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.
[d] Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge.
[e] Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA.
[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. 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.

Significance

Measurements of prehistoric human skeletal remains provide a record of changes in height and other anthropometric traits, over time. Often, these changes are interpreted in terms of plastic developmental response to shifts in diet, climate or other environmental factors. These changes can also be genetic in origin but, until recently, it has been impossible to separate the effects of genetics and environment. Here we use ancient DNA to directly estimate genetic changes in phenotypes and to identify changes driven not by genetics, but by environment. We show that changes over the past 35,000 years are largely predicted by genetics, but also identify specific shifts that are more likely to be environmentally driven.

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). Height can respond rapidly in a developmentally plastic manner to changes in environment, as demonstrated by large increases in Europe, and worldwide, during the secular trends of the 19th and 20th centuries (1, 4). In European countries today, mean adult male height is ~170-180 cm (1). It is broadly agreed that prehistoric changes were likely to have been driven by a combination of environmental (e.g. climate or diet) and genetic factors including drift, admixture and selection (4, 7-9), although the effects of these 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

70 other populations (19). The extent to which differences in PRS between populations are
71 predictive of population-level differences in phenotype is currently unclear (20). Recent studies
72 have demonstrated that such differences may partly be artifacts of correlation between
73 environmental and genetic structure in the original GWAS (21, 22). These studies also
74 suggested best practices for PRS comparisons, including the use of GWAS summary statistics
75 from large homogenous studies (instead of meta-analyses), and replication of results using
76 summary statistics derived from within-family analyses that are robust to population
77 stratification.
78
79 Bearing these caveats in mind, PRS can be applied to ancient populations thanks to recent
80 technological developments that have dramatically increased ancient DNA (aDNA) sample sizes.
81 These have provided remarkable insights into the demographic and evolutionary history of both
82 modern and archaic humans across the world (23-25), particularly in Europe, and allow us to
83 track the evolution of variants underlying phenotypes ranging from pigmentation to diet (26-
84 29). In principle, PRS applied to ancient populations could similarly allow us to make inference
85 about the evolution of complex traits. A few studies have used PRS to make predictions about
86 the relative statures of ancient populations (29-31) but looked at only a few hundred samples in
87 total and did not compare their predictions with stature measured from skeletons. Here, we
88 compare measured skeletal data to genetic predictions and directly investigate the genetic
89 contribution to height independent of environmental effects acting during development.

90 Results

91 PRS and skeletal measurements

92 We collected published aDNA data from 1071 ancient individuals from Western Eurasia (west of
93 50° E), dated to between 38,000 and 1100 years before present (BP) (27, 29, 30, 32-57). Using
94 GWAS summary statistics for height from the UK Biobank (generated and made available by the
95 Neale lab: http://nealelab.is/), we computed height PRS for each individual, using a P-value
96 cutoff of 10^{-6}, clumping variants in 250kb windows, and replacing missing genotypes with the
97 mean across individuals (Methods). We refer to this as PRS(GWAS). Because of concerns about

98 GWAS effect sizes being inflated by residual population stratification, we also computed a PRS
99 where we used GWAS P-values to select SNPs, but computed the PRS using effect sizes
100 estimated from a within-family test for ~17,000 sibling pairs from UK Biobank (Methods) which
101 we refer to as PRS(GWAS/Sibs), and which should be unaffected by stratification. We also
102 obtained stature estimates from 1159 individuals dating to between 33,700 and 1100 BP taken
103 from a larger dataset of 2177 individuals with stature and body proportion estimates from
104 substantially complete skeletons (4, 58). There is limited overlap in these datasets (12
105 individuals), but they cover the same time periods and broadly the same geographic locations
106 (Supplementary Fig. 1), although the genetic data contain more individuals from further east
107 (30-50° E) compared to the skeletal data. We divided these individuals into five groups based
108 on date: Early Upper Paleolithic (>25,000 BP; EUP), Late Upper Paleolithic (25,000-11,000 BP;
109 LUP), Mesolithic (11,000-5500 BP), Neolithic (8500-3900 BP) and post-Neolithic (5000-1100 BP,
110 including the Copper and Bronze Ages, plus later periods), resolving individuals in the
111 overlapping periods using either archaeological or genetic context (Methods). These groups
112 broadly correspond to transitions in both archaeological culture and genetic ancestry (33, 38,
113 59) (Supplementary Fig. 1c-d, Supplementary Table 1).
114

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

116 Both PRS and skeletal stature decreased from the EUP to Mesolithic periods and increased
117 between the Neolithic and post-Neolithic (Supplementary Fig. 2). Fitting group (time period) as
118 a covariate, we found a significant effect on PRS(GWAS) (ANOVA P= 1.9×10^{-9}), PRS(GWAS/Sibs)
119 (P=0.045) and skeletal stature (P=2.8×10^{-11}). There was no evidence of difference between
120 LUP, Mesolithic and Neolithic groups (Supplementary Fig. 3a-b), so we merged these three
121 groups (we refer to the merged group as LUP-Neolithic). We find that PRS(GWAS) in the LUP-
122 Neolithic period is 0.47 standard deviations (SD) lower than in the EUP (P=0.002), and 0.40 SD
123 lower (P= 8.7×10^{-11}) than in the post-Neolithic period (Fig. 1a). PRS(GWAS/Sib) shows a very
124 similar pattern (Fig. 1b), demonstrating that this is not a result of differential relatedness of the
125 ancient individuals to the structured present-day GWAS populations. Skeletal stature shows a
126 qualitatively similar pattern to the genetic predictions (Fig. 1c), with a 1.5 SD (9.6cm; P=$2.9\times10^{-}$

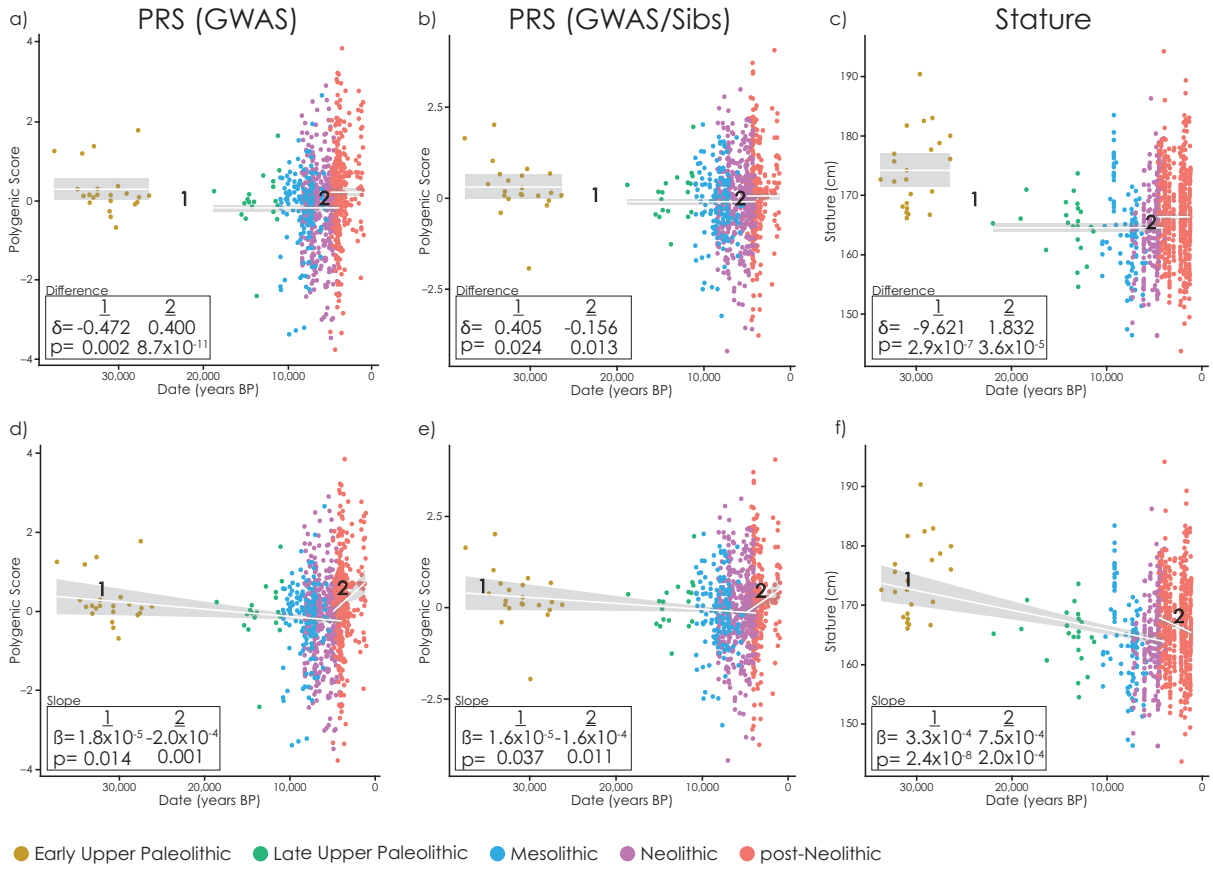

Figure 1: Changes in height PRS and stature though time. Each point is an ancient individual, white lines show fitted values, grey area is the 95% confidence interval, and boxes show parameter estimates and p-values for difference in means (δ) and slopes (β). **a-c**) PRS(GWAS) (a), PRS(GWAS/Sibs) (b) 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.

^{7}) difference between EUP and LUP-Neolithic and a 0.27 SD (1.8cm; P=3.6x10^{-5}) difference between LUP-Neolithic and post-Neolithic. Broad patterns of change in stature over time are therefore consistent with genetic predictions.

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=2.4×10^{-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=2.0×10^{-4}).

To further explore these trends, we fitted a broader range of piecewise linear models to both datasets (Methods; Supplementary Table 1; Supplementary Fig. 4-6). 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 2). 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. 4-6). Some models suggest that the increase in stature–but not PRS–may have started during the Neolithic (Supplementary Fig. 6a-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. 7-8).

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

Geographic variation in standing height

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

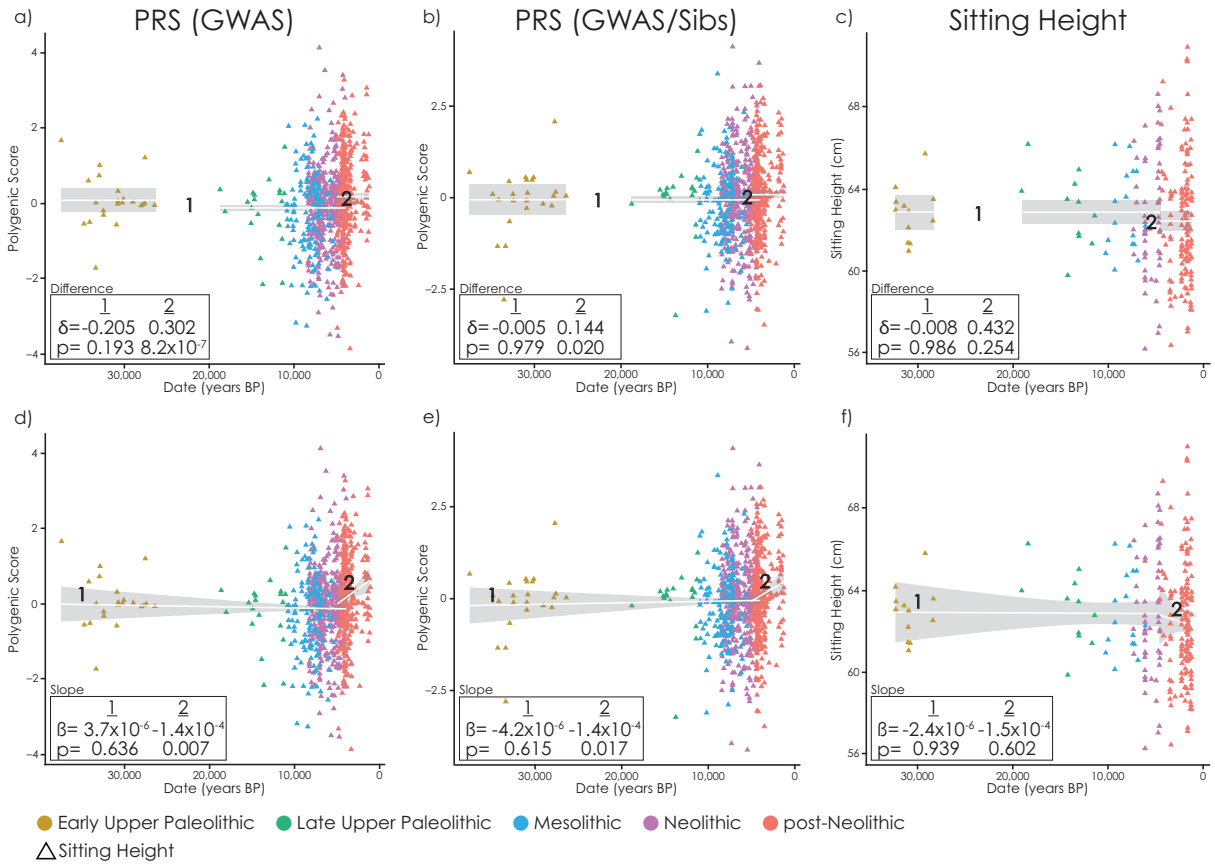

Figure 2: Changes in sitting height PRS and sitting height though time. Each point is an ancient individual, lines show fitted values, grey area is the 95% confidence interval, and boxes show parameter estimates and p-values for difference in means (δ) and slopes (β). **a-c**) PRS(GWAS) (a), PRS(GWAS/Sibs) (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.

184 geographic variation in PRS could explain these geographic trends, as it partially explains
185 temporal trends.
186
187 We regressed the residuals from our fitted linear height model (the model shown in Fig. 1d-f)
188 on longitude and latitude. Stature increases significantly with latitude ($P=1.2\times10^{-10}$) in the post-
189 Neolithic period. PRS(GWAS) increases in the post-Neolithic (P=0.006) although this is not
190 replicated by PRS(GWAS/Sibs) (P=0.557). PRS does not increase significantly with latitude in the
191 EUP-Neolithic period. There is some evidence of a modest trend in stature in the EUP-Neolithic
192 period (Fig. 3c). However, there is only evidence for this in the Neolithic, not in the EUP-
193 Mesolithic (Supplementary Fig. 9a). Further, because time and geography are correlated in our
194 Neolithic sample, this can also be explained by a temporal increase during the Neolithic, in
195 which case there is no geographic trend (Supplementary Fig. 9b).
196
197 In contrast to latitude, there is a significant increasing trend of stature with longitude before
198 but not during the Neolithic (0.36 cm/degree $P=1.6\times10^{-7}$; Fig. 3, Supplementary Fig. 9c). This is
199 partly driven by a small number of samples from a single site, but still persists if these samples
200 are removed (0.20 standardized residuals per degree, P=0.004; Supplementary Fig. 9d). There is
201 little or no trend (0.06 cm/degree; P=0.047) in the post-Neolithic period (Fig. 3f). We find no
202 evidence for longitudinal clines in PRS. In summary, we find that stature increases with latitude
203 in the post-Neolithic, possibly in the Neolithic, but not before. This cline may have a genetic
204 basis. Stature also increases with longitude, particularly in the Mesolithic, but this cline is not
205 predicted by genetics.
206

207 Correlated changes in bone density PRS and femoral bending strength

208 Beyond stature, we wanted to investigate the utility of using PRS to interpret other measurable
209 phenotypes in ancient individuals. Decreased mobility though time, associated with large-scale
210 lifestyle transitions between hunting-gathering, agriculture, and ultimately modern
211 industrialism, is well documented through declines in lower limb bone diaphyseal strength and
212 trabecular density (4, 65, 66). Today, heel bone mineral density (hBMD) is often used as an

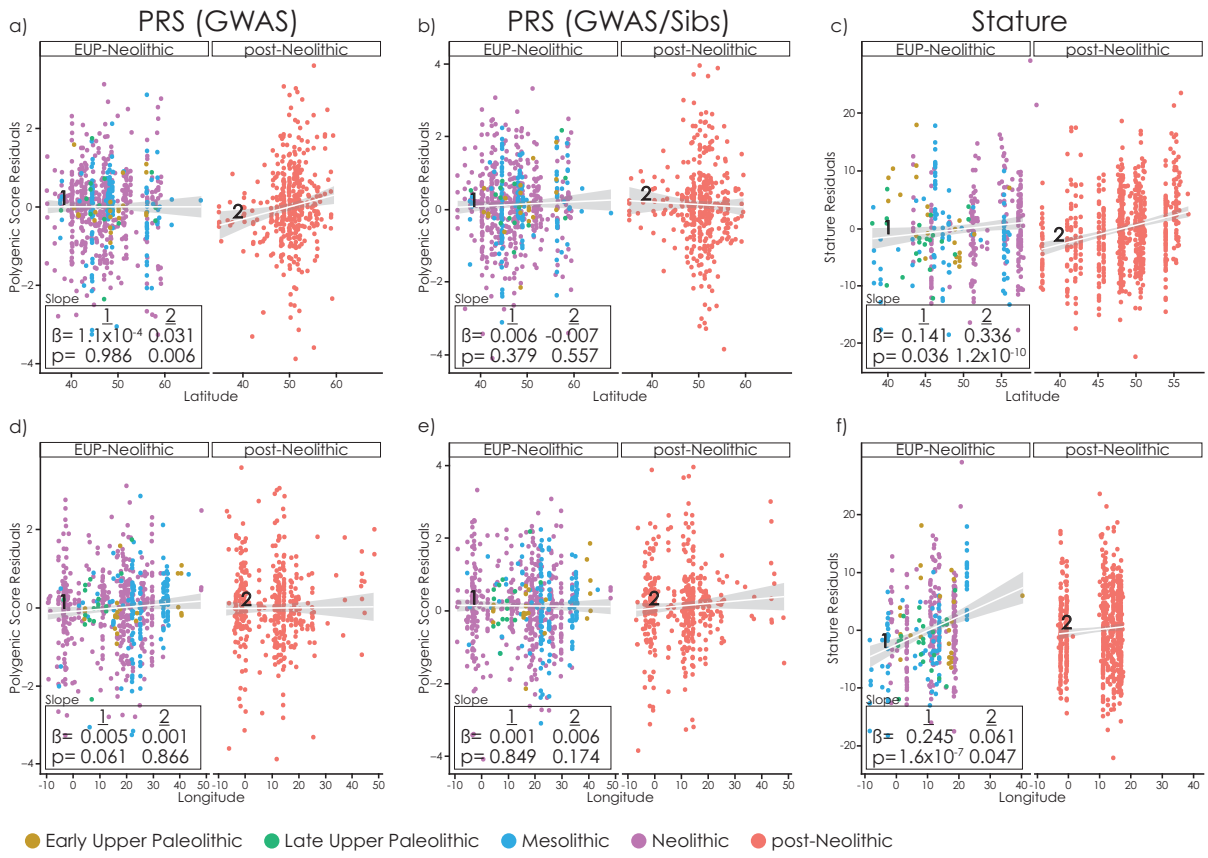

Figure 3: Geographic variation in PRS and skeletal standing height. Residuals for the linear height model (Fig. 1 d-f) against **a-c**) latitude and **d-f**) longitude. Each point is an ancient individual, lines show fitted values, grey area is the 95% confidence interval, and boxes show parameter estimates (β) and p-values for slopes.

213 indicator of general activity levels in younger people (67) and of osteoporosis in older
214 individuals (68, 69); UK Biobank has GWAS data for this trait, indirectly estimated by
215 ultrasound. However, evaluating differences in BMD in archaeological and paleontological
216 specimens can be problematic. In the short term soil leaches bone minerals, while later the
217 bone begins to fossilize, leading to unpredictable patterns of density in ancient remains (70)
218 and requiring special processing methods (65) that are difficult to apply to large samples.
219 However, femoral diaphyseal bending strength can be calculated from bone cross-sectional
220 geometric measurements that are not as affected by bone preservation (71). Here we focus on
221 anteroposterior bending strength (section modulus) of the midshaft femur (FZx), which has
222 been linked specifically to mobility (72). Since both trabecular density and diaphyseal strength
223 should respond to mobility and activity levels, we reasoned that they would be likely to show
224 correlated patterns of temporal change. Following established protocols (71), we standardized
225 FZx first by sex, then the product of estimated body mass and femoral length (4).
226 Qualitatively, PRS(GWAS) and FZx show similar patterns, decreasing through time (Fig. 4,
227 Supplementary Fig. 2-3). There is a significant drop in FZx (Figure 4c) from the Mesolithic to
228 Neolithic ($P= 1.2\times10^{-8}$) and again from the Neolithic to post-Neolithic ($P=1.5\times10^{-13}$). PRS(GWAS)
229 for hBMD decreases significantly from the Mesolithic to Neolithic (Fig. 4a; $P=5.5\times10^{-12}$), which is
230 replicated in PRS(GWAS/Sibs) ($P=7.2x10^{-10}$; Fig. 4b); neither PRS shows evidence of decrease
231 between the Neolithic and post-Neolithic. We hypothesize that both FZx and hBMD responded
232 to the reduction in mobility that accompanied the adoption of agriculture (72). In particular, the
233 lower genetic hBMD and skeletal FZx of Neolithic compared to Mesolithic populations may
234 represent adaptation to the same change in environment although we do not know the extent
235 to which the change in FZx was driven by genetic or plastic developmental response to
236 environmental change. On the other hand, FZx continues to decrease between the Neolithic
237 and post-Neolithic (Fig. 4c,f)–which is not reflected in the hBMD PRS (Fig. 4 a-b,d-e). One
238 possibility is that the two phenotypes responded differently to the post-Neolithic intensification
239 of agriculture. Another is that the non-genetic component of hBMD, which we do not capture
240 here, also continued to decrease.
241

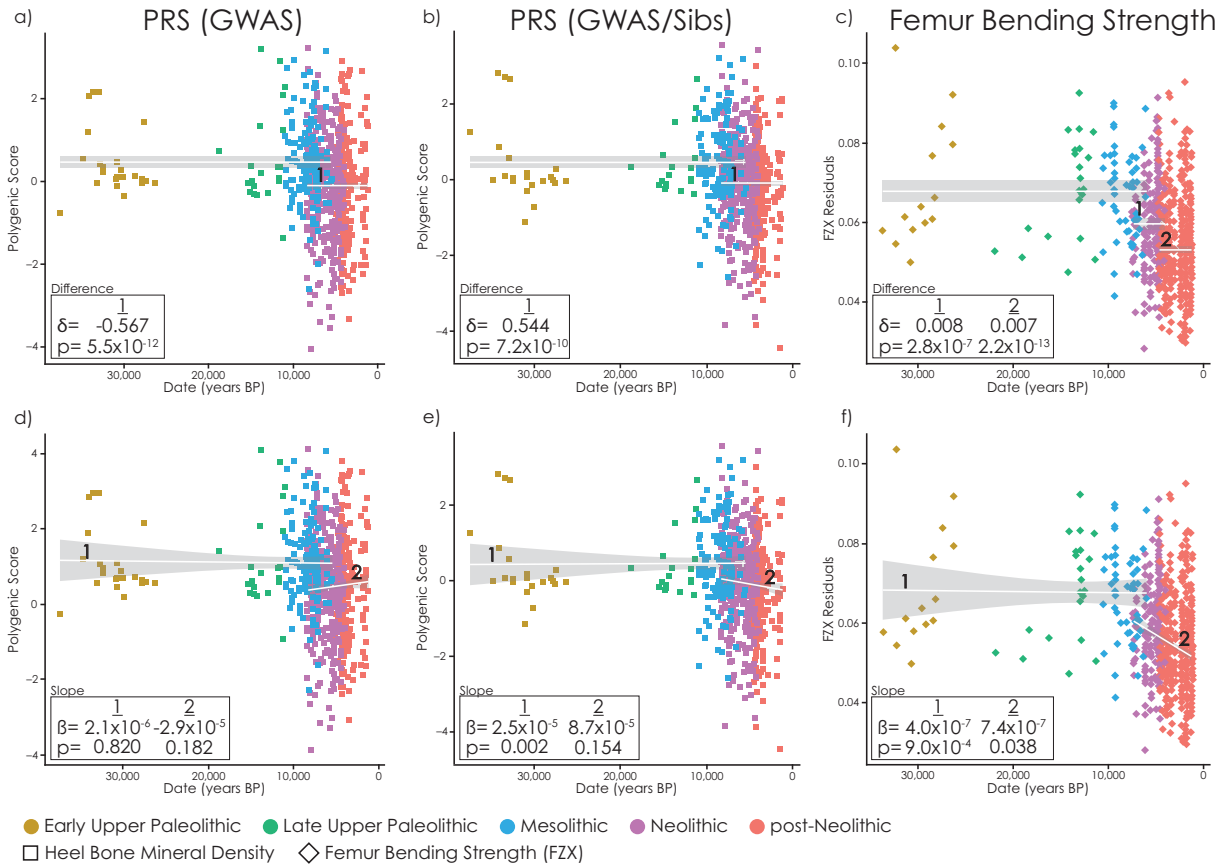

Figure 4: Changes in heel bone mineral density (hBMD) PRS and femur bending strength (FZx) though time. Each point is an ancient individual, lines show fitted values, grey area is the 95% confidence interval, and boxes show parameter estimates and p-values for difference in means (δ) and slopes (β). **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.

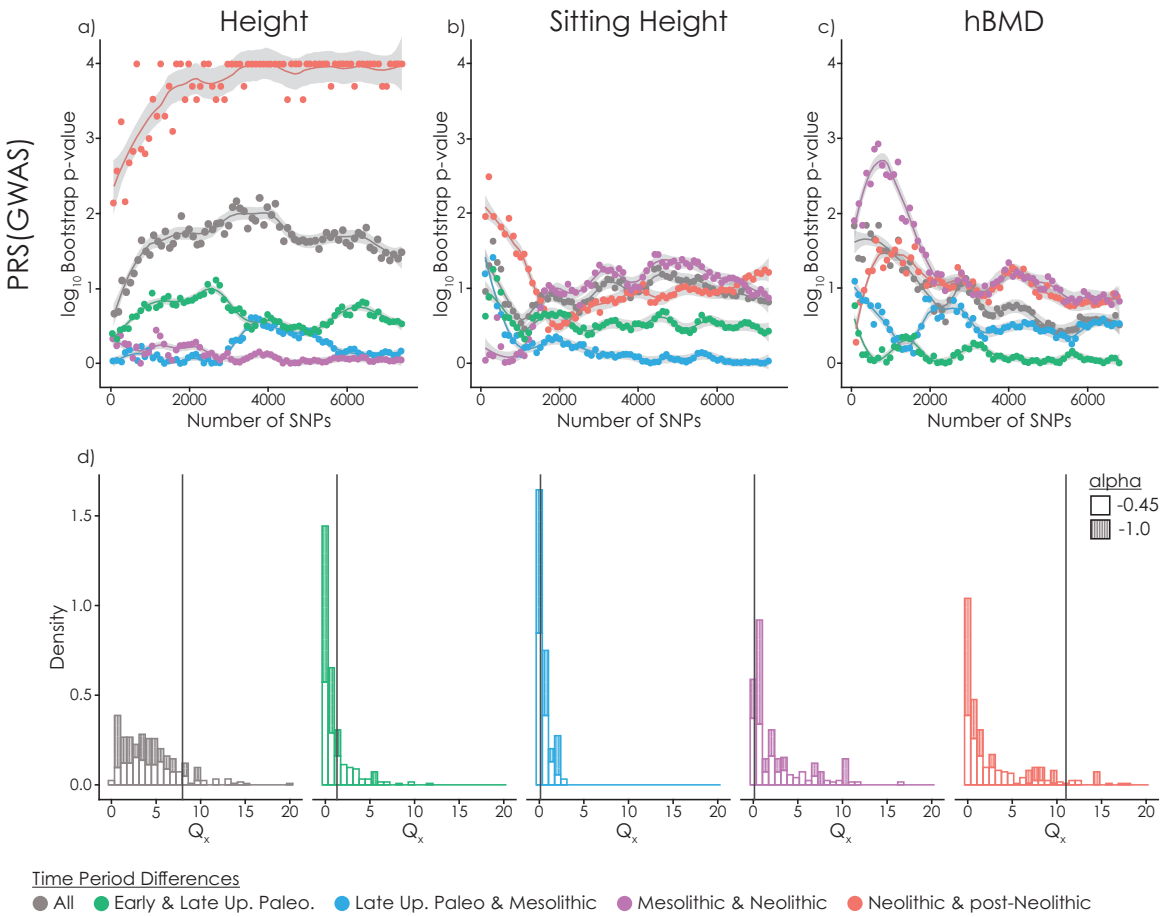

Figure 5: Signals of selection on standing height, sitting height and bone mineral density. **a-c**) Log_{10} bootstrap P-values for the Q_x statistics (y-axis, capped at 4) for GWAS signals. We tested each pair of adjacent populations, and the combination of all of them ("All"). We ordered PRS SNPs by increasing P-value and tested the significance of Q_x for increasing numbers of SNPs (x-axis). **d**) Distribution of Q_x statistics in simulated data (Methods). Observed height values for 6800 SNPs shown by vertical lines.

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

243 The Q_x statistic (73) can be used to test for polygenic selection. We computed it for increasing
244 numbers of SNPs from each PRS (Fig. 5a-c), between each pair of adjacent time periods and
245 over all time periods. We estimated empirical P-values by replacing allele frequencies with
246 random derived allele frequency-matched SNPs from across the genome, while keeping the
247 same effect sizes. To check these Q_x results, we simulated a GWAS from the UK Biobank dataset
248 (Methods), then used these effect sizes to compute simulated Q_x statistics. The Q_x test suggests
249 selection between the Neolithic and Post-Neolithic for stature ($P<1 \times 10^{-4}$; Fig. 5a), which
250 replicates using effect sizes estimated within siblings ($10^{-4}<P<10^{-2}$; Supplementary Fig. 10). The
251 reduction in the sibling effect compared to the GWAS effect sizes is consistent with the
252 reduction expected from the lower sample size (Supplementary Figure 10). However, several
253 (11/200) simulated datasets produce higher Q_x values than observed in the real data (Fig. 5d).
254 This suggests that re-estimating effect sizes between siblings may not fully control for the effect
255 of population structure and ascertainment bias on the Q_x test. The question of whether
256 selection contributes to the observed differences in height PRS remains unresolved.
257
258 For sitting height, we find little evidence of selection in any time period ($P<10^{-2}$) We conclude
259 that there was most likely selection for increased standing but not sitting height in the Steppe
260 ancestors of Bronze Age European populations, as previously proposed (29). One potential
261 caveat is that, although we re-estimated effect sizes within siblings, we still used the GWAS
262 results to identify SNPs to include. This may introduce some subtle confounding, which remains
263 a question for future investigation. Finally, using GWAS effect sizes, we identify some evidence
264 of selection on heel BMD when comparing Mesolithic and Neolithic populations ($10^{-3}<P<10^{-2}$;
265 Fig. 5c). However, this signal is relatively weak when using within-sibling effect sizes, and
266 disappears when we include more than about 2000 SNPs.

267 ## Discussion

268 We showed that the well-documented temporal and geographic trends in stature in Europe
269 between the Early Upper Paleolithic and the post-Neolithic period are broadly consistent with

270 those that would be predicted by polygenic risk scores (PRS) computed using present-day
271 GWAS results combined with ancient DNA. However, because of the limited predictive power of
272 current PRS, we cannot provide a quantitative estimate of how much of the variation in
273 phenotype between populations might be explained by variation in PRS. Similarly, we cannot
274 say whether the changes were continuous, reflecting evolution through time, or discrete,
275 reflecting changes associated with known episodes of replacement or admixture of populations
276 that have diverged genetically over time. Finally, we find cases where predicted genetic
277 changes are discordant with observed phenotypic changes–emphasizing the role of
278 developmental plasticity in response to environmental change and the difficulty in interpreting
279 differences in PRS in the absence of phenotypic data.
280
281 Our results indicate two major episodes of change in genetic height. First, there was a reduction
282 in standing height PRS–but not sitting height PRS–between the Early and Late Upper Paleolithic,
283 coinciding with an almost total population replacement (33). These genetic changes are
284 consistent with the decrease in stature–driven by leg length–observed in skeletons during this
285 time period (4, 64, 74, 75). One possibility is that the stature decrease in the ancestors of the
286 LUP populations could have been adaptive, driven by changes in resource availability (76) or to
287 a colder climate (61). Comparison between patterns of phenotypic and genetic variation
288 suggest that, on a broad scale, variation in body proportions among present-day people reflects
289 adaptation to environment largely along latitudinal gradients (77, 78). Early Upper Paleolithic
290 populations in Europe would have migrated relatively recently from more southern latitudes
291 and had body proportions that are typical of present-day tropical populations (75). The
292 populations that replaced them would have had more time to adapt to the colder climate of
293 northern latitudes. On the other hand, we do not find genetic evidence for selection on stature
294 during this time period–suggesting that the changes may have been neutral and not adaptive.
295
296 The second episode of change in height is either between the Neolithic and post-Neolithic, or
297 during the post-Neolithic period. This period is characterized by the eastward movement of
298 substantial amounts of “Steppe ancestry” into Central and Western Europe (27, 30, 38, 50). Our

299 results are thus consistent with previous results that migration and admixture from Bronze Age
300 populations of the Eurasian steppe increased genetic height in Europe (29, 30). Whether this
301 increase was driven by selection in the ancestors of these populations remains unresolved. The
302 geographic gradient of increasing skeletal stature is unclear in the Paleolithic, largely West-East
303 in the Mesolithic (7, 64) and largely South-North by the Bronze Age (4, 7, 9). Latitudinal, but not
304 longitudinal, patterns are qualitatively consistent with geographic patterns in PRS suggesting
305 that, like temporal variation, both genetics and environment contribute to geographic variation.
306
307 A major confounding factor in analysis of temporal and geographic variation in PRS, particularly
308 in the Bronze Age, is genetic population structure. Present-day European population structure
309 is correlated with geography and largely driven by variation in Steppe ancestry proportion, with
310 more Steppe ancestry in Northern Europe and less in Southern Europe (38). Suppose that
311 environmental variation in stature is also correlated with geography, and that Northern
312 Europeans are taller than Southern Europeans for entirely non-genetic reasons. Then, GWAS
313 that do not completely correct for stratification will find that genetic variants that are more
314 common in Steppe populations than Neolithic populations are associated with increased height.
315 When these GWAS results are used to compute PRS for ancient populations, they will predict
316 that Steppe ancestry populations were genetically taller simply because they are more closely
317 related to present-day Northern Europeans (21, 22). In this study, we attempted to avoid this
318 confounding in two ways: first, by using GWAS effect sizes from the UK Biobank–a homogenous
319 dataset that should be well-controlled for population stratification, and second, by replicating
320 our results after re-estimating the effect sizes within siblings, which should be robust to
321 population stratification, though less precise. However, we cannot exclude the possibility that
322 some confounding remains, for example because although we re-estimated effect sizes using
323 the within-siblings design, we still ascertained loci using the GWAS results. A related concern is
324 that, even in the absence of bias, variance explained by the PRS is likely to decrease as we move
325 back in time and ancient populations become less closely related to present-day populations
326 (19). However, our results indicate that the PRS still captures enough of the genetic variance to

327 predict changes in phenotype even though, presumably, we would gain even more resolution
328 using a PRS that captured more of the genetic variation in the phenotype.
329
330 As well as genetic contributions to phenotype, our results shed light on possible environmental
331 contributions. In some cases, we can make hypotheses about the relationship between
332 environmental or lifestyle changes, and genetic change. For example, if we interpret change in
333 femur bending strength as reflecting a decrease in mobility, the coincident Mesolithic/Neolithic
334 change in heel bone mineral density PRS can be seen as a genetic response to this change.
335 However, in the Neolithic/post-Neolithic periods, the two observations are decoupled. This
336 emphasizes the role of developmental plasticity in response to changes in environment, and of
337 joint interpretation of phenotypic and genetic variables. Even when looking at the same
338 phenotype, we find cases where genetic predictions and phenotypic data are discordant–for
339 example in post-Neolithic sitting height. We must therefore be cautious in the interpretation of
340 predicted genetic patterns where phenotypes cannot be directly measured, even if it is possible
341 to control stratification. Predicted genetic changes should be used as a baseline, against which
342 non-genetic effects can be measured and tested.

343 # Methods

344 ## Ancient DNA and polygenic risk score construction

345 We collected published ancient DNA data from 1122 ancient individuals, taken from 29
346 publications. The majority of these individuals had been genotyped using an in-solution capture
347 reagent ("1240k") that targets 1.24 million single nucleotide polymorphisms (SNPs) across the
348 genome. Because of the low coverage of most of these samples, the genotype data are pseudo-
349 haploid. That is, there is only a single allele present for each individual at each site, but alleles at
350 adjacent sites may come from either of the two chromosomes of the individual. For individuals
351 with shotgun sequence data, we selected a single read at each 1240k site. We obtained the
352 date of each individual from the original publication. Most of the samples have been directly
353 radiocarbon dated, or else are securely dated by context. We summarized the genetic
354 relationships between ancient and present-day groups by computing F_{ST} using smartpca v16000
355 (79) (Supplementary Table 1), multidimensional scaling using pairwise distances computed
356 using plink v1.90b5.3 (options --distance flat-missing 1-ibs) (80) (Supplementary Fig. 1c) and
357 unsupervised ADMXITURE (81) (Supplementary Fig. 1d).
358
359 We obtained GWAS results from the Neale lab UK Biobank page (http://www.nealelab.is/uk-
360 biobank/; Round 1, accessed February and April 2018). To compute PRS, we first took the
361 intersection of the 1240k sites and the association summary statistics. We then selected a list of
362 SNPs to use in the PRS by selecting the SNP with the lowest P-value, removing all SNPs within
363 250kb, and repeating until there were no SNPs remaining with P-value less than 10^{-6}. We then
364 computed PRS for each individual by taking the sum of genotype multiplied by effect size for all
365 included SNPs. Where an individual was missing data at a particular SNP, we replaced the SNP
366 with the average frequency of the SNP across the whole dataset. This has the effect of shrinking
367 the PRS towards the mean and should be conservative for the identification of differences in
368 PRS. We confirmed that there was no correlation between missingness and PRS, to make sure
369 that missing data did not bias the results (correlation between missingness and PRS ρ=0.02;
370 P=0.44, Supplementary Fig. 11). Finally, we normalized the PRS across individuals to have mean
371 0 and standard deviation 1.

372 We estimated within-family effect sizes from 17,358 sibling pairs in the UK Biobank to obtain
373 effect estimates that are unaffected by stratification. Pairs of individuals were identified as
374 siblings if estimates of IBS0 were greater than 0.0018 and kinship coefficients were greater than
375 0.185. Of those pairs, we only retained those where both siblings were classified by UK Biobank
376 as "white British", and randomly picked two individuals from families with more than two
377 siblings. We used Hail (82) to estimate within-sibling pair effect sizes for 1,284,881 SNPs by
378 regressing pairwise phenotypic differences between siblings against the difference in genotype.
379 We included pairwise differences of sex (coded as 0/1) and age as covariates, and inverse-rank-
380 normalized the phenotype before taking the differences between siblings. To combine the
381 GWAS and sibling results, we first restricted the GWAS results to sites where we had estimated
382 a sibling effect size and replaced the GWAS effect sizes by the sibling effects. We then restricted
383 to 1240k sites and constructed PRS in the same way as for the GWAS results.
384
385 To test whether the differences in the GWAS and GWAS/Sibs PRS results can be explained by
386 differences in power, we created subsampled GWAS estimates which matched the sibling in the
387 expected standard errors, by determining the equivalent sample size necessary and randomly
388 sampling N_{sub} individuals. $N_{sub} = \frac{N_{sib}}{2\,var(\delta_{sib})}$ where δ_{sib} is the difference in normalized
389 phenotype between siblings after accounting for the covariates age and sex.
390

391 ## Stature data

392 We obtained stature data from Ruff (2018) (4) (data file and notes available at
393 http://www.hopkinsmedicine.org/fae/CBR.html), which also includes estimated body mass,
394 femoral midshaft anteroposterior strength (FZx), and other osteometric dimensions. Statures
395 and body masses were calculated from linear skeletal measurements using anatomical
396 reconstruction or sample-specific regression formulae (4, 58). We calculated sitting height as
397 basion-bregma (cranial) height (BBH) plus vertebral column length (VCL). Analysis was restriced
398 to 1159 individuals dated earlier than 1165 BP (651 males and 508 females), of which 1130 had
399 estimates for stature, 1014 for FZx and 236 for sitting height. Sitting and standing height were
400 standardized for sex by adding the mean difference between male and female estimates to all

401 the female values. Sex differences in stature remain relatively constant over time (4), making it
402 reasonable to adjust all female heights by the same mean value. Male/Female counts in each
403 period were: EUP 13/10, LUP 15/8, Mesolithic 56/37, Neolithic 130/91 and post-Neolithic
404 437/362. For FZx we first standardized for sex as we did for stature then divided each by
405 estimated body mass multiplied by biomechanical femur length (4).
406

407 ## Grouping

408 We grouped individuals into broad categories based on date and, in some cases, archeological
409 and genetic context. All individuals were assigned to one time period group, based on median
410 age estimates of the sample obtained from the original publications. Date ranges for each time
411 period are based on a combination of historical, climatic, and archaeological factors. The Early
412 Upper Paleolithic comprises all samples older than 25,000 BP, which roughly coincides with the
413 end of the last glacial maximum (LGM). The Late Upper Paleolithic begins when the European
414 glaciers are beginning to recede (25,000 BP) and extends until 11,000 BP and a shift in lithic
415 technology that is traditionally used to delineate the beginning of the Mesolithic period.
416 Transitions between the Mesolithic, Neolithic, and Bronze Age are staggered throughout
417 Europe, so creating universally applicable date ranges is not possible. We instead defined
418 overlapping transition periods between the Mesolithic and Neolithic periods (8500-5500 BP)
419 and between the Neolithic and post-Neolithic (5000-3900 BP). For the genetic data, samples in
420 the overlapping periods were assigned based on genetic population affiliation, inferred using
421 supervised ADMIXTURE (81) which, in most of Western Europe, corresponds closely to
422 archaeological context (38, 48). In particular, the Mesolithic/Neolithic overlap was resolved
423 based on whether each individual had more (Neolithic) or less (Mesolithic) than 50% ancestry
424 related to northwest Anatolian Neolithic Farmers. The Neolithic/post-Neolithic overlap was
425 resolved based on whether individuals had more than 25% ancestry related to Bronze Age
426 Steppe populations ("Steppe ancestry"; See (83) for details). For the skeletal data, group
427 assignment in the overlapping periods was determined by the archaeology of each site.
428 Broadly, sites belonging to the Neolithic have transitioned to agricultural subsistence. Similarly,
429 post-Neolithic populations are broadly defined by evidence of metal working (Copper, Bronze

430 and Iron Ages, and later periods). In particular, we included Late Eneolithic (Copper Age) sites
431 associated with Corded Ware and Bell Beaker material culture in the post-Neolithic category
432 but for consistency with the genetic classifications, we included 8 Early Eneolithic (before 4500
433 BP) individuals in the Neolithic category, since this precedes the appearance of Steppe ancestry
434 in Western Europe. We excluded samples more recent than 1165 BP.
435

436 ## Linear models

437 We fitted a series of linear models to changes in both PRS and stature data with time. In the
438 most general model, we allow both the intercept and slope to vary between groups. We then
439 either force some of the slopes to be zero, or some of the adjacent groups to have identical
440 parameters. We describe the models using underscores to indicate changes in parameters,
441 lowercase to indicate slopes (change with respect to time) fixed to zero, and upper case to
442 indicate free slopes (i.e. linear trends with time). For example, "E_L_M_N_B" is the most
443 general model, "elmnb" indicates that all groups have the same mean and there is no change
444 with time, and "ELMN_B" indicates that the first four groups share the same parameters, and
445 the post-Neolithic has different parameters. The models shown in Figures 1 and 2 are
446 "e_lmn_b" (panels a-b), "e_lm_nb" (panel c), "ELMN_B" (panels d-e) and "ELM_NB" (panel f).
447 To analyze geographic variation, we used the residuals of the "ELMN_B" model for the PRS and
448 "ELM_NB" for skeletal stature, and fitted regressions against latitude and longitude.
449

450 ## Polygenic selection test

451 We computed bootstrap P-values for the Q_x statistic (73) by re-computing the statistic for
452 random sets of SNPs in matched 5% derived allele frequency bins (polarized using the
453 chimpanzee reference gnome panTro2). For each bootstrap replicate, we keep the original
454 effect sizes but replace the frequencies of each SNP with one randomly sampled from the same
455 bin. Unlike the PRS calculations, we ignored missing data, since the Q_x statistic uses only the
456 population-level estimated allele frequencies and not individual-level data. We tested a series
457 of nested sets of SNPs (x-axis in Fig. 5), adding SNPs in 100 SNP batches, ordered by increasing
458 P-value, down to a P-value of 0.1.

Simulated GWAS data

We simulated GWAS, generating causal effects at a subset of around 159,385 SNPs in the intersection of SNPs which passed QC in the UK Biobank GWAS, are part of the 1240k capture, and are in the POBI dataset (84). We assumed that the variance of the effect size of an allele of frequency *f* was proportional to $[f(1-f)]^{\alpha}$, where the parameter α measures the relationship between frequency and effect size (85). We performed 100 simulations with α=-1 (the most commonly used model where each SNP explains the same proportion of phenotypic variance) and 100 with α=-0.45 as estimated for height (85). We then added an equal amount of random noise to the simulated genetic values, so that the SNP heritability equaled 0.5. We tested for association between these SNPs and the simulated phenotypes. Using these results as summary statistics, we computed PRS and Q_x tests using the pipeline described above.

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 [KA2018-98559]. 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 research made use of the UK Biobank Resource under Application Number 33923. The 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* doi:10.1007/s12520-12019-00850-12523.
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.
13. Silventoinen K*, et al.* (2003) Heritability of adult body height: a comparative study of twin cohorts in eight countries. *Twin Res* 6(5):399-408.
14. Visscher PM, Hill WG, & Wray NR (2008) Heritability in the genomics era--concepts and misconceptions. *Nat Rev Genet* 9(4):255-266.
15. Wood AR*, et al.* (2014) Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet* 46(11):1173-1186.
16. Yengo L*, et al.* (2018) Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. *Hum Mol Genet* 27(20):3641-3649.
17. Lango Allen H*, et al.* (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 467(7317):832-838.
18. Marouli E*, et al.* (2017) Rare and low-frequency coding variants alter human adult height. *Nature* 542(7640):186-190.

527 19. Martin AR*, et al.* (2019) Clinical use of current polygenic risk scores may exacerbate
528 health disparities. *Nat Genet* 51(4):584-591.
529 20. Martin AR*, et al.* (2017) Human Demographic History Impacts Genetic Risk Prediction
530 across Diverse Populations. *Am J Hum Genet* 100(4):635-649.
531 21. Berg JJ*, et al.* (2019) Reduced signal for polygenic adaptation of height in UK Biobank.
532 *Elife* 8:e39725.
533 22. Sohail M*, et al.* (2019) Polygenic adaptation on height is overestimated due to
534 uncorrected stratification in genome-wide association studies. *Elife* 8:e39702.
535 23. Skoglund P & Mathieson I (2018) Ancient Genomics of Modern Humans: The First
536 Decade. *Annu Rev Genomics Hum Genet* 19:381-404.
537 24. Marciniak S & Perry GH (2017) Harnessing ancient genomes to study the history of
538 human adaptation. *Nat Rev Genet* 18(11):659-674.
539 25. Slatkin M & Racimo F (2016) Ancient DNA and human history. *Proc Natl Acad Sci U S A*
540 113(23):6380-6387.
541 26. Wilde S*, et al.* (2014) Direct evidence for positive selection of skin, hair, and eye
542 pigmentation in Europeans during the last 5,000 y. *Proceedings of the National Academy*
543 *of Sciences* 111(13):4832-4837.
544 27. Allentoft ME*, et al.* (2015) Population genomics of Bronze Age Eurasia. *Nature*
545 522(7555):167-172.
546 28. Burger J, Kirchner M, Bramanti B, Haak W, & Thomas MG (2007) Absence of the lactase-
547 persistence-associated allele in early Neolithic Europeans. *Proc Natl Acad Sci U S A*
548 104(10):3736-3741.
549 29. Mathieson I*, et al.* (2015) Genome-wide patterns of selection in 230 ancient Eurasians.
550 *Nature* 528(7583):499-503.
551 30. Martiniano R*, et al.* (2017) The population genomics of archaeological transition in west
552 Iberia: Investigation of ancient substructure using imputation and haplotype-based
553 methods. *PLoS Genet* 13(7):e1006852.
554 31. Berg JJ, Zhang X, & Coop G (2017) Polygenic Adaptation has Impacted Multiple
555 Anthropometric Traits. *biorXiv*:https://doi.org/10.1101/167551.
556 32. Cassidy LM*, et al.* (2016) Neolithic and Bronze Age migration to Ireland and
557 establishment of the insular Atlantic genome. *Proceedings of the National Academy of*
558 *Sciences* 113(2):368-373.
559 33. Fu Q*, et al.* (2016) The genetic history of Ice Age Europe. *Nature* 534(7606):200-205.
560 34. Gamba C*, et al.* (2014) Genome flux and stasis in a five millennium transect of European
561 prehistory. *Nat Commun* 5.
562 35. González-Fortes G*, et al.* (2017) Paleogenomic Evidence for Multi-generational Mixing
563 between Neolithic Farmers and Mesolithic Hunter-Gatherers in the Lower Danube
564 Basin. *Current Biology*.
565 36. Gunther T*, et al.* (2018) Population genomics of Mesolithic Scandinavia: Investigating
566 early postglacial migration routes and high-latitude adaptation. *PLoS Biol*
567 16(1):e2003703.
568 37. Gunther T*, et al.* (2015) Ancient genomes link early farmers from Atapuerca in Spain to
569 modern-day Basques. *Proc Natl Acad Sci U S A* 112(38):11917-11922.

38. Haak W, *et al.* (2015) Massive migration from the steppe is a source for Indo-European languages in Europe. *Nature* 522(7555):207-211.
39. Hofmanová Z, *et al.* (2016) Early farmers from across Europe directly descended from Neolithic Aegeans. *Proceedings of the National Academy of Sciences* 113(25):6886-6891.
40. Jones ER, *et al.* (2015) Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nature communications* 6.
41. Jones ER, *et al.* (2017) The Neolithic Transition in the Baltic Was Not Driven by Admixture with Early European Farmers. *Curr Biol* 27(4):576-582.
42. Keller A, *et al.* (2012) New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nat Commun* 3:698.
43. Kılınç Gülşah M, *et al.* (2016) The Demographic Development of the First Farmers in Anatolia. *Current Biology* 26(19):2659-2666.
44. Lazaridis I, *et al.* (2017) Genetic origins of the Minoans and Mycenaeans. *Nature* 548(7666):214-218.
45. Lazaridis I, *et al.* (2014) Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 513(7518):409-413.
46. Lipson M, *et al.* (2017) Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature* 551(7680):368-372.
47. Martiniano R, *et al.* (2016) Genomic signals of migration and continuity in Britain before the Anglo-Saxons. *Nature communications* 7:10326.
48. Mathieson I, *et al.* (2018) The genomic history of southeastern Europe. *Nature* 555(7695):197-203.
49. Olalde I, *et al.* (2014) Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* advance online publication.
50. Olalde I, *et al.* (2018) The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* 555(7695):190-196.
51. Olalde I, *et al.* (2015) A Common Genetic Origin for Early Farmers from Mediterranean Cardial and Central European LBK Cultures. *Molecular Biology and Evolution* 32(12):3132-3142.
52. Omrak A, *et al.* (2016) Genomic Evidence Establishes Anatolia as the Source of the European Neolithic Gene Pool. *Curr Biol* 26(2):270-275.
53. Saag L, *et al.* (2017) Extensive Farming in Estonia Started through a Sex-Biased Migration from the Steppe. *Curr Biol* 27(14):2185-2193 e2186.
54. Schiffels S, *et al.* (2016) Iron age and Anglo-Saxon genomes from East England reveal British migration history. *Nature communications* 7:10408.
55. Sikora M, *et al.* (2017) Ancient genomes show social and reproductive behavior of early Upper Paleolithic foragers. *Science* 358(6363):659-662.
56. Skoglund P, *et al.* (2014) Genomic Diversity and Admixture Differs for Stone-Age Scandinavian Foragers and Farmers. *Science* 344(6185):747-750.
57. Veeramah KR, *et al.* (2018) Population genomic analysis of elongated skulls reveals extensive female-biased immigration in Early Medieval Bavaria. *Proc Natl Acad Sci U S A* 115(13):3494-3499.
58. Ruff CB, *et al.* (2012) Stature and body mass estimation from skeletal remains in the European Holocene. *Am J Phys Anthropol* 148(4):601-617.

59. Skoglund P, *et al.* (2012) Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science* 336(6080):466-469.
60. Chan Y, *et al.* (2015) Genome-wide Analysis of Body Proportion Classifies Height-Associated Variants by Mechanism of Action and Implicates Genes Important for Skeletal Development. *Am J Hum Genet* 96(5):695-708.
61. Ruff CB (2002) Variation in Human Body Size and Shape. *Annual Review of Anthropology* 31:211-232.
62. Allen JA (1887) The influence of physical conditions in the genesis of species. *Radical Review* 1:108-140.
63. Bergman C (1847) Über die Verhältnisse der Wärmeökonomie der Thiere zu ihrer Grösse. *Göttinger Studien* 3:595-708.
64. Formicola V & Giannecchini M (1999) Evolutionary trends of stature in upper Paleolithic and Mesolithic Europe. *J Hum Evol* 36(3):319-333.
65. Chirchir H, *et al.* (2015) Recent origin of low trabecular bone density in modern humans. *Proc Natl Acad Sci U S A* 112(2):366-371.
66. Chirchir H, Ruff CB, Junno JA, & Potts R (2017) Low trabecular bone density in recent sedentary modern humans. *Am J Phys Anthropol* 162(3):550-560.
67. Pettersson U, Nilsson M, Sundh V, Mellstrom D, & Lorentzon M (2010) Physical activity is the strongest predictor of calcaneal peak bone mass in young Swedish men. *Osteoporos Int* 21(3):447-455.
68. Bauer DC, *et al.* (2007) Quantitative ultrasound predicts hip and non-spine fracture in men: the MrOS study. *Osteoporos Int* 18(6):771-777.
69. Bauer DC, *et al.* (1997) Broadband ultrasound attenuation predicts fractures strongly and independently of densitometry in older women. A prospective study. Study of Osteoporotic Fractures Research Group. *Arch Intern Med* 157(6):629-634.
70. Kneissel M, *et al.* (1994) Age- and sex-dependent cancellous bone changes in a 4000y BP 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. Patterson N, Price AL, & Reich D (2006) Population structure and eigenanalysis. *PLoS Genet* 2(12):e190.
80. Chang CC*, et al.* (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4:7.
81. Alexander DH, Novembre J, & Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19(9):1655-1664.
82. Hail Team (2019) Hail. pp https://github.com/hail-is/hail/releases/tag/0.2.13.
83. Mathieson S & Mathieson I (2018) FADS1 and the Timing of Human Adaptation to Agriculture. *Mol Biol Evol* 35(12):2957-2970.
84. Leslie S*, et al.* (2015) The fine-scale genetic structure of the British population. *Nature* 519(7543):309-314.
85. Schoech AP*, et al.* (2019) Quantification of frequency-dependent genetic architectures in 25 UK Biobank traits reveals action of negative selection. *Nat Commun* 10(1):790.