Supporting Information

for

Genetic evidence for natural selection in humans in the contemporary United States

Jonathan P. Beauchamp[1*]

[1] Department of Economics, Harvard University, Cambridge, MA 02138

* Corresponding Author
jpbeauch@fas.harvard.edu

SI Materials and Methods

The Study Sample and the Cohorts

The Health and Retirement Study (HRS) is a longitudinal panel study for which a representative sample of approximately 20,000 Americans have been surveyed every two years since 1992 (the HRS oversamples certain minority groups, but this does not affect the current study which only uses data on individuals of European ancestry.) All individuals were born between 1900 and 1992, with more than 95% of the individuals who have been successfully genotyped born in 1953 or earlier. All primary respondents were over the age of 50 when enrolled; spouses of the primary respondents were also interviewed, regardless of age. DNA samples have been collected for a subsample of the HRS participants between 2006 and 2008.

My main analyses focus on individuals born between 1931 and 1953. To reduce the risks of confounding by population stratification, I restrict the analyses to unrelated individuals of European ancestry (i.e., non-Hispanic White individuals; some of those unrelated individuals are spouses). To ensure that the lifetime reproductive success (LRS) variable is a good proxy for completed fertility, I only include females who were at least 45 years old when asked the number of children they ever gave birth to and males who were at least 50 years old when asked the number of children they ever fathered. (Although the HRS contains variables on self-reported age at menopause, I do not use those variables to exclude females who had not completed menopause when asked the number of children they ever gave birth to, because this could induce a selection bias and because those variables are imprecisely measured.) Further, to ensure that the sample of individuals who have been successfully genotyped is comparable to the sample of individuals who have not, I only include individuals who were enrolled in the HRS and asked the number of children they ever gave birth to or fathered in 2008 or earlier (all but two individuals who have been successfully genotyped were enrolled in the HRS and asked this question in 2008 or earlier, but a large number of individuals who have not been successfully genotyped were enrolled later). This left 6,414 females and 5,436 males with phenotypic data and 3,416 unrelated females and 2,571 unrelated males who have been successfully genotyped and who passed the quality control filters described below (and for whom I could thus construct polygenic scores). I refer to the resulting sample as the "study sample."

For some specifications, I divided the study sample into cohorts based on the individuals' birth years. The HRS1 cohort contains individuals born from 1931 to 1941; the HRS2 cohort contains individuals born from 1942 to 1947; and the HRS3 cohort contains individuals born from 1948 to 1953. Table S6 provides more details on these cohorts and on the HRS0 cohort of individuals born from 1924 to 1930. My definition of the cohorts closely resembles the HRS', except for the fact that the HRS assigns the primary respondents' spouses or partners to the primary respondents' cohorts regardless of the spouses' years of birth; hence, except for the spouses of a few primary respondents, my HRS0, HRS1, HRS2, and HRS3 cohorts are identical to the HRS' Children of the Depression, Initial HRS, War Baby, and Early Baby Boomer cohorts. The HRS cohorts were added to the HRS at different times, and dividing the study sample into cohorts allowed me to test the robustness of my results across cohorts.

To mitigate the risk of selection bias based on mortality, I excluded the HRS0 cohort from the study sample, but my main results are robust to the inclusion of that cohort (Table S4). As shown in Table S6, only 69% of females and 60% of males in the HRS0 cohort survived to 2008, the last year when individuals were genotyped. Also, for the study sample the estimates from the regressions of rLRS on the phenotypic variables are very similar between the sample of genotyped individuals and the sample of all individuals (Tables 1 and S4). By contrast, as can be seen in Table S8, for the HRS0 cohort the coefficient on phenotypic EA is negative and significant in the sample of all females ($P = 0.031$), but the corresponding coefficient in the sample of genotyped females is positive. That latter coefficient is significantly different from the coefficient in the sample of females who have not been genotyped ($P = 0.009$, *t*-test of the interaction between the coefficient on EA and a dummy for genotyped individuals), suggesting that sample selection bias is at play for the HRS0 cohort. Also to mitigate the risk of selection bias based on mortality, I excluded individuals born prior to 1924 from the study sample. I also excluded individuals born after 1953 from the study sample, as very few of them have been genotyped.

Phenotypic Variables

For my baseline analyses, I operationalize relative fitness with the relative LRS (rLRS) variable. Using rLRS instead of LRS as the measure of fitness helps control for the effects of time trends in LRS and makes it possible to interpret my estimates as rates of natural selection (1, 21). (My results are robust to using LRS instead of rLRS as the measure of fitness.) The LRS variable is the number of children females ever gave birth to or the number of children males ever fathered. Most individuals were asked this question in their first HRS interview. As Fig. S1 shows, LRS for females and males declined gradually between 1931 and 1953, from around 3 children in the early 1930s to 2 children around 1950. To construct the rLRS variable, I proceeded as follows: for any given birth year, I first calculated the mean LRS for all females born in the preceding, the given, and the following birth years; then, I obtained rLRS for females born in the given birth year by dividing their LRS by this mean LRS. I proceeded analogously for males.

The HRS contains phenotypic variables for body mass index (BMI), educational attainment (EA), height (HGT), and total cholesterol (TC). The TC phenotypic variable is an indicator variable for a self-reported health problem with high cholesterol in 1992, and not plasma concentrations of total cholesterol as in the genome-wide association study (GWAS) of TC. BMI was measured in each wave of the HRS, and the BMI phenotypic variable for females is the mean across waves of female BMI residualized on birth year dummies, plus the mean female BMI across all waves. The BMI phenotypic variable for males is defined analogously. HGT was also measured in each wave of the HRS and the HGT phenotypic variable is also defined analogously for both sexes. The EA variable is from the RAND HRS data. Table S1 presents summary statistics for these phenotypic variables and for birth year, LRS and childlessness (a dummy that is equal to one if an individual is childless). The HRS does not contain phenotypic variables for fasting glucose concentration (GLU), schizophrenia (SCZ), and age at menarche (AAM; in females).

Quality Control of the Genotypic Data and Polygenic Scores

Following (31), the individuals' genotyped (as opposed to imputed) single nucleotide polymorphisms (SNPs) were used to compute the polygenic scores. The HRS-provided binary PLINK-format data files that exclude chromosome anomalies greater than 10 MB (among other

things) were used. Following the HRS recommendations regarding the use of the genotypic data ("Quality Control Report for Genotypic Data"), only individuals in the HRS-provided "hwe_eur_keep.txt" file were used; this effectively only keeps a set of unrelated individuals of European-ancestry, with missing call rate of less than 2%, who self-identified as White, and falling within 1 SD of all self-identified non-Hispanic Whites on the first two principal components of the genetic relatedness matrix of all unrelated individuals. Also following the HRS recommendations, only SNPs in the HRS-provided "SNP_qual_maf_filter_extract.txt" file were used; among other things, this effectively removed SNPs with minor allele frequency (MAF) less than 1%, with *P*-value less than $1x10^{-4}$ on the test for Hardy–Weinberg equilibrium, and with missing call rate greater than 2%. In addition, the following filters were applied to the resulting sample, which only includes the "hwe_eur_keep.txt" individuals and the "SNP_qual_maf_filter_extract.txt" SNPs: SNPs with MAF less than 1%, with *P*-value less than $1x10^{-4}$ on the test for Hardy–Weinberg equilibrium, and (following (31)) with missing call rate greater than 1%, were excluded. 1,411,964 SNPs passed these quality control filters. For each set of GWAS summary statistics, SNPs with MAF less than 1% in the summary statistics were also dropped.

The individuals' genotyped SNPs that passed the above quality control filters and that were present in the phenotype's summary statistics files were used to construct the polygenic scores. Depending on the phenotype, there were between 505,254 and 544,493 such overlapping SNPs (except for GLU, for which there were only 22,895 such overlapping SNPs, because the GLU GWAS was conducted with only ~66,000 SNPs). The average sample sizes across the SNPs in the summary statistics used to construct the polygenic scores are $\overline{N}_{BMI} = 232{,}186$, $\overline{N}_{EA} = 386{,}098$, $\overline{N}_{HGT} = 243{,}630$, and $\overline{N}_{TC} = 92{,}793$ individuals. The summary statistics for GLU, SCZ, and MEN did not contain sample size information, but the reported samples sizes for the main GWAS of these phenotypes are $N_{GLU} = 133{,}010$, $N_{SCZ} \approx 80{,}000$, and $N_{MEN} = 132{,}989$ individuals. To avoid overfitting, I ensured that the GWAS summary statistics used to construct the polygenic scores for each phenotype are based on meta-analyses that exclude the HRS dataset (30). The HRS was not included in the GWAS of BMI, GLU, HGT, SZC, TC, and AAM. For EA, whose GWAS included the HRS (20), summary statistics based on a meta-analysis that excludes the HRS were used to construct the scores.

For the main analysis, I used LDpred (31) to construct the polygenic scores; for a robustness check, I also constructed a set of polygenic scores with PLINK (32). (For EA, polygenic scores were constructed and directly provided to me by the Social Science Genetic Association Consortium (SSGAC), following the procedure described here and which I used to construct the other scores.) For a given phenotype and a given individual i, both the LDpred and PLINK scores are calculated as the weighted sums of individual i's SNPs:

$$\text{PGS}_i = \sum_{j=1}^{m} \hat{\beta}_j g_{ji},$$

where PGS_i is individual i*'s* polygenic score, $\hat{\beta}_j$ is an estimate of SNP j's effect size (i.e., the effect of having one more copy of the reference allele at SNP j), and g_{ij} is the genotype of individual i at SNP j (coded as having 0, 1 or 2 copies of the reference allele at SNP j). For the PLINK scores, the $\hat{\beta}_j$ for a given SNP j is simply the GWAS estimate for SNP j. Because nearby SNPs tend to be in linkage disequilibrium (LD) (i.e., correlated), that GWAS estimate captures the causal effects both of SNP j and of SNPs that are in LD with SNP j. As a result, PLINK polygenic scores effectively count the causal effects of SNPs that are in LD with other SNPs multiple times. To correct for this multiple counting problem, LDpred uses information on LD between SNPs from a reference panel together with a prior on the SNPs' effect sizes, and adjusts the GWAS estimate for SNP j to obtain an estimate of the causal effect of SNP j independent of the effects of other SNPs. LDpred then uses that estimate as the $\hat{\beta}_j$ for SNP j in the above formula. The resulting LDpred score for individual i is therefore the sum of i's genotype across all SNPs, weighted by the LDpred estimates of the SNPs' causal effects; it is an unbiased predictor of the true genetic score (and of the phenotype itself) for individual i, conditional on the model assumptions and the data (31).

For the LDpred scores, the study sample individuals' genotyped SNPs that passed the above quality control filters were used as the reference panel to calculate the LD between the SNPs. (LDpred requires that the reference panel be from a population that is similar to that in which the GWAS summary statistics were estimated (31); here, the study sample individuals and almost all the individuals in the various GWAS whose summary statistics are used are of European ancestry.) The LDpred prior on the SNPs' effect sizes depends on an assumed Gaussian mixture

weight, which corresponds to the assumed fraction of causal markers. For each phenotype, LDpred scores were constructed for each of the following Gaussian mixture weights: 0.0001, 0.0003, 0.001, 0.003, 0.01, 0.03, 0.1, 0.3, 1. For BMI, EA, HGT, TC—for which there are phenotypic variables in the HRS—I selected the weights that maximize the incremental R^2 of each score in an ordinary least squares (OLS) regression of the phenotypic variable on the score and on variables for sex, birth year, birth year squared, and the top 20 principal components of the genetic relatedness matrix. For each of GLU, SCZ, AAM—for which there are no phenotypic variables in the HRS—I selected the weights that maximize the correlations between the score and known correlates of the phenotype, controlling for sex, birth year, birth year squared, and the top 20 principal components. For GLU, I selected the weight that maximizes the correlation with a variable indicating if an individual ever had diabetes or high blood sugar; for SCZ, I selected the weight that maximizes the correlations with neuroticism (59) and cognitive ability (60); for AAM, I selected the weight that maximizes the correlations with HGT and BMI (61). For each phenotype, I verified that the correlation between the score and the phenotype or the known correlates has the expected direction. Table S7 shows the parameters used to construct the scores and the sources for each phenotype's summary statistics. The scores were standardized to have mean zero and a standard deviation of one.

To construct the PLINK scores, the PLINK's "score" command was used with the default options; I then standardized the resulting scores so that they have mean zero and a standard deviation of one.

Association Analyses

For each of BMI, EA, HGT, and TC—for which there are phenotypic variables in the HRS—I regressed rLRS on the corresponding phenotypic variable, separately for males and females; these regressions included birth year dummies and HRS-defined cohort dummies, and were estimated by OLS. For all phenotypes, I also regressed rLRS on the polygenic score of the phenotype in various samples; the regressions included birth year dummies, HRS-defined cohort dummies, and the top 20 principal components of the genetic relatedness matrix, and were also estimated by OLS. (The top principal components of the genetic relatedness matrix capture the main dimensions along which the ancestry of the individuals in the dataset vary and are

commonly used to control for population stratification (27); Section 5 of the Supplemental Material of ref. (56) demonstrates how controlling for the top principal components can eliminate some spurious associations. I did not control for the top 20 principal components in the regressions of rLRS on the phenotypic variables because it is not possible to compute these for the individuals who have not been successfully genotyped.) For the regressions in the sample of females and males together, I also controlled for sex and I only included the respondent with the lowest person number (PN, an HRS identifier) in each household, as spouses very often have the same number of children, which induces a complex correlation structure between the error terms (the results for the score of EA are robust to alternative ways of selecting one respondent per household).

In all results tables, I report the coefficient estimates and standard errors, with stars to indicate statistical significance; *P*-values are included in the log files available on my website. As Lande and Arnold (16) note, the estimates and standard errors from these OLS regressions are unbiased; however, rLRS is not a continuous variable and the error terms from these regressions are not normally distributed, so the *P*-values based on the OLS *t*-statistics are only asymptotically valid. Here, given the large sample sizes, the *P*-values should be informative. To verify, I used the nonparametric bootstrap method to bootstrap the *t*-statistics of the coefficients whose estimates are reported in Tables 1 and 2, using 10,000 bootstrap samples; I then calculated the percentile of the actual *t*-statistics of each coefficient in its corresponding bootstrapped distribution to obtain the coefficient's bootstrapped *P*-value. The bootstrapped *P*-values were very similar to the *P*-values implied by the stars in Tables 1 and 2.

I also note that the standard errors and *P*-values implied by the stars in the tables reporting my estimates from regressions of rLRS on the LDpred scores do not account for the uncertainty stemming from the selection of the Gaussian mixture weights for the LDpred scores. However, all my results are robust to the use of the PLINK scores instead of the selected LDpred scores; my results for EA are robust to the use of the alternative weights of 0.3 and 1 instead of 0.1 (for lower weights, the scores of EA have much lower incremental R^2); my results for AAM are actually much stronger and more robust with the alternative weight of 1 instead of 0.3; and my insignificant results for the other phenotypes remain insignificant with different weights. All of

this implies that my main results are not driven by the weight selection procedure.

Directional Selection Differentials

The directional selection differential of a character is the change in its mean value due to natural selection in one generation. The Robertson-Price identity (33, 34) equates the directional selection differential of a character to the genetic covariance between the character and relative fitness:

$$\Delta\bar{z} = \overline{z^*} - \bar{z} = \mathrm{Cov}_a(w, z),$$

where z is the character of interest before selection, z^* is the character of interest after one generation of selection, w is relative fitness, and $\mathrm{Cov}_a(\,.\,)$ is the genetic covariance.

If we define the polygenic score of EA (as opposed to EA, as would be usual) as the character of interest, then

$$\Delta\bar{z}_{\text{PGS of EA}} = \mathrm{Cov}_a(\text{rLRS}, \text{PGS of EA}) = \mathrm{Cov}(\text{rLRS}, \text{PGS of EA})$$
$$= \frac{\mathrm{Cov}(\text{rLRS}, \text{PGS of EA})}{\mathrm{Var}(\text{PGS of EA})} = \beta_{\text{rLRS on PGS of EA}},$$

where "PGS of EA" is the polygenic score of EA. The second equality follows because the non-genetic component of rLRS is independent of the genetic component of EA (by definition) and thus of PGS of EA, the third equality follows from the fact that the score of EA has been standardized to have unit variance, and the last equality holds if the other covariates in the regression of rLRS on PGS of EA are uncorrelated with PGS of EA (which is a reasonable approximation). (This can also be derived from the framework of Lande and Arnold (16), in which the directional selection differential $\Delta\bar{z}$ is given by $\Delta\bar{z} = \overline{z^*} - \bar{z} = \boldsymbol{G} \cdot \boldsymbol{P^{-1}} \cdot \mathrm{Cov}(w, z)$, where $\boldsymbol{G}$ and $\boldsymbol{P}$ are the genotypic and phenotypic variance-covariance matrices. If we treat the score of EA as the single character of interest—as opposed to the phenotype, as would be usual—then $\boldsymbol{G} = \boldsymbol{P}$ and $\mathrm{Cov}(w, z) = \mathrm{Cov}_a(w, z)$ and the Robertson-Price identity follows.)

Hence, the directional selection differential of the score of EA is equal to the coefficient on the score of EA in the regression of rLRS on the score of EA. In other words, my estimates of the coefficients on the score of EA in Table 2 can be interpreted as directional selection differentials, or as the implied changes in the mean values of the score of EA that will occur in one generation

as a result of natural selection. Furthermore, because the score of EA has a standard deviation of one, these implied changes are expressed in Haldanes (one Haldane is one standard deviation per generation). This also applies to the scores of BMI, GLU, HGT, SCZ, TC, and AAM.

To express the estimates of the directional selection differential for the score of EA in years of education per generation instead of in Haldanes, I first rescale the score of EA in years of education and calculate its standard deviation in years of education. If we let $\widetilde{\text{PGS of EA}}$ denote the rescaled polygenic score of EA, then $\text{EA} = \widetilde{\text{PGS of EA}} + \varepsilon_{PGS}$ and $R^2_{\widetilde{\text{PGS of EA}}} = \frac{\text{Var}(\widetilde{\text{PGS of EA}})}{\text{Var}(\text{EA})}$ and it follows that $\sigma_{\widetilde{\text{PGS of EA}}} = \sigma_{\text{EA}} \cdot \sqrt{R^2_{\widetilde{\text{PGS of EA}}}}$. We can then express the directional selection differential for the score of EA in years of education per generation by multiplying the estimates of $\Delta\bar{z}_{\text{PGS of EA}} = \hat{\beta}_{\text{rLRS on PGS of EA}}$ (expressed in Haldanes) by $\sigma_{\widetilde{\text{PGS of EA}}}$:

$$\Delta\bar{z}_{\widetilde{\text{PGS of EA}}} \equiv \Delta\bar{z}_{\text{PGS of EA}} \cdot \sigma_{\widetilde{\text{PGS of EA}}}.$$

Using this formula and using the nonparametric bootstrap method with 1,000 bootstrap samples to estimate percentile confidence intervals, I obtain estimates of the rescaled directional selection differential for the score of EA of $\Delta\bar{z}_{\widetilde{\text{PGS of EA}}} = -0.022$ (95% CI: -0.036 to -0.009) and $\Delta\bar{z}_{\widetilde{\text{PGS of EA}}} = -0.022$ (95% CI: -0.040 to -0.004) years of education per generation for females and for males, or about minus one week of education per generation for both sexes. (These calculations can in principle also apply to the other phenotypes, but are not performed here because I do not have an estimate of $R^2_{\widetilde{\text{PGS of AAM}}}$ and because my estimates for the scores of BMI, GLU, HGT, SCZ, and TC are not significant.)

For a given phenotype, LDpred calculates the posterior mean of the true genetic score of the phenotype; the result is an unbiased predictor of the true genetic score of the phenotype (and of the phenotype itself), conditional on the model assumptions and the data (31). (Here, the scores are *rescaled* unbiased predictors of the true genetic scores, but this does not affect the present derivations.) Thus, if we focus on EA, we can write

$$\widetilde{g_{\text{EA}}} = \widetilde{\text{PGS of EA}} + \widetilde{U},$$

where $\widetilde{g_{\text{EA}}}$ is the true genetic score of EA, $\widetilde{U}$ is orthogonal to $\widetilde{\text{PGS of EA}}$, and where $\widetilde{g_{\text{EA}}}$,

$\widetilde{\text{PGS of EA}}$, and $\widetilde{U}$ are expressed in years of education.

To obtain an estimate of the directional selection differential of EA (or, equivalently, of the true genetic score of EA—rather than of the score of EA) expressed in years of education per generation, observe that

$$\begin{aligned}\widetilde{\Delta\bar{z}_{\text{EA}}} &\equiv \text{Cov}_a(\text{rLRS}, \text{EA}) = \text{Cov}(\text{rLRS}, \widetilde{g_{\text{EA}}}) \\ &= \frac{\text{Cov}(\text{rLRS}, \widetilde{g_{\text{EA}}})}{\text{Var}(\widetilde{g_{\text{EA}}})} \cdot \frac{\text{Var}(\widetilde{g_{\text{EA}}})}{\text{Var}(\text{EA})} \cdot \frac{\text{Var}(\text{EA})}{\text{Var}\left(\widetilde{\text{PGS of EA}}\right)} \cdot \text{Var}\left(\widetilde{\text{PGS of EA}}\right) \\ &= \frac{\text{Cov}(\text{rLRS}, \widetilde{g_{\text{EA}}})}{\text{Var}(\widetilde{g_{\text{EA}}})} \cdot h^2_{\text{EA}} / R^2_{\widetilde{\text{PGS of EA}}} \cdot \text{Var}\left(\widetilde{\text{PGS of EA}}\right) \\ &= \hat{\beta}_{\text{rLRS on } \widetilde{g_{\text{EA}}}} \cdot h^2_{\text{EA}} / R^2_{\widetilde{\text{PGS of EA}}} \cdot \text{Var}\left(\widetilde{\text{PGS of EA}}\right),\end{aligned}$$

where h^2_{EA} is the heritability of EA and $R^2_{\widetilde{\text{PGS of EA}}}$ is the R^2 of the score. The second equality follows because the non-genetic component of rLRS is independent of the genetic component of EA (by definition).

Under the assumption that $\text{E}\left[\text{rLRS}\middle|\widetilde{\text{PGS of EA}}, \widetilde{U}\right] = \beta \cdot \widetilde{\text{PGS of EA}} + \beta \cdot \widetilde{U} = \beta \cdot \widetilde{g_{\text{EA}}}$ (or, equivalently, that $\beta_{\text{rLRS on } \widetilde{g_{\text{EA}}}} = \beta_{\text{rLRS on } \widetilde{\text{PGS of EA}}}$), the following holds: $\beta_{\text{rLRS on } \widetilde{g_{\text{EA}}}} = \beta_{\text{rLRS on } \widetilde{\text{PGS of EA}}} = \beta_{\text{rLRS on PGS of EA}} / \sigma_{\widetilde{\text{PGS of EA}}} = \Delta\bar{z}_{\text{PGS of EA}} / \sigma_{\widetilde{\text{PGS of EA}}}$. It follows that

$$\begin{aligned}\widetilde{\Delta\bar{z}_{\text{EA}}} &= \Delta\bar{z}_{\text{PGS of EA}} / \sigma_{\widetilde{\text{PGS of EA}}} \cdot h^2_{\text{EA}} / R^2_{\widetilde{\text{PGS of EA}}} \cdot \text{Var}\left(\widetilde{\text{PGS of EA}}\right) \\ &= \Delta\bar{z}_{\text{PGS of EA}} \cdot \sigma_{\widetilde{\text{PGS of EA}}} \cdot h^2_{\text{EA}} / R^2_{\widetilde{\text{PGS of EA}}} \\ &= \Delta\widetilde{\bar{z}_{\text{PGS of EA}}} \cdot h^2_{\text{EA}} / R^2_{\widetilde{\text{PGS of EA}}}.\end{aligned}$$

In other words, under the assumption that the coefficient on the score of EA in a regression of rLRS on the score is equal to the coefficient on the true genetic score of EA in a regression of rLRS on the true genetic score, it follows that the directional selection differential of EA is equal to the directional selection differential of the score multiplied by the ratio of the heritability of EA to the R^2 of the score of EA.

Estimates of the heritability of EA vary substantially across studies and countries, but a recent meta-analysis of existing heritability estimates of EA obtained a mean of ~0.40 (35). Assuming

that $h^2_{\text{EA}} = 0.40$, using my earlier estimates of $R^2_{\text{PGS of EA}}$ and $\Delta\bar{z}_{\widetilde{\text{PGS of EA}}}$, and using the nonparametric bootstrap method with 1,000 bootstrap samples to estimate percentile confidence intervals, I obtain estimates of the directional selection differential of EA of $\widehat{\Delta\bar{z}_{\text{EA}}} = -0.108$ (95% CI: -0.177 to -0.045) and $\widehat{\Delta\bar{z}_{\text{EA}}} = -0.128$ (95% CI: -0.237 to -0.025) years of education per generation for females and for males, respectively—which is equivalent to -1.30 (95% CI: -2.12 to -0.54) and -1.53 (95% CI: -2.85 to -0.31) months of education per generation for females and males, respectively. (These estimates of the confidence intervals do not account for the uncertainty in the value of the heritability of EA, nor for the uncertainty stemming from the selection of the Gaussian mixture weights for the LDpred scores.)

Testing for Nonlinear Selection

Lande and Arnold (16) show that, under the assumption that the characters have a multivariate normal distribution before selection, the estimates from a quadratic regression of relative fitness on all the characters and their squares and interactions together can inform whether the characters are under either of the three types of nonlinear selection: stabilizing, disruptive, or correlational selection. In that framework, the coefficients on the characters capture the forces of directional selection; the coefficients on the squared characters capture the forces of stabilizing and disruptive selection acting directly on the variances of the characters and will be positive for the characters that are under disruptive selection and negative for the characters that are under stabilizing selection; and the coefficients on the interacted characters capture the forces of correlational selection (9) as well as the impact of selection on the covariance between the characters. All these coefficients are needed to project evolutionary changes over more than one generation.

I treat the polygenic scores themselves (as opposed to the phenotypes, as is usual) as the characters of interest and estimated the quadratic regression of rLRS on all the scores and their squares and interactions together, along with birth year dummies, HRS-defined cohort dummies, and the top 20 principal components of the genetic relatedness matrix, separately for females and males in the study sample, by OLS.

Table S5 reports the results. For both females and males, few of the coefficients on the squared

and interacted scores are significant at the 5%-level. The most significant estimate is that of the coefficient on the interaction of the scores of SCZ and TC in females ($P = 0.007$), but it is not significant after Bonferroni correction (it is not significant after Bonferroni correction for more than seven tests; coefficients on 28 squared or interacted scores are tested for females, and 21 are tested for males, so 49 tests are conducted in total). There is thus no solid evidence that stabilizing, disruptive, or correlational selection has been operating on the genetic variants associated with the various phenotypes. Comparing Table S5 to Table 2, the coefficient estimates on the scores from the quadratic regressions are very similar to those from the regressions of rLRS on each score individually. This is not surprising, since the correlations between most scores are low and the coefficients on the squared and interacted scores are small. I emphasize that these null results could be attributable to the polygenic scores being imperfect proxies for the true genetic scores; they do not prove that there has been no stabilizing, disruptive, or correlational selection of the genetic variants associated with the various phenotypes.

Additional acknowledgements

The Health and Retirement Study (HRS) is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan.

The polygenic scores of EA were accessed under Section 4 of the Data Sharing Agreement of the Social Science Genetic Association Consortium (SSGAC). I contributed to the GWAS of EA reported in Okbay et al. (20); in accordance with SSGAC policy, I acknowledge the remaining authors of that paper:

Aysu Okbay, Mark A. Fontana, James J. Lee, Tune H. Pers, Cornelius A. Rietveld, Patrick Turley, Guo-Bo Chen, Valur Emilsson, S. Fleur W. Meddens, Sven Oskarsson, Joseph K. Pickrell, Kevin Thom, Pascal Timshel, Ronald de Vlaming, Abdel Abdellaoui, Tarunveer S. Ahluwalia, Jonas Bacelis, Clemens Baumbach, Gyda Bjornsdottir, Johannes H. Brandsma, Maria Pina Concas, Jaime Derringer, Nicholas A. Furlotte, Tessel E. Galesloot, Giorgia Girotto, Richa Gupta, Leanne M. Hall, Sarah E. Harris, Edith Hofer, Momoko Horikoshi, Jennifer E. Huffman, Kadri Kaasik, Ioanna P. Kalafati, Robert Karlsson, Augustine Kong, Jari Lahti, Sven J. van der Lee, Christiaan de Leeuw, Penelope A. Lind, Karl-Oskar Lindgren, Tian Liu, Massimo Mangino, Jonathan Marten, Evelin Mihailov, Michael B. Miller, Peter J. van der Most, Christopher Oldmeadow, Antony Payton, Natalia Pervjakova, Wouter J. Peyrot, Yong Qian, Olli Raitakari, Rico Rueedi, Erika Salvi, Börge Schmidt, Katharina E. Schraut, Jianxin Shi, Albert V. Smith, Raymond A. Poot, Beate St Pourcain, Alexander Teumer, Gudmar Thorleifsson, Niek Verweij, Dragana Vuckovic, Juergen Wellmann, Harm-Jan Westra, Jingyun Yang, Wei Zhao, Zhihong Zhu, Behrooz Z. Alizadeh, Najaf Amin, Andrew Bakshi, Sebastian E. Baumeister, Ginevra Biino, Klaus Bønnelykke, Patricia A. Boyle, Harry Campbell, Francesco P. Cappuccio, Gail Davies, Jan-Emmanuel De Neve, Panos Deloukas, Ilja Demuth, Jun Ding, Peter Eibich, Lewin Eisele, Niina Eklund, David M. Evans, Jessica D. Faul, Mary F. Feitosa, Andreas J. Forstner, Ilaria Gandin, Bjarni Gunnarsson, Bjarni V. Halldórsson, Tamara B. Harris, Andrew C. Heath, Lynne J. Hocking, Elizabeth G. Holliday, Georg Homuth, Michael A. Horan, Jouke-Jan Hottenga, Philip L. de Jager, Peter K. Joshi, Astanand Jugessur, Marika A. Kaakinen, Mika Kähönen, Stavroula Kanoni, Liisa Keltigangas-Järvinen, Lambertus A.L.M. Kiemeney, Ivana Kolcic, Seppo Koskinen, Aldi T. Kraja, Martin Kroh, Zoltan Kutalik, Antti Latvala, Lenore J. Launer, Maël P. Lebreton, Douglas F. Levinson, Paul Lichtenstein, Peter Lichtner, David C.M.

Liewald, LifeLines Cohort Study, Anu Loukola, Pamela A. Madden, Reedik Mägi, Tomi Mäki-Opas, Riccardo E. Marioni, Pedro Marques-Vidal, Gerardus A. Meddens, George McMahon, Christa Meisinger, Thomas Meitinger, Yusplitri Milaneschi, Lili Milani, Grant W. Montgomery, Ronny Myhre, Christopher P. Nelson, Dale R. Nyholt, William E.R. Ollier, Aarno Palotie, Lavinia Paternoster, Nancy L. Pedersen, Katja E. Petrovic, David J. Porteous, Katri Räikkönen, Susan M. Ring, Antonietta Robino, Olga Rostapshova, Igor Rudan, Aldo Rustichini, Veikko Salomaa, Alan R. Sanders, Antti-Pekka Sarin, Helena Schmidt, Rodney J. Scott, Blair H. Smith, Jennifer A. Smith, Jan A. Staessen, Elisabeth Steinhagen-Thiessen, Konstantin Strauch, Antonio Terracciano, Martin D. Tobin, Sheila Ulivi, Simona Vaccargiu, Lydia Quaye, Frank J.A. van Rooij, Cristina Venturini, Anna A.E. Vinkhuyzen, Uwe Völker, Henry Völzke, Judith M. Vonk, Diego Vozzi, Johannes Waage, Erin B. Ware, Gonneke Willemsen, John R. Attia, David A. Bennett, Klaus Berger, Lars Bertram, Hans Bisgaard, Dorret I. Boomsma, Ingrid B. Borecki, Ute Bultmann, Christopher F. Chabris, Francesco Cucca, Daniele Cusi, Ian J. Deary, George V. Dedoussis, Cornelia M. van Duijn, Johan G. Eriksson, Barbara Franke, Lude Franke, Paolo Gasparini, Pablo V. Gejman, Christian Gieger, Hans-Jörgen Grabe, Jacob Gratten, Patrick J.F. Groenen, Vilmundur Gudnason, Pim van der Harst, Caroline Hayward, David A. Hinds, Wolfgang Hoffmann, Elina Hyppönen, William G. Iacono, Bo Jacobsson, Marjo-Riitta Järvelin, Karl-Heinz Jöckel, Jaakko Kaprio, Sharon L.R. Kardia, Terho Lehtimäki, Steven F. Lehrer, Patrik K.E. Magnusson, Nicholas G. Martin, Matt McGue, Andres Metspalu, Neil Pendleton, Brenda W.J.H. Penninx, Markus Perola, Nicola Pirastu, Mario Pirastu, Ozren Polasek, Danielle Posthuma, Christine Power, Michael A. Province, Nilesh J. Samani, David Schlessinger, Reinhold Schmidt, Thorkild I.A. Sørensen, Tim D. Spector, Kari Stefansson, Unnur Thorsteinsdottir, A. Roy Thurik, Nicholas J. Timpson, Henning Tiemeier, Joyce Y. Tung, André G. Uitterlinden, Veronique Vitart, Peter Vollenweider, David R. Weir, James F. Wilson, Alan F. Wright, Dalton C. Conley, Robert F. Krueger, George Davey Smith, Albert Hofman, David I. Laibson, Sarah E. Medland, Michelle N. Meyer, Jian Yang, Magnus Johannesson, Peter M. Visscher, Tõnu Esko, Philipp D. Koellinger, David Cesarini, and Daniel J. Benjamin.

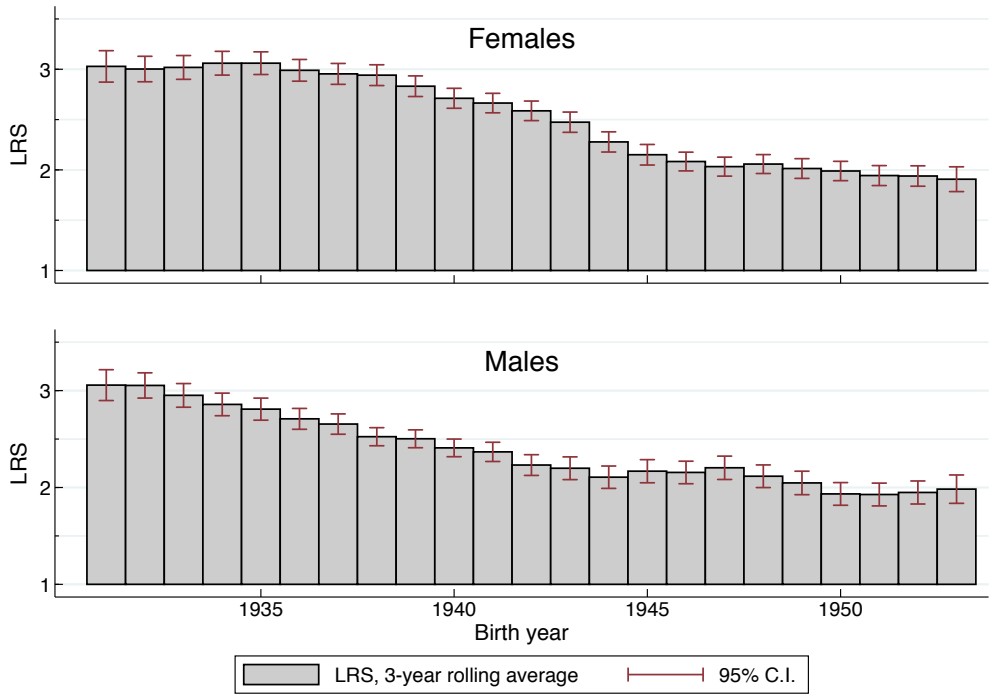

Fig. S1. 3-year rolling average of LRS by birth year, for females and males in the study sample. The rolling average for each year was calculated using the year's data together with data from the previous and following years.

Table S1. Summary statistics for the phenotypic variables

			All individuals (genotyped and not genotyped)			Genotyped individuals only		
Phenotype	**Comment**	**Unit**	*N*	Mean	Std. Dev.	*N*	Mean	Std. Dev.
		Females						
Birth year	---	Year	6,414	1941.21	6.41	3,416	1941.35	6.40
LRS	Number of children ever given birth to	Children	6,414	2.57	1.63	3,416	2.58	1.58
BMI	Mean across waves of female BMI residualized on birth year dummies, plus mean female BMI across all waves and females	BMI points	6,396	27.01	5.73	3,413	27.18	5.56
EA	---	Years of education	6,403	12.98	2.38	3,410	13.12	2.33
HGT	Mean across waves of female HGT residualized on birth year dummies, plus mean female HGT across all waves and females	Centimeters	6,411	162.95	6.36	3,416	163.14	6.25
TC	Indicator variable for a self-reported health problem with high cholesterol in 1992	1 = Yes	4,152	0.25	0.43	2,217	0.25	0.44
Childlessness	Indicator variable for a childless individual (LRS=0)	1 = Childless	6,414	0.10	0.31	3,416	0.10	0.31
		Males						
Birth year	---	Year	5,436	1940.83	6.50	2,571	1941.07	6.50
LRS	Number of children ever fathered	Children	5,435	2.45	1.60	2,571	2.47	1.57
BMI	Mean across waves of male BMI residualized on birth year dummies, plus mean male BMI across all waves and males	BMI points	5,432	27.81	4.53	2,571	27.79	4.49
EA	---	Years of education	5,420	13.27	2.82	2,566	13.49	2.70
HGT	Mean across waves of male HGT residualized on birth year dummies, plus mean male HGT across all waves and males	Centimeters	5,436	178.07	6.63	2,571	178.18	6.48
TC	Indicator variable for a self-reported health problem with high cholesterol in 1992	1 = Yes	3,078	0.24	0.42	1,441	0.27	0.44
Childlessness	Indicator variable for a childless individual (LRS=0)	1 = Childless	5,435	0.12	0.32	2,571	0.11	0.31

This table shows the summary statistics for the study sample. For the main analyses, I use rLRS (relative LRS), not LRS. The phenotypic variable for TC is an indicator for a self-reported health problem with high cholesterol, and not plasma concentrations of total cholesterol as in the GWAS of TC. The HRS does not contain phenotypic variables for GLU, AAM, and SCZ.

Table S2. Estimates from separate regressions of rLRS on each phenotypic variable, for the genotyped individuals only

	Females		Males	
	Coefficient estimate	*N*	Coefficient estimate	*N*
BMI	0.010*** (0.002)	3,413	0.010*** (0.003)	2,571
EA	-0.055*** (0.004)	3,410	-0.020*** (0.005)	2,566
HGT	-0.009*** (0.002)	3,416	-0.001 (0.002)	2,571
TC	-0.002 (0.028)	2,217	-0.020 (0.036)	1,441

This table mirrors Table 1 in the main text, but shows results for the sample of genotyped individuals only—instead of for the sample of all individuals (genotyped and not genotyped)—in the study sample. It shows estimates of the coefficients on the phenotypic variables and their standard errors (in parentheses) from separate regressions of rLRS on each phenotypic variable. Each estimate comes from a different regression and every regression included birth year dummies and HRS-defined cohort dummies. The HRS does not contain phenotypic variables for GLU, AAM, and SCZ.

*$P < 0.10$, **$P < 0.05$, ***$P < 0.01$

Table S3. Estimates from separate regressions of rLRS on the polygenic score of each phenotype, for the study sample and for each cohort separately, and by sex and for females and males together

	Study sample (born 1931-53)	**HRS1 cohort** (born 1931-41)	**HRS2 cohort** (born 1942-47)	**HRS3 cohort** (born 1948-53)
		Females		
Score of BMI	0.006 (0.010)	0.017 (0.014)	0.004 (0.020)	-0.034 (0.025)
Score of EA	-0.033*** (0.010)	-0.038*** (0.014)	-0.044** (0.020)	-0.001 (0.025)
Score of GLU	0.009 (0.010)	0.002 (0.014)	0.029 (0.021)	0.004 (0.024)
Score of HGT	-0.011 (0.014)	-0.025 (0.019)	0.024 (0.028)	-0.019 (0.033)
Score of SCZ	-0.001 (0.011)	-0.012 (0.015)	0.053** (0.022)	-0.035 (0.025)
Score of TC	-0.012 (0.011)	-0.022 (0.014)	-0.005 (0.021)	0.010 (0.025)
Score of AAM	0.018* (0.011)	0.008 (0.014)	0.021 (0.021)	0.037 (0.024)
N	3,416	1,840	811	765
		Males		
Score of BMI	0.016 (0.013)	0.015 (0.016)	-0.025 (0.029)	0.049 (0.032)
Score of EA	-0.031** (0.012)	-0.050*** (0.015)	-0.010 (0.028)	0.012 (0.031)
Score of GLU	-0.013 (0.013)	0.009 (0.016)	-0.023 (0.028)	-0.051* (0.031)
Score of HGT	-0.005 (0.018)	0.016 (0.022)	-0.043 (0.040)	-0.024 (0.042)
Score of SCZ	0.009 (0.013)	0.033** (0.016)	-0.033 (0.032)	-0.009 (0.033)
Score of TC	-0.003 (0.013)	-0.010 (0.016)	0.010 (0.030)	0.022 (0.033)
N	2,571	1,493	506	572
		Females and males together (one person per household)		
Score of BMI	0.018* (0.010)	0.025** (0.012)	0.013 (0.018)	-0.014 (0.022)
Score of EA	-0.041*** (0.009)	-0.047*** (0.012)	-0.039** (0.018)	-0.008 (0.021)
Score of GLU	-0.003 (0.009)	0.011 (0.012)	0.006 (0.018)	-0.037* (0.021)
Score of HGT	-0.015 (0.013)	-0.008 (0.016)	0.001 (0.026)	-0.011 (0.028)
Score of SCZ	0.011 (0.010)	0.011 (0.013)	-0.016 (0.020)	-0.004 (0.022)
Score of TC	-0.008 (0.010)	-0.012 (0.012)	-0.005 (0.019)	0.007 (0.022)
N	4,361	2,647	1,135	1,121

This table mirrors Table 2 in the main text, but also shows the results for each cohort separately as well as for females and males together (the results for the study sample for females and for males separately are the same as those shown in Table 2). The table shows estimates of the coefficients on the polygenic scores and their standard errors (in parentheses) from separate regressions of rLRS on the polygenic score of each phenotype. Each estimate comes from a different regression. All regressions included birth year dummies, HRS-defined cohort dummies, and the top 20 principal components of the genetic relatedness matrix, and all regressions for a cohort and sex had the same number of observations. The regressions for females and males together also included a sex dummy and only included the respondent with the lowest person number (PN, an HRS identifier) in each household. (The results for the score of EA are robust to alternative ways of selecting one respondent per household, but the significant estimates for the score of BMI are not.) The coefficients can be interpreted as directional selection differentials of the scores, expressed in Haldanes—i.e., each coefficient equals the implied change in the score that will occur due to natural selection in one generation, expressed in standard deviations of the score.
*$P < 0.10$, **$P < 0.05$, ***$P < 0.01$

Table S4. Robustness checks for the regressions reported in Table 2

	LRS (instead of rLRS) as the dependent variable	PLINK (instead of LDpred) polygenic scores	Females aged 50-70 and males aged 55-70 only	Study sample together with the HRS0 cohort
		Females		
Score of BMI	0.020 (0.027)	0.002 (0.011)	-0.007 (0.014)	0.002 (0.009)
Score of EA	-0.085*** (0.027)	-0.028*** (0.011)	-0.030** (0.014)	-0.021** (0.009)
Score of GLU	0.019 (0.026)	-0.003 (0.010)	0.014 (0.013)	0.000 (0.009)
Score of HGT	-0.030 (0.036)	-0.011 (0.013)	-0.007 (0.018)	-0.016 (0.013)
Score of SCZ	-0.001 (0.028)	-0.010 (0.012)	0.001 (0.014)	0.001 (0.010)
Score of TC	-0.036 (0.027)	-0.001 (0.010)	-0.009 (0.014)	-0.005 (0.010)
Score of AAM	0.045* (0.027)	0.022** (0.010)	0.021 (0.014)	0.013 (0.010)
N	3,416	3,416	2,065	4,182
		Males		
Score of BMI	0.047 (0.031)	0.021 (0.013)	0.003 (0.021)	0.013 (0.011)
Score of EA	-0.079** (0.031)	-0.031** (0.013)	-0.074*** (0.020)	-0.021* (0.011)
Score of GLU	-0.030 (0.030)	-0.014 (0.013)	-0.019 (0.020)	-0.003 (0.011)
Score of HGT	0.007 (0.040)	0.001 (0.017)	-0.026 (0.029)	0.007 (0.016)
Score of SCZ	0.022 (0.036)	0.004 (0.015)	0.018 (0.022)	0.015 (0.012)
Score of TC	-0.009 (0.031)	-0.001 (0.013)	-0.016 (0.021)	-0.008 (0.012)
N	2,571	2,571	959	3,173

This table mirrors Table 2 in the main text, but shows the results for alternative specifications and samples. Column 2 shows the estimates and standard errors (in parentheses) from separate regressions of LRS (instead of rLRS) on the polygenic score of each phenotype for the study sample; column 3 shows the estimates and standard errors from separate regressions of rLRS on the polygenic score of each phenotype constructed with PLINK (instead of LDpred) for the study sample; column 4 shows estimates and standard errors from separate regressions of rLRS on the polygenic score of each phenotype, but only for individuals in the study sample who were aged no more than 70 in 2008 (the last year for genotyping) and at least 50 years old when asked the number of children they ever gave birth to (for females) or at least 55 years old when asked the number of children they ever fathered (for males); column 5 shows estimates and standard errors from separate regressions of rLRS on the polygenic score of each phenotype, but for the study sample and the HRS0 cohort together. All regressions included birth year dummies, HRS-defined cohort dummies, and the top 20 principal components of the genetic relatedness matrix, and all regressions for each specification and sex had the same number of observations. The coefficients in column 2 can be interpreted as the effects of one-standard deviation increases in the scores on the number of children ever given birth to (for females) or the number of children ever fathered (for males). The coefficients in columns 3-5 can be interpreted as directional selection differentials of the scores, expressed in Haldanes—i.e., each coefficient equals the implied change in the score that will occur due to natural selection in one generation, expressed in standard deviations of the score.
*$P < 0.10$, **$P < 0.05$, ***$P < 0.01$

Table S5. Estimates from the quadratic regression from Lande and Arnold (16) of rLRS on all the polygenic scores of the different phenotypes, all the squared polygenic scores, and all their interactions

	Females	Males
Score of BMI	0.003 (0.011)	0.012 (0.013)
Score of EA	-0.035*** (0.011)	-0.029** (0.013)
Score of GLU	0.010 (0.010)	-0.012 (0.013)
Score of HGT	-0.008 (0.014)	0.004 (0.018)
Score of SCZ	-0.001 (0.011)	0.006 (0.014)
Score of TC	-0.013 (0.011)	-0.004 (0.013)
Score of AAM	0.022** (0.011)	---
Squared score of BMI	-0.006 (0.008)	-0.001 (0.010)
Squared score of EA	-0.012 (0.008)	-0.004 (0.009)
Squared score of GLU	-0.007 (0.007)	-0.002 (0.009)
Squared score of HGT	-0.009 (0.007)	0.000 (0.009)
Squared score of SCZ	0.010 (0.007)	0.022** (0.009)
Squared score of TC	-0.007 (0.007)	0.000 (0.009)
Squared score of AAM	-0.004 (0.008)	---
Score of BMI x score of EA	0.005 (0.011)	0.004 (0.013)
Score of BMI x score of GLU	0.002 (0.011)	-0.011 (0.013)
Score of BMI x score of HGT	-0.008 (0.012)	-0.015 (0.013)
Score of BMI x score of SCZ	0.017 (0.011)	0.015 (0.013)
Score of BMI x score of TC	0.009 (0.011)	0.016 (0.014)
Score of BMI x score of AAM	-0.012 (0.011)	---
Score of EA x score of GLU	-0.006 (0.011)	0.017 (0.012)
Score of EA x score of HGT	-0.003 (0.011)	0.001 (0.013)
Score of EA x score of SCZ	0.001 (0.011)	-0.032** (0.013)
Score of EA x score of TC	-0.010 (0.011)	0.000 (0.013)
Score of EA x score of AAM	0.003 (0.011)	---
Score of GLU x score of HGT	-0.020* (0.011)	-0.002 (0.013)
Score of GLU x score of SCZ	-0.008 (0.011)	0.008 (0.013)
Score of GLU x score of TC	-0.007 (0.010)	0.013 (0.013)
Score of GLU x score of AAM	-0.015 (0.011)	---
Score of HGT x score of SCZ	0.004 (0.011)	0.022 (0.014)
Score of HGT x score of TC	0.000 (0.011)	-0.010 (0.013)
Score of HGT x score of AAM	0.003 (0.012)	---
Score of SCZ x score of TC	-0.028*** (0.010)	-0.004 (0.013)
Score of SCZ x score of AAM	-0.019* (0.011)	---
Score of TC x score of AAM	0.002 (0.011)	---
N	3,416	2,571

This table shows estimates (and their standard errors, in parentheses) from the quadratic regression from Lande and Arnold (16) of rLRS on all the polygenic scores of the different phenotypes, all the squared polygenic scores, and all their interactions, for the study sample. (I thus treat the scores themselves—as opposed to the phenotypes, as is usual—as the characters of interest in Lande and Arnold's quadratic framework.) The estimates for females come from one single regression and the estimates for males come from another single regression. Each regression included birth year dummies, HRS-defined cohort dummies, and the top 20 principal components of the genetic relatedness matrix.

*$P < 0.10$, **$P < 0.05$, ***$P < 0.01$

Table S6. Description of cohorts

	HRS0	HRS1	HRS2	HRS3
Included in the study sample?	No	Yes	Yes	Yes
Years of birth	1924-30	1931-41	1942-47	1948-53
Females				
N (genotyped and not genotyped)	1,736	3,505	1,528	1,381
Age asked #children, 1st-99th percentiles	65-77	55-73	49-66	45-62
Age when genotyped (in 2006-08)	76-84	65-77	59-66	53-60
Fraction survived to 2008	0.69	0.82	0.90	0.96
Fraction asked to be genotyped (in 2006-08)	0.59	0.70	0.74	0.73
Fraction consented to be genotyped (in 2006-08)	0.51	0.61	0.62	0.63
Fraction in sample of genotyped individuals	0.44	0.52	0.53	0.55
Males				
N (genotyped plus not genotyped)	1,562	3,231	1,068	1,137
Age asked #children, 1st-99th percentiles	66-80	55-75	50-68	50-62
Age when genotyped (in 2006-08)	76-84	65-77	59-66	53-60
Fraction survived to 2008	0.60	0.76	0.86	0.97
Fraction asked to be genotyped (in 2006-08)	0.52	0.63	0.65	0.70
Fraction consented to be genotyped (in 2006-08)	0.47	0.54	0.55	0.58
Fraction in sample of genotyped individuals	0.39	0.46	0.47	0.50

For each cohort, I include all individuals of European ancestry born in the cohort's years of birth, who were enrolled in the HRS and asked the number of children they ever gave birth to or fathered in 2008 or earlier, and who were at least 45 years old when asked the number of children they ever gave birth to (for females) or at least 50 years old when asked the number of children they ever fathered (for males). "Fraction in sample of genotyped individuals" indicates the fraction of individuals who are in the sample of unrelated genotyped individuals who passed the quality control filters described in *Supporting Information* (the sample used in the analyses with the genotyped individuals). The HRS0 cohort is not included in the study sample because of the high mortality among its members by 2008 (the last year when individuals were genotyped) and because of evidence of selection bias for the genotyped individuals in that cohort. The main results are robust to the inclusion of that cohort.

Table S7. Summary information on the polygenic scores of the different phenotypes

	Optimal Gaussian mixture weight for LDpred	LDpred window size (# of SNPs)	Number of SNPs used	Assumed GWAS sample size	GWAS article	Source for GWAS summary statistics	R^2, previously reported	Incremental R^2 (Std. Error), est. in HRS
Score of BMI	0.1	170	505,254	232,186	Ref. (19)	http://www.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files#GWAS_Anthropometric_2015_BMI	0.06	0.089*** (0.007)
Score of EA	0.1	180	544,493	386,098	Ref. (20)	SSGAC (Summary statistics from a meta-analysis excluding the HRS were used)	0.039	0.074*** (0.006)
Score of GLU	0.03	10	22,894	120,000	Ref. (21)	www.magicinvestigators.org/downloads/	---	---
Score of HGT	1	170	510,411	243,630	Ref. (22)	http://www.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files#GWAS_Anthropometric_2014_Height	0.17	0.174*** (0.009)
Score of SCZ	0.3	180	544,225	75,000	Ref. (23)	http://www.med.unc.edu/pgc/files/resultfiles	0.070	---
Score of TC	0.3	175	530,012	92,793	Ref. (24)	www.broadinstitute.org/mpg/pubs/lipids2010/	---	0.012*** (0.004)
Score of AAM	0.3	170	506,120	120,000	Ref. (25)	http://www.reprogen.org/data_download.html	---	---

For every phenotype, the polygenic score was constructed using LDpred (31), using the individuals' genotyped SNPs that passed quality control filters and overlapped with the SNPs in the phenotype's summary statistics file. The optimal Gaussian mixture weights (the assumed fractions of causal markers) were selected to maximize each score's R^2 with respect to the corresponding phenotype or to maximize the correlations between each score and known correlates of the corresponding phenotype. As recommended, LDpred windows approximately equal to the number of used SNPs divided by 3,000 were used. The assumed GWAS sample sizes are the assumed sample sizes for LDpred, based on mean sample sizes across the used SNPs (when SNP-level sample sizes are reported) or based on the reported GWAS sample sizes (slighlty reduced to account for missing observations). The previously reported R^2 and the incremental R^2 estimated in the HRS are the numerical values of the results presented in Fig. 1 in the main text (with standard errors instead of 95% C.I.). The incremental R^2 estimated in the HRS for the score of EA is substantially higher than the previously reported R^2 because the score I used in the HRS is based on summary statistics from a meta-analysis that includes one additional large cohort (the UK Biobank).

Table S8. Estimates for the HRS0 cohort (born 1924-30)

Panel A: regressions of rLRS on the phenotypic variables				
	All individuals (genotyped and not genotyped)		Genotyped individuals only	
Females				
	Coefficient estimate	*N*	Coefficient estimate	*N*
BMI	0.015*** (0.003)	1,731	0.008* (0.005)	766
EA	-0.013** (0.006)	1,736	0.008 (0.009)	766
---	---	---	---	---
HGT	-0.000 (0.003)	1,735	-0.001 (0.004)	766
---	---	---	---	---
TC	-0.024 (0.122)	93	---	---
---	---	---	---	---
Males				
	Coefficient estimate	*N*	Coefficient estimate	*N*
BMI	0.003 (0.004)	1,562	0.007 (0.006)	602
EA	-0.003 (0.005)	1,561	-0.001 (0.008)	602
---	---	---	---	---
HGT	-0.003 (0.002)	1,562	-0.004 (0.004)	602
---	---	---	---	---
TC	-0.067 (0.059)	590	-0.102 (0.091)	220

Panel B: regressions of rLRS on the scores		
	Genotyped individuals only	
Females		
	Coefficient estimate	*N*
Score of BMI	-0.020 (0.023)	766
Score of EA	0.031 (0.023)	766
Score of GLU	-0.044* (0.023)	766
Score of HGT	-0.046 (0.031)	766
Score of SCZ	0.006 (0.025)	766
Score of TC	0.020 (0.023)	766
Score of AAM	-0.010 (0.024)	766
Males		
	Coefficient estimate	*N*
Score of BMI	0.002 (0.026)	602
Score of EA	0.029 (0.026)	602
Score of GLU	0.044* (0.026)	602
Score of HGT	0.054 (0.034)	602
Score of SCZ	0.037 (0.026)	602
Score of TC	-0.026 (0.025)	602

Panel A mirrors Table 1 in the main text and Table S2, but shows results for the HRS0 cohort. It shows estimates of the coefficients on the phenotypic variables and their standard errors (in parentheses) from separate regressions of rLRS on each phenotypic variable. Each estimate comes from a different regression and every regression included birth year dummies and HRS-defined cohort dummies. The HRS does not contain phenotypic variables for GLU, AAM, and SCZ; the results are not reported for TC for genotyped females because there are only 36 genotyped females with TC data in the HRS0 cohort. **Panel B** mirrors Table 2 in the main text, but shows results for the HRS0 cohort. The table shows estimates of the coefficients on the polygenic scores and their standard errors (in parentheses) from separate regressions of rLRS on the polygenic score of each phenotype. Each estimate comes from a different regression. All regressions included birth year dummies, HRS-defined cohort dummies, and the top 20 principal components of the genetic relatedness matrix. The coefficients can be interpreted as directional selection differentials of the scores, expressed in Haldanes—i.e., each coefficient equals the implied change in the score that will occur due to natural selection in one generation, expressed in standard deviations of the score.
*$P < 0.10$, **$P < 0.05$, ***$P < 0.01$