *Nature communications* 7:10408.
581 55. Sikora M*, et al.* (2017) Ancient genomes show social and reproductive behavior of early
582 Upper Paleolithic foragers. *Science* 358(6363):659-662.
583 56. Skoglund P*, et al.* (2014) Genomic Diversity and Admixture Differs for Stone-Age
584 Scandinavian Foragers and Farmers. *Science* 344(6185):747-750.
585 57. Veeramah KR*, et al.* (2018) Population genomic analysis of elongated skulls reveals
586 extensive female-biased immigration in Early Medieval Bavaria. *Proc Natl Acad Sci U S A*
587 115(13):3494-3499.
588 58. Ruff CB*, et al.* (2012) Stature and body mass estimation from skeletal remains in the
589 European Holocene. *Am J Phys Anthropol* 148(4):601-617.
590 59. Skoglund P*, et al.* (2012) Origins and genetic legacy of Neolithic farmers and hunter-
591 gatherers in Europe. *Science* 336(6080):466-469.
592 60. Chan Y*, et al.* (2015) Genome-wide Analysis of Body Proportion Classifies Height-
593 Associated Variants by Mechanism of Action and Implicates Genes Important for
594 Skeletal Development. *Am J Hum Genet* 96(5):695-708.
595 61. Ruff CB (2002) Variation in Human Body Size and Shape. *Annual Review of Anthropology*
596 31:211-232.
597 62. Allen JA (1887) The influence of physical conditions in the genesis of species. *Radical*
598 *Review* 1:108-140.
599 63. Bergman C (1847) Über die Verhältnisse der Wärmeökonomie der Thiere zu ihrer
600 Grösse. *Göttinger Studien* 3:595-708.
601 64. Formicola V & Giannecchini M (1999) Evolutionary trends of stature in upper Paleolithic
602 and Mesolithic Europe. *J Hum Evol* 36(3):319-333.
603 65. Chirchir H*, et al.* (2015) Recent origin of low trabecular bone density in modern humans.
604 *Proc Natl Acad Sci U S A* 112(2):366-371.
605 66. Chirchir H, Ruff CB, Junno JA, & Potts R (2017) Low trabecular bone density in recent
606 sedentary modern humans. *Am J Phys Anthropol* 162(3):550-560.
607 67. Pettersson U, Nilsson M, Sundh V, Mellstrom D, & Lorentzon M (2010) Physical activity is
608 the strongest predictor of calcaneal peak bone mass in young Swedish men. *Osteoporos*
609 *Int* 21(3):447-455.
610 68. Bauer DC*, et al.* (2007) Quantitative ultrasound predicts hip and non-spine fracture in
611 men: the MrOS study. *Osteoporos Int* 18(6):771-777.
612 69. Bauer DC*, et al.* (1997) Broadband ultrasound attenuation predicts fractures strongly
613 and independently of densitometry in older women. A prospective study. Study of
614 Osteoporotic Fractures Research Group. *Arch Intern Med* 157(6):629-634.
615 70. Kneissel M*, et al.* (1994) Age- and sex-dependent cancellous bone changes in a 4000y BP
616 population. *Bone* 15(5):539-545.

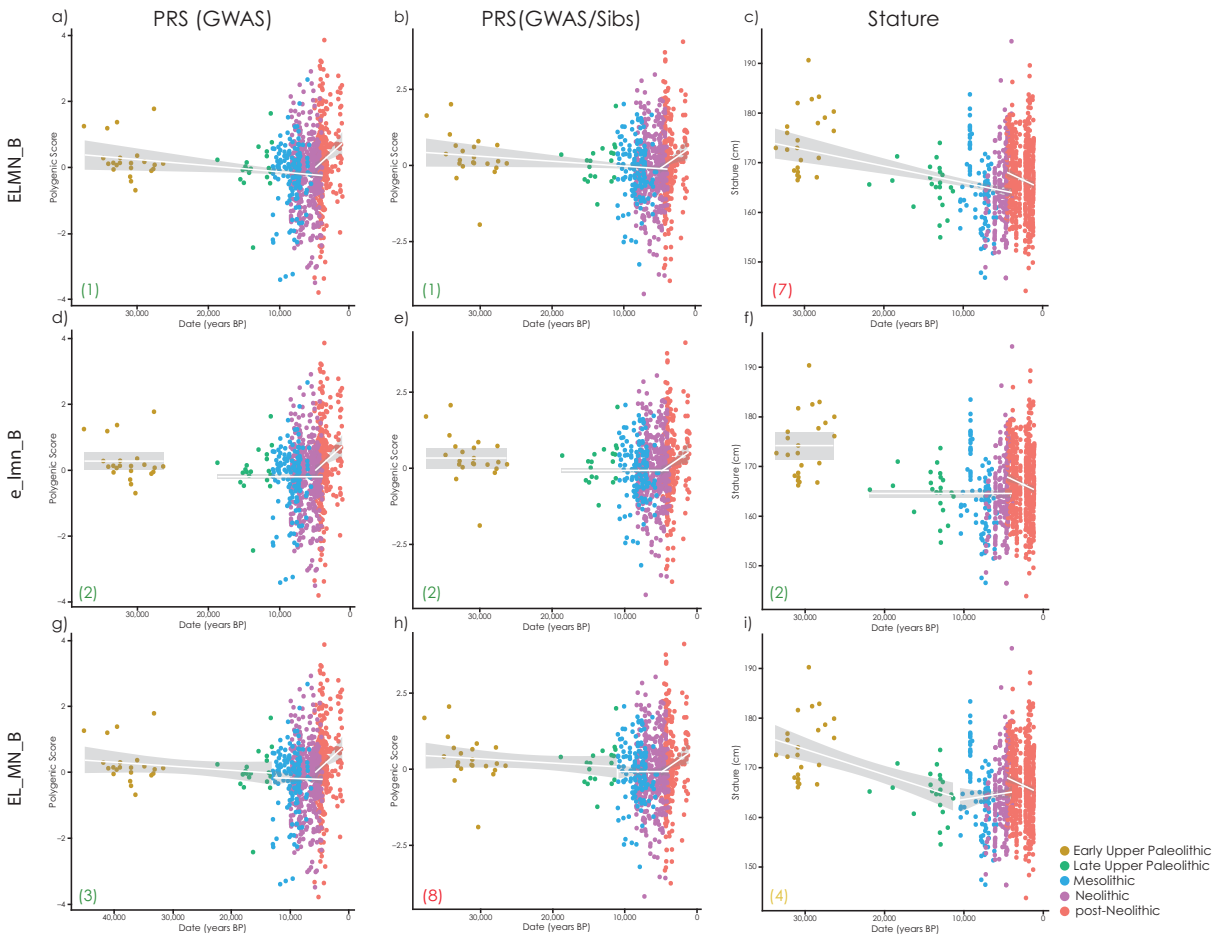
71. Ruff CB (2019) Biomechanical analyses of archaeological human skeletal samples. *Biological Anthropology of the Human Skeleton*, eds Katzenburg MA & Grauer AL (John Wiley and Sons, New York), 3rd Ed, pp 189-224.
72. Ruff CB*, et al.* (2015) Gradual decline in mobility with the adoption of food production in Europe. *Proc Natl Acad Sci U S A* 112(23):7147-7152.
73. Berg JJ & Coop G (2014) A population genetic signal of polygenic adaptation. *PLoS Genet* 10(8):e1004412.
74. Ruff CB, Trinkaus E, & Holliday TW (1997) Body mass and encephalization in Pleistocene Homo. *Nature* 387(6629):173-176.
75. Holliday TW (1997) Body proportions in Late Pleistocene Europe and modern human origins. *J Hum Evol* 32(5):423-448.
76. Formicola V & Holt BM (2007) Resource availability and stature decrease in Upper Palaeolithic Europe. *Journal of Anthropological Sciences* 85:147-155.
77. Roseman CC & Auerbach BM (2015) Ecogeography, genetics, and the evolution of human body form. *J Hum Evol* 78:80-90.
78. Savell KR, Auerbach BM, & Roseman CC (2016) Constraint, natural selection, and the evolution of human body form. *Proc Natl Acad Sci U S A* 113(34):9492-9497.
79. Poznik GD*, et al.* (2016) Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. *Nat Genet* 48(6):593-599.
80. Karmin M*, et al.* (2015) A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Res* 25(4):459-466.
81. Gilman A (1981) The Development of Social Stratification in Bronze Age Europe. *Current Anthropology* 22:1-23.
82. Mallory J & Adams DQ eds (1997) *Encyclopedia of Indo-European Culture* (Routledge, London).
83. Novembre J*, et al.* (2008) Genes mirror geography within Europe. *Nature* 456(7219):274-274.
84. Hail Team (2019) Hail. pp https://github.com/hail-is/hail/releases/tag/0.2.13.
85. Alexander DH, Novembre J, & Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19(9):1655-1664.
86. Mathieson S & Mathieson I (2018) FADS1 and the Timing of Human Adaptation to Agriculture. *Mol Biol Evol* 35(12):2957-2970.



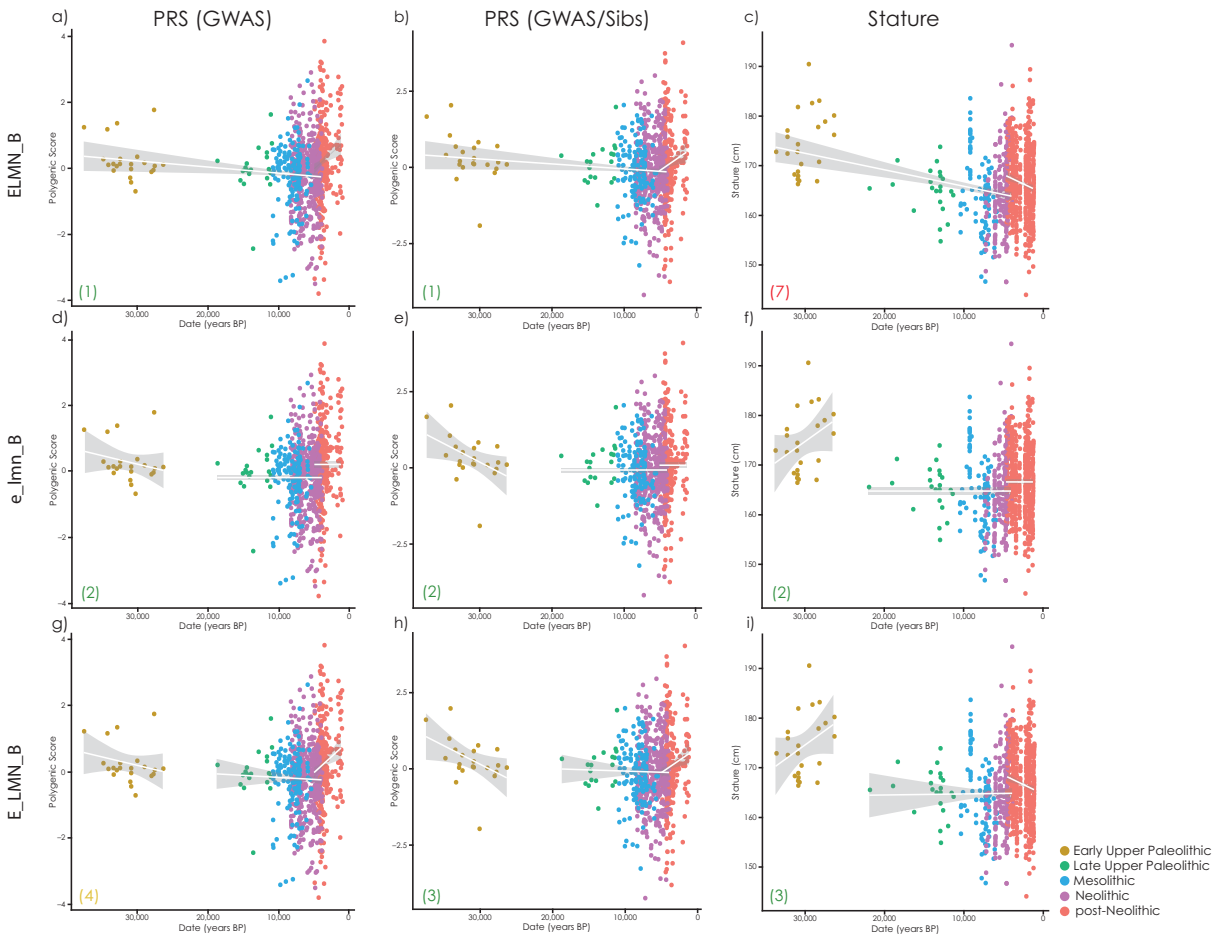
Supplementary Figure 1: Locations of samples, colored by time period. **a**) ancient DNA samples; **b**) skeletal samples.



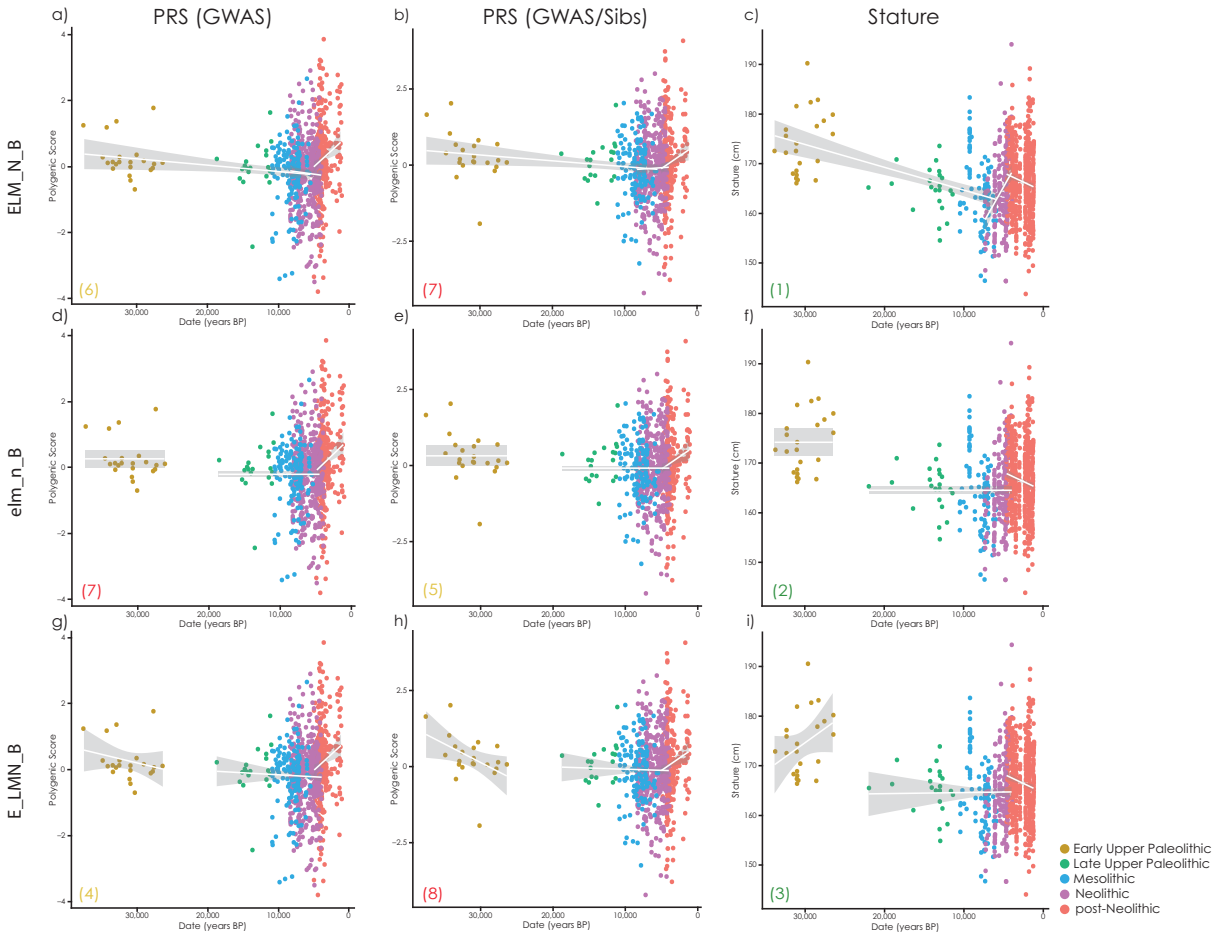
Supplementary Figure 2: Changes in PRS and skeletal phenotypes through time, with constant values in each time period. Each point is an ancient individual, lines show fitted values, grey area is the 95% confidence interval, and boxes show p-values for difference in means between adjacent groups. **a**) Standing height PRS(GWAS); **b**) Standing height PRS(GWAS/Sibs); **c**) Stature (skeletal); **d**) Sitting height PRS(GWAS); **e**) Sitting height PRS(GWAS/Sibs); **f**) Sitting height (skeletal); **g**) Heel bone mineral density PRS(GWAS); **h**) Heel bone mineral density PRS(GWAS/Sibs); **i**) Femur bending strength (skeletal).



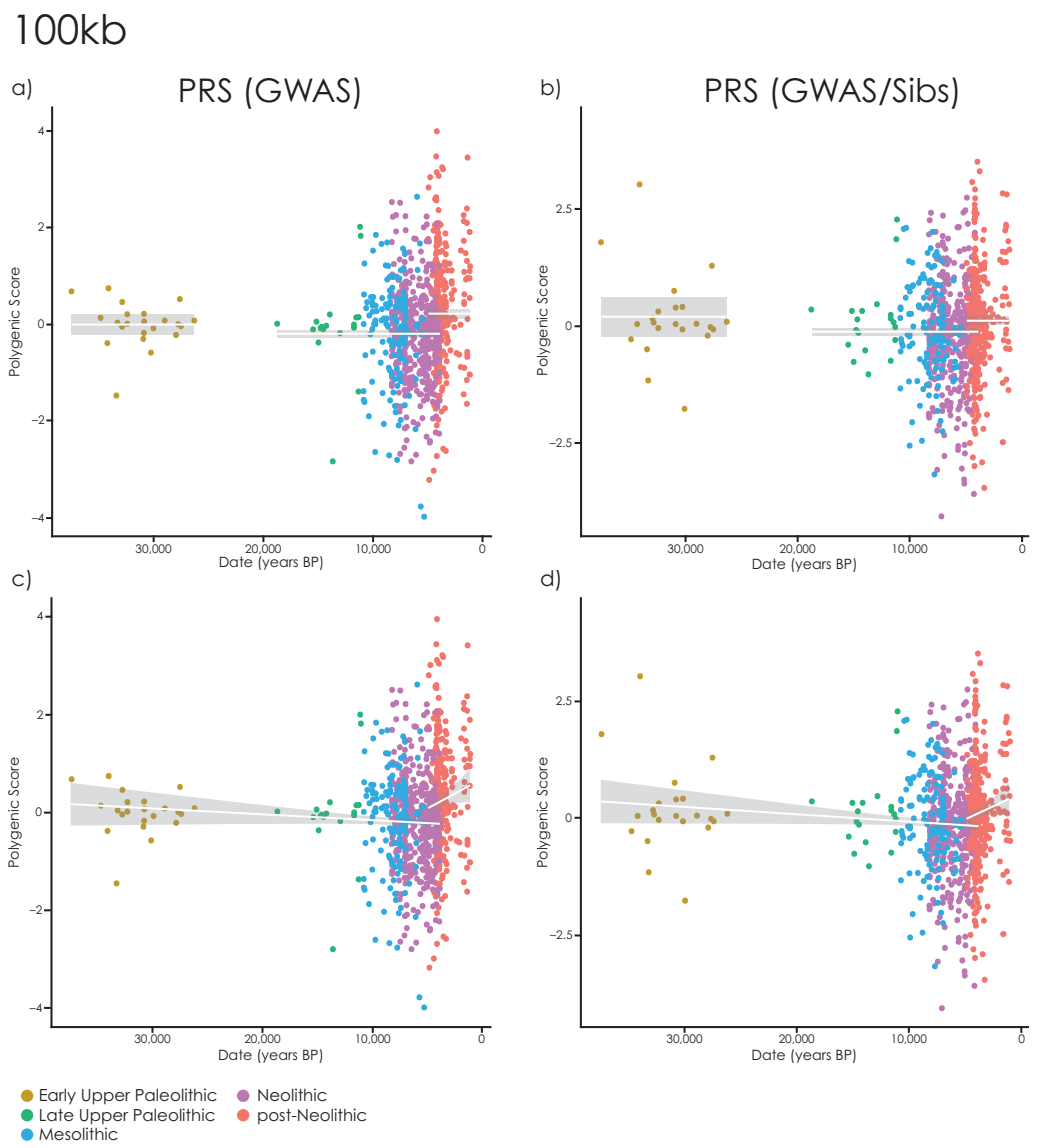
Supplementary Figure 3: Top three AIC models for PRS(GWAS), and the corresponding models for PRS(GWAS/Sibs) and skeletal stature. Row name indicates the model being tested, lowercase letters use fixed values for that time period, uppercase letters indicate the values were allowed to vary linearly with time. Number in the lower left corner of each plot indicates its place in the AIC ranking for PRS(GWAS), PRS(GWAS/Sibs) and Stature, green color indicates a good fit (rank 1-3/29 models), yellow a medium fit (rank 4-6/29 models), and red a poor fit (rank 7 or lower out of 29 models).



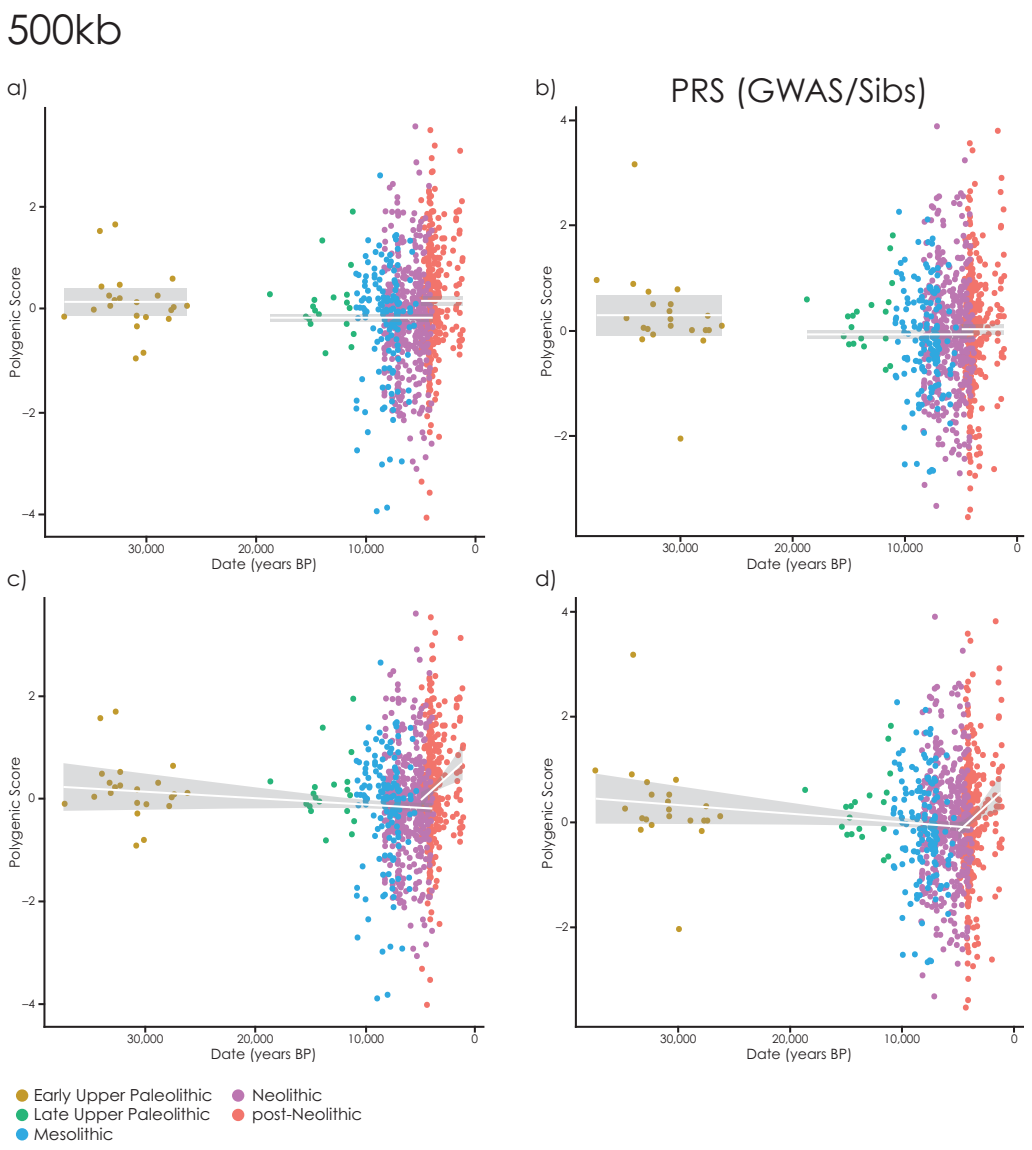
Supplementary Figure 4: Top three AIC models for PRS(GWAS/Sibs), and the corresponding models for PRS(GWAS) and skeletal stature. Row name indicates the model being tested, lowercase letters use fixed values for that time period, uppercase letters indicate the values were allowed to vary linearly with time. Number in the lower left corner of each plot indicates its place in the AIC ranking for PRS(GWAS), PRS(GWAS/Sibs) and Stature, green color indicates a good fit (rank 1-3/29 models), yellow a medium fit (rank 4-6/29 models), and red a poor fit (rank 7 or lower out of 29 models).



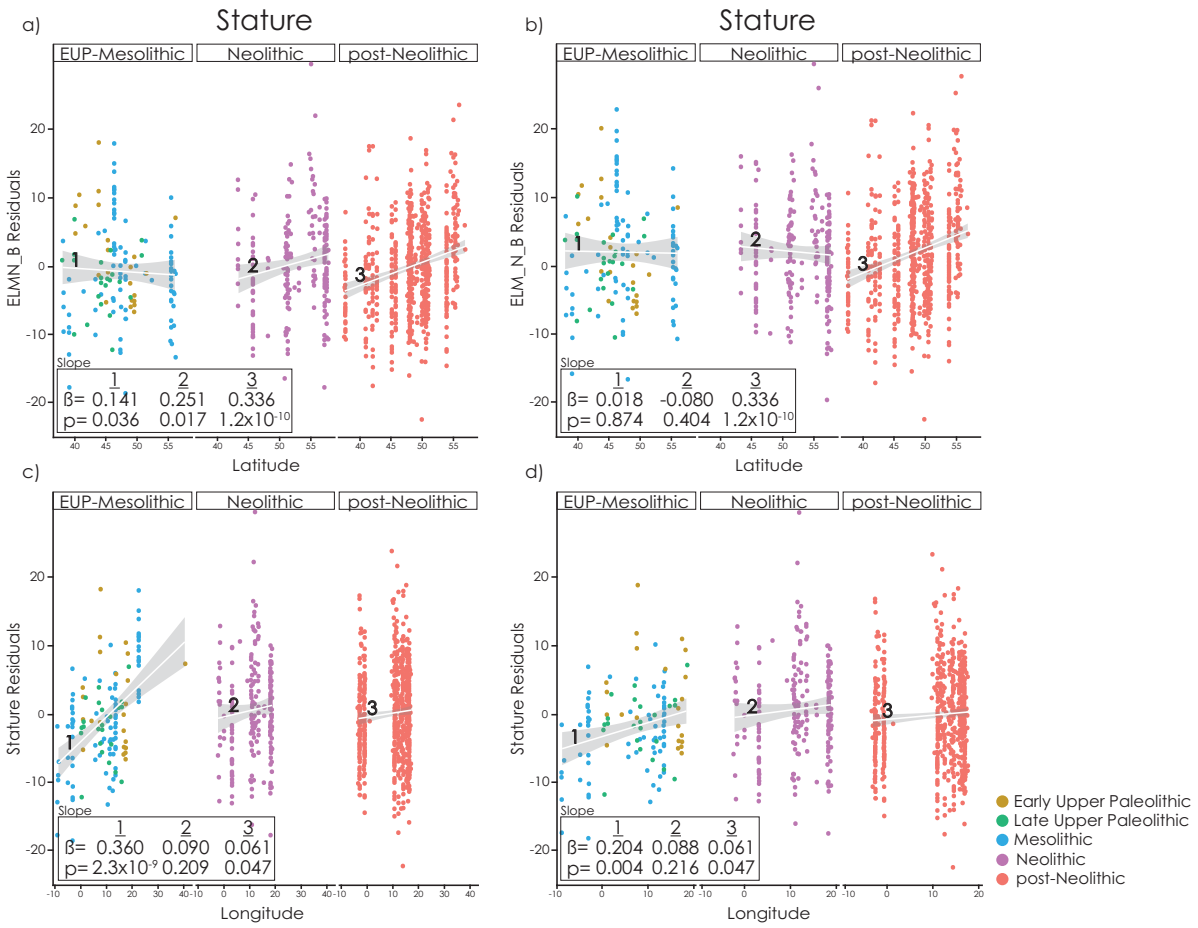
Supplementary Figure 5: Top three AIC models for stature, and the corresponding models for PRS(GWAS) and PRS(GWAS/Sibs). Row name indicates the model being tested, lowercase letters use fixed values for that time period, uppercase letters indicate the values were allowed to vary linearly with time. Number in the lower left corner of each plot indicates its place in the AIC ranking for PRS(GWAS), PRS(GWAS/Sibs) and Stature, green color indicates a good fit (rank 1-3/29 models), yellow a medium fit (rank 4-6/29 models), and red a poor fit (rank 7 or lower out of 29 models).



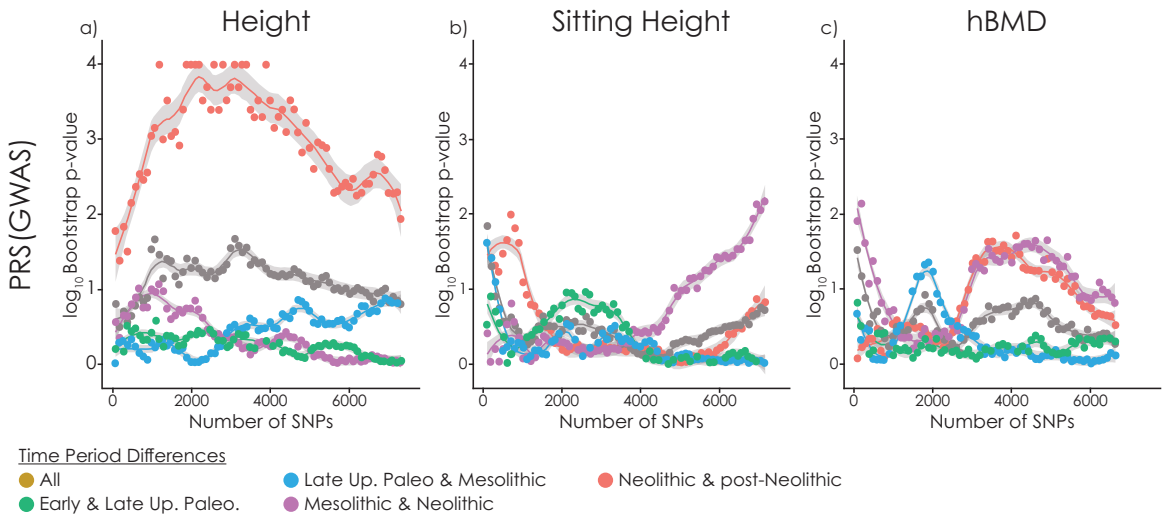
Supplementary Figure 6: Changes in standing height PRS though time with PRS constructed using 100kb clumping windows. Each point is an ancient individual, lines show fitted values, grey area is the 95% confidence interval. **a-b**) Constant values in the EUP, LUP-Neolithic and post-Neolithic; **c-d**) A linear trend with time between EUP-Neolithic and a different trend in the post-Neolithic.



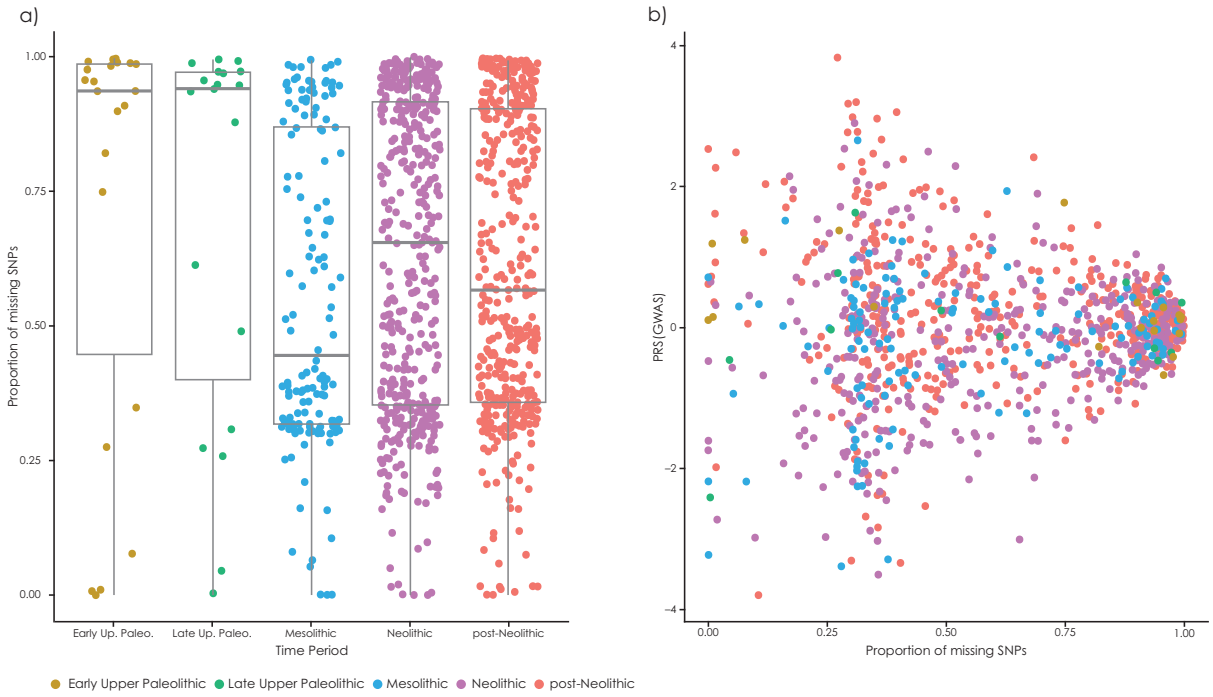
Supplementary Figure 7: Changes in standing height PRS though time with PRS constructed using 500kb clumping windows. Each point is an ancient individual, lines show fitted values, grey area is the 95% confidence interval. **a-b**) Constant values in the EUP, LUP-Neolithic and post-Neolithic; **c-d**) A linear trend with time between EUP-Neolithic and a different trend in the post-Neolithic.



Supplementary Figure 8: Geographic gradients in stature under different models. Here we show EUP-Mesolithic, Neolithic and post-Neolithic periods separately, instead of EUP-Neolithic and post-Neolithic as in the main text. **a**) Latitudinal gradient using residuals of the ELMN_B model (Fig. 1f). **b**) Latitudinal gradient using residuals of the ELM_N_B model (Supplementary Fig. 5c). Note that there is no longer a gradient in the Neolithic, so the apparent geographic gradient can equally be explained by temporal change interacting with sampling. **c**) Longitudinal gradient using residuals of the ELMN_B model (Fig. 1f); the gradient is steepest in the Mesolithic and earlier. **d**) Longitudinal gradient with relatively tall Eastern Mesolithic and Paleolithic samples removed.



Supplementary Figure 9: Selection test as in Figure 6, but using GWAS results generated on a subsample of individuals so that the standard error of the effect size estimates is the same as the standard error of the within-sibling pair estimates.



Supplementary Figure 10: Effect of missing data on PRS. **a**) proportion of missing data as a function of group. **b**) PRS(GWAS) as a function of missing data proportion.