www.sciencemag.org/cgi/content/full/science.aag0776/DC1

Supplementary Materials for

Detection of human adaptation during the past 2000 years

Yair Field,* Evan A Boyle, Natalie Telis, Ziyue Gao, Kyle J. Gaulton, David Golan, Loic Yengo, Ghislain Rocheleau, Philippe Froguel, Mark I. McCarthy, Jonathan K. Pritchard*

*Corresponding author. Email: yairf@stanford.edu (Y.F.); pritch@stanford.edu (J.K.P.)

Published 13 October 2016 on *Science* First Release
DOI: 10.1126/science.aag0776

This PDF file includes:

Other Supporting Online Material for this manuscript includes the following:
(available at www.sciencemag.org/cgi/content/full/science.aag0776/DC1)

Table of contents:

List of Figures and Tables:

Supplementary Methods

The Singleton Density Score (SDS) – theoretical motivation: The history of a sample of haplotypes going back to their common ancestor can be described by a branching structure known as the coalescent, that represents the genealogy of the sample at a particular site (*26*). While we cannot observe the sample genealogy directly, population genetics theory tells us how, for a random sample, the expected times of coalescence in the genealogy depend on the underlying population size history. In turn, the distribution of mutations depends on branch lengths: specifically, mutations are modeled as Poisson-distributed with rate equal to the product of mutation rate times branch length (*26*). This basic connection between distribution of sequence variants in a sample and the historical demographic model of the population underlies much of modern theory and methods in population genetics.

Our goal in this paper is to detect very recent allele frequency changes at SNPs. The key idea underlying SDS is that recent frequency changes generate differences in the distributions of coalescence times on the two allelic backgrounds. We designed SDS to detect differences between the derived and ancestral alleles with respect to their most recent coalescent times – namely, the external branches of the genealogies (also known as "tip branches"). Under a constant population size model with effective (diploid) population size N_e, the expected length of a tip branch in a sample of $2n$ haplotypes equals $\frac{2N_e}{n}$ (*26*). Consider a test SNP with sample derived allele frequency (DAF) of f, and assume for the purpose of illustration that this sample frequency equals the population frequency, and that the population frequency hasn't changed over history. Then under this simplistic neutral model, the tip-branches of the derived and ancestral alleles have the same expected length: $2\frac{N_e f}{nf} = 2\frac{N_e(1-f)}{n(1-f)} = 2\frac{N_e}{n}$. However, if one allele has been increasing in frequency, then its expected tip branch should be shorter than the neutral expectation, as going back in time the effective allele population size is smaller, and so the probability to coalesce increases. Likewise, for the alternative allele, whose frequency has been decreasing, we expect longer tip branch lengths. This is illustrated in

Figures 1 and S1 by a more realistic simulation (e.g., accounting for population growth, drift, etc. See simulation details in the Methods section: *Simulations for the illustration figure*).

SDS detects recent changes in allele frequency by inferring and contrasting the expected tip-branch lengths of the derived and ancestral alleles. We do so by modeling the local distribution of singletons, those variants that originated along tip-branches. The effective timescale of SDS is therefore roughly the average tip-branch length of neutral genealogies (i.e., $\sim\frac{2N_e}{n}$ in a constant population size), which gets shorter for larger samples. In comparison, related methods such as the Extended Haplotype Homozygosity (EHH) (*27*) and the integrated Haplotype Score (iHS) (*5*) have a much older timescale that does not depend on the sample size. These methods could be conceptualized as detecting recent changes in frequency by contrasting the average pairwise coalescent times of the derived and ancestral alleles, whose expected times are on the order of N_e.

The Singleton Density Score Model: Consider a bi-allelic test SNP. We want to evaluate whether there is a difference in average tip lengths between the genealogies of the two alleles. For each diploid individual $i = 1 \ldots n$, we compute the distance d_i between the nearest upstream singleton and nearest downstream singleton, relative to the test SNP. Note that singletons are defined with respect to the entire sample of n individuals: i.e., a singleton is a variant that is seen exactly once in the sample of n. Further, since current sequencing data typically do not allow phasing of singletons, we compute d_i using the distances to the nearest singletons in the diploid genotype without regard to phase.

The SDS score is computed from a maximum likelihood estimate (MLE) of the log-ratio of mean tip branch lengths on the two allelic backgrounds, standardized to have mean 0 and variance 1. To compute this likelihood, we use the following approximate generative model for each distance d_i.

We assume that tip branches of allele h (where h=derived or ancestral) are drawn independently and identically from an allele-specific distribution of tip-branch lengths. These distributions will be denoted φ_h. Thus, at each site, each individual has two tip lengths drawn from the appropriate distributions, given their genotype.

Then, ignoring recombination, we further assume that for each haplotype, given that the tip branch of the test SNP is of length t and the local mutation rate is μ, a Poisson process with rate $t\mu$ describes the occurrence of singletons around the test SNP. It follows that the distance from the test SNP to the nearest downstream (or equivalently upstream) singleton over the haplotype is an *Exponential* random variable with rate $t\mu$. In turn, for an individual with test-SNP tip-lengths t_1 and t_2, the distance to the first downstream (or equivalently upstream) singleton in the diploid, without regard to phase, is an *Exponential* random variable with rate $(t_1 + t_2)\mu$. Therefore, given the tip lengths for the test SNP in individual i, the distance d_i between the nearest upstream singleton and the nearest downstream singleton is a sum of two *Exponential* random variables with equal rates, and is given by a *Gamma* distribution with shape parameter 2 and rate parameter $(t_1 + t_2)\mu$.

Integrating over the uncertainty of the two tip-lengths of an individual diploid, the likelihood of d_i is given by:

$$L(d_i; \theta_{ancestral}, \theta_{derived}, \mu)$$
$$= \iint_{t_1,t_2} \mathcal{G}(d_i; 2, (t_1 + t_2)\mu) \cdot \varphi_{h[i,1]}(t_1; \theta_{h[i,1]}) \cdot \varphi_{h[i,2]}(t_2; \theta_{h[i,2]})\, dt_1\, dt_2$$

where $\theta_{ancestral}$ and $\theta_{derived}$ are the parameters of the tip length distributions, $\varphi_{ancestral}(t)$ and $\varphi_{derived}(t)$; $\mathcal{G}(x; \alpha, \beta)$ is the *Gamma* distribution with shape and rate parameters α and β; and $h[i,1]$ and $h[i,2]$ are the ancestral/derived annotations for the two test-SNP haplotypes for individual i.

To account for imperfect mapping of singletons, and the variability in sequencing depth among individuals, we further assume that for each individual i, singletons are observed with probability p_i. Events that follow a *Poisson* process with rate λ and are then

observed with probability p follow a *Poisson* process with rate λp. Thus, the likelihood of d_i is given by:

$$L(d_i; \theta_{ancestral}, \theta_{derived}, \mu, p_i)$$
$$= \iint_{t_1,t_2} \mathcal{G}(d_i; 2, (t_1 + t_2)\mu p_i) \cdot \varphi_{h[i,1]}(t_1; \theta_{h[i,1]}) \cdot \varphi_{h[i,2]}(t_2; \theta_{h[i,2]})\, dt_1\, dt_2$$

As we motivate later on in this section, we choose to model the tip length distributions of the ancestral and derived alleles by *Gamma* distributions:

$$\varphi_{ancestral}(t; \theta_{ancestral}) = \mathcal{G}(t; \alpha_{ancestral}, \beta_{ancestral})$$
$$\varphi_{derived}(t; \theta_{derived}) = \mathcal{G}(t; \alpha_{derived}, \beta_{derived})$$

The likelihood then becomes:

$$L(d_i; \alpha_{ancestral}, \beta_{ancestral}, \alpha_{derived}, \beta_{derived}, \mu, p_i)$$
$$= \iint_{t_1,t_2} \left[\frac{\left((t_1 + t_2)\mu p_i\right)^2}{\Gamma(2)} d_i e^{-((t_1+t_2)\mu p_i)d_i} \right] \cdot \left[\frac{{\beta_{h[i,1]}}^{\alpha_{h[i,1]}}}{\Gamma(\alpha_{h[i,1]})} {t_1}^{\alpha_{h[i,1]}-1} e^{-\beta_{h[i,1]}t_1} \right]$$
$$\cdot \left[\frac{{\beta_{h[i,2]}}^{\alpha_{h[i,2]}}}{\Gamma(\alpha_{h[i,2]})} {t_2}^{\alpha_{h[i,2]}-1} e^{-\beta_{h[i,2]}t_2} \right] dt_1\, dt_2$$

We are ultimately interested to learn the ratio between the mean tip lengths of the ancestral and derived alleles, which is given by $\frac{\hat{t}_{ancestral}}{\hat{t}_{derived}} = \frac{\alpha_{ancestral}/\beta_{ancestral}}{\alpha_{derived}/\beta_{derived}}$. This ratio of tip lengths is not dependent on the local mutation rate μ, but do we need to know – or explicitly learn – this parameter? The answer is no. We only care about $\frac{\beta_{ancestral}}{\mu}$ and $\frac{\beta_{derived}}{\mu}$, because both the ratio of mean tip lengths and the likelihood are dependent only on these compounded parameters, and not on $\beta_{ancestral}$, $\beta_{derived}$ and μ individually. For the ratio of mean tip lengths this is immediate: $\frac{\hat{t}_{ancestral}}{\hat{t}_{derived}} = \frac{\alpha_{ancestral}/\left(\frac{\beta_{ancestral}}{\mu}\right)}{\alpha_{derived}/\left(\frac{\beta_{derived}}{\mu}\right)}$. To see it for the likelihood, make a change of variables $T_1 = \mu t_1$ and $T_2 = \mu t_2$ to get:

$$L(d_i; \alpha_{ancestral}, \beta_{ancestral}, \alpha_{derived}, \beta_{derived}, \mu, p_i)$$

$$= \iint_{T_1,T_2} \left[\frac{\left((T_1+T_2)p_i\right)^2}{\Gamma(2)} d_i e^{-((T_1+T_2)p_i)d_i} \right] \cdot \left[\frac{\left(\frac{\beta_{h[i,1]}}{\mu}\right)^{a_{h[i,1]}}}{\Gamma\left(\alpha_{h[i,1]}\right)} {T_1}^{\alpha_{h[i,1]}-1} e^{-\left(\frac{\beta_{h[i,1]}}{\mu}\right)T_1} \right]$$

$$\cdot \left[\frac{\left(\frac{\beta_{h[i,2]}}{\mu}\right)^{a_{h[i,2]2}}}{\Gamma\left(\alpha_{h[i,2]}\right)} {T_2}^{\alpha_{h[i,2]}-1} e^{-\left(\frac{\beta_{h[i,2]}}{\mu}\right)T_2} \right] dT_1\, dT_2$$

$$= L\left(d_i; \alpha_{ancestral}, \frac{\beta_{ancestral}}{\mu}, \alpha_{derived}, \frac{\beta_{derived}}{\mu}, 1, p_i\right)$$

We assume that we know *a priori* the individual-specific probabilities to observe singletons due to incomplete sequencing depth, p_i, up to an unknown constant $c > 0$. We refer to $o_i = p_i \cdot c$ as the "singleton observability" constants, and we present how we estimated them for the UK10K data later on (in Methods section: *Computing SDS for UK10K data*). As we now show, in the maximum likelihood framework we can account for the p_i's by simply scaling the observed distances between singletons by o_i – that is, by working with $d_i \cdot o_i$ instead of d_i. This follows from the fact that for any $c > 0$:

$$L\left(d_i; \alpha_{ancestral}, \frac{\beta_{ancestral}}{\mu}, \alpha_{derived}, \frac{\beta_{derived}}{\mu}, 1, p_i\right)$$

$$= p_i c \iint_{t_1,t_2} \left[\frac{\left((t_1+t_2)\frac{1}{c}\right)^2}{\Gamma(2)} (d_i p_i c) e^{-(t_1+t_2)\frac{1}{c}(d_i p_i c)} \right] \cdot \left[\frac{\left(\frac{\beta_{h[i,1]}}{\mu}\right)^{a_{h[i,1]}}}{\Gamma\left(\alpha_{h[i,1]}\right)} {t_1}^{\alpha_{h[i,1]}-1} e^{-\left(\frac{\beta_{h[i,1]}}{\mu}\right)t_1} \right]$$

$$\cdot \left[\frac{\left(\frac{\beta_{h[i,2]}}{\mu}\right)^{a_{h[i,2]2}}}{\Gamma\left(\alpha_{h[i,2]}\right)} {t_2}^{\alpha_{h[i,2]}-1} e^{-\left(\frac{\beta_{h[i,2]}}{\mu}\right)t_2} \right] dt_1\, dt_2$$

$$= o_i \cdot L\left(d_i o_i; \alpha_{ancestral}, \frac{\beta_{ancestral}}{\mu}, \alpha_{derived}, \frac{\beta_{derived}}{\mu}, \frac{1}{c}, 1\right)$$

$$= o_i \cdot L\left(d_i o_i; \alpha_{ancestral}, \frac{\beta_{ancestral}}{c/\mu}, \alpha_{derived}, \frac{\beta_{derived}}{c/\mu}, 1, 1\right)$$

Since the o_i term that scales the Likelihood above is a constant, it does not affect the maximum likelihood estimation and we can ignore it. Thus without loss of generality we think of the observations d_i as already being scaled by o_i. We further simplify notation by using ${\beta'}_{ancestral} \stackrel{\text{def}}{=} \frac{\beta_{ancestral}}{c/\mu}$ and ${\beta'}_{derived} \stackrel{\text{def}}{=} \frac{\beta_{derived}}{c/\mu}$. The likelihood is then given by:

$$L\left(d_i; \alpha_{ancestral}, \beta'_{ancestral}, \alpha_{derived}, \beta'_{derived}\right)$$
$$= \left(\frac{\beta'_{h[i,1]}}{d_i + \beta'_{h[i,1]}}\right)^{\alpha_{h[i,1]}} \cdot \left(\frac{\beta'_{h[i,2]}}{d_i + \beta'_{h[i,2]}}\right)^{\alpha_{h[i,2]}} \cdot d_i$$
$$\cdot \left(\frac{\alpha_{h[i,1]}\left(\alpha_{h[i,1]} + 1\right)}{\left(d_i + \beta'_{h[i,1]}\right)^2} + 2\frac{\alpha_{h[i,1]}\alpha_{h[i,2]}}{\left(d_i + \beta'_{h[i,1]}\right)\left(d_i + \beta'_{h[i,2]}\right)} \right.$$
$$\left. + \frac{\alpha_{h[i,2]}\left(\alpha_{h[i,2]} + 1\right)}{\left(d_i + \beta'_{h[i,2]}\right)^2} \right)$$

The derivation of the above closed-form solution to the likelihood is somewhat long and we skip it here. We note that the basic idea in solving the double integral is the observation that integrals of the form $\int_{x=0}^{\infty} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx$ equal 1 because $\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$ is the probability density function of the *Gamma* distribution.

We now further simplify the model by assuming that we know *a priori* the shape parameters $\alpha_{ancestral}$ and $\alpha_{derived}$. Notice that for a *Gamma* distribution $\mathcal{G}(x; \alpha, \beta)$, the mean and variance are given by $E = \frac{\alpha}{\beta}$ and $V = \frac{\alpha}{\beta^2}$, so the shape parameter describes the variance as function of the mean: $\alpha = \frac{E^2}{V}$. The tip length distribution depends in general on the demographic model, the exact allele frequency trajectory, and the allele sample size. We therefore chose to fix the Gamma-shape parameters to best match the shapes of tip-length distributions that we simulate, for a given demographic model, sample size and present-day allele frequency, under a model of neutral drift. The estimates for each shape parameter were taken to be the average of squared mean over variance, over the simulated tip-length distributions. This is described in more details below (in Methods section: *Computing SDS for UK10K data*).

We can now better motivate the choice for a *Gamma* distribution to model the allele-specific tip length distribution. First, we found empirically that simulated genealogies under a model of neutral drift are reasonably described by a *Gamma* distribution. Second, the choice of *Gamma* distribution allowed us to solve the

double integral in the likelihood into an explicit form, thus simplifying the maximum likelihood optimization. Third, in our model, the two parameters of the *Gamma* distribution split nicely into two components – one parameter (rate) absorbs all the locus-specific and individual-specific effects, including the regional mutation rate, and the probability to detect singletons due to global sequencing depth as well as due to any other local effect (e.g., locus-specific challenges in mapping sequencing reads due to specific base composition or repetitive nature of the sequence); whereas the other parameter (shape) is more global in nature, which motivates its pre-fixation based on neutral simulations. Thus in sum, this naïve approximation is nevertheless a mathematically convenient and effective choice for the tip-length distributions. We note that a somewhat similar use of a *Gamma* distribution to describe a tip-length distribution was previously presented (*28*).

Therefore, the model depends on only two parameters, $\beta'_{ancestral}$ and $\beta'_{derived}$. The objective of SDS, the ratio of means of the tip length distributions for the ancestral and derived alleles, is then given by: $\frac{\hat{t}_{ancestral}}{\hat{t}_{derived}} = \frac{\alpha_{ancestral}/\beta'_{ancestral}}{\alpha_{derived}/\beta'_{derived}}$.

We thus define:

$$\left(\beta'^{*}_{ancestral}, \beta'^{*}_{derived}\right) \stackrel{\text{def}}{=} \underset{\beta'_{ancestral},\beta'_{derived}}{\operatorname{argmax}} \prod_{i=1}^{n} L\left(d_i; \beta'_{ancestral}, \beta'_{derived}\right)$$

We infer these maximum likelihood estimates (MLEs) using a standard gradient descent optimization, implemented in R using the *nlm* function. Using both simulations and real data we verified that the optimization indeed reaches optimal solutions by comparing optimization results with those obtained by exhaustive grid searches (which is easy to visualize given that there are only two parameters). We define the "raw SDS" statistic to be the log-ratio of mean tip lengths as obtained by the MLE estimates above:

$$raw\ SDS \stackrel{\text{def}}{=} \log\left(\frac{\alpha_{ancestral}/\beta'^{*}_{ancestral}}{\alpha_{derived}/\beta'^{*}_{derived}}\right) = \log\left(\frac{\hat{t}^{*}_{ancestral}}{\hat{t}^{*}_{derived}}\right)$$

Notice that $raw\ SDS$ is also the direct MLE of our model. This is because in the Maximum Likelihood estimation framework, it is mathematically equivalent to re-parameterize the model directly in terms of this quantity of interest, e.g., by $\log\left(\frac{\alpha_{ancestral}/\beta'_{ancestral}}{\alpha_{derived}/\beta'_{derived}}\right)$ and $\alpha_{derived}/\beta'_{derived}$, and report raw SDS as:

$$raw\ SDS \stackrel{\text{def}}{=} \underset{\log\left(\frac{\alpha_{ancestral}/\beta'_{ancestral}}{\alpha_{derived}/\beta'_{derived}}\right)}{\text{argmax}} \prod_{i=1}^{n} L\left(d_i; \beta'_{ancestral}, \beta'_{derived}\right)$$

The Singleton Density Score (SDS) that we finally report is a standardized version of the raw SDS, relative to the genome wide predictions with similar DAF. That is, we define a partition of all SNPs genome-wide into derived allele frequency bins (in practice for the UK10K analysis we defined bins of 1% DAF; see details in Method section: *Computing SDS for UK10K data*). Then, assuming a test SNP s_j with derived allele frequency f belongs to bin $B(f)$, we define the standardized SDS to be:

$$SDS[s_j] \stackrel{\text{def}}{=} \frac{raw\ SDS[s_j] - \mu_{B(f)}}{\sigma_{B(f)}}$$

where $\mu_{B(f)}$ and $\sigma_{B(f)}$ are the mean and standard deviation estimates of raw SDS over all SNPs s_k that belong in bin $B(f)$. That is,

$$\mu_{B(f)} = \frac{1}{|B(f)|} \sum_{s_k \in B(f)} raw\ SDS[s_k]$$

$$\sigma_{B(f)} = \sqrt{\frac{1}{|B(f)| - 1} \sum_{s_k \in B(f)} \left(raw\ SDS[s_k] - \mu_{B(f)}\right)^2}$$

In summary, large positive SDS corresponds to a lower density of singletons associated with the derived allele and hence that derived allele tip-branches are much shorter than the tip-branches of the ancestral allele, and thus implies a recent increase in frequency of the derived allele.

Modeling considerations and possible future extensions: One key simplification of the SDS model is that we ignore the effect of recombination. The main effect of recombination is that it scrambles distant singletons across allelic backgrounds. This effect will tend to bias true signal back toward the null hypothesis whenever the nearest recombination event (on the tip branch) is closer than the nearest mutation, but should not create false positives.

A practical issue with current data is that detection of singletons may be incomplete, especially in low-coverage data sets such as UK10K. Again, we regard this as an effect that reduces our power to detect true signals; however in most cases we do not expect singleton detection to differ systematically by genotype at the test SNP, and thus this should not systematically create false positives. (See below for a discussion of this issue in the context of the MHC region.) More generally, in SDS the two allelic backgrounds provide a natural control against each other for a variety of possible shortcomings in the data and models. Another strength of SDS compared to earlier methods such as iHS, is that the information comes from a large number of nearly independent tip branches, thus explaining why the statistic is well-behaved under the null.

In principle, modifiers of mutation rate could also generate significant SDS scores (*29*). At the present time, no common alleles are known that modify mutation rates, and we anticipate that such signals are likely to be less numerous than signals due to positive selection.

There are numerous possible extensions to SDS. First, it would be of value to consider alleles at different frequencies – e.g., doubletons, tripletons, etc. Consideration of variants in different frequency classes will make it possible to explore more detailed trajectories of frequency over time. These will require substantial modifications of the current statistical model. Second, it would be valuable to infer recombination break points on each tip. The singleton distances become uninformative past the nearest recombination event, and these censoring points could be incorporated into the likelihood. Moreover, the recombination distances are themselves informative about

branch lengths. Recombination-based analysis could allow extension of SDS to genotype or exome data, neither of which contain enough singleton information for application of the current SDS method. Third, SDS might be combined with allele frequencies from other populations to gain further power.

We also investigated other metrics for quantifying the number of singletons, such as the number that occur within a window of fixed size. However we found that such statistics were heavily impacted by outliers. Further, in humans, recombination and mutation occur at roughly similar rates per base pair, this implies that on average only ~1 singleton per individual is truly informative.

Simulating genealogies: To study the specificity of SDS to recent history, for a sample comparable to the UK10K data we analyzed here, we performed simulations of samples of n=3,000 individuals (2n=6,000 haplotypes) using a recent demographic model of a European population (*14*). To this end, we examined first the average tip-branch length as a function of sample size by this demographic model, using 1,000 simulated genealogies per sample size, for samples ranging from 2n=100 to 2n=100,000 haplotypes (Fig. 2A). We used the coalescent simulator *ms* (*30*), with a population size model of the European population from (*14*). The average tip-branch length (in generations) decreases within this range from 360±54 (2n=100) to 13.3±0.04 (2n=100,000). For 2n=6,000 haplotypes, comparable to the size of the UK10K data sample, the average tip is 74.5±0.68 generations. In the Methods section *Estimating the mean tip-branch length in the UK10K data* we consider additional demographic models and provide a more thorough description of the uncertainty about the actual mean tip length in the UK10K sample.

Simulation of power and specificity of SDS to recent history: To study the power of SDS with n=3,000 individuals and its specificity to selection within the past 100 generations (Fig. 2B), we simulated random samples of sequences under different models of selection using a backward coalescent approach. Given the present day derived allele frequency (DAF), and models of the history of the population size and relative fitness of the selected allele, we first simulated the DAF trajectory backward in time using

simuPOP (*31*). To this end, we used the population size model of a European population from (*14*), and considered one of three additive selection models (i.e., with dominance parameter h=0.5): (1) constant selection throughout history, (2) constant selection starting 100 generations ago, and (3) constant selection ending 100 generations ago. For each model, we examined varying selection coefficients, ranging from very low s=0.001 to extremely high s=0.10. Main text results all correspond to present-day DAF=0.7, but simulations with additional initial frequencies from DAF=0.1 to 0.9 are shown in fig. S4. Given the simulated allele frequency trajectory, as well as the input population size trajectory, we then used *mbs2* (*32*) to simulate samples of 2n=6,000 haplotype sequences of length 10Mb, and we randomly paired them to get samples of n=3,000 diploid genotypes. For this purpose, we set the per-generation mutation and recombination rates to 2.36e-8 and 1e-8, respectively. For each combination of initial DAF, selection model, and selection coefficient, we ran 100 simulations. We further ran additional simulations of neutral drift to confirm that SDS is indeed distributed roughly normally under the null: 750 simulations for each DAF=0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.8, and 0.9, and 2,000 simulations for DAF=0.7 (Shapiro-Wilks p-value=0.52, Kolmogorov-Smirnov p-value=0.93; fig. S2).

Simulations for the illustration figure: The simulations presented in Figures 1 and S1 were generated using the backward coalescent pipeline as described above (in Methods section: *Simulation of power and specificity of SDS to recent history*). Specifically, we used the population size model of (*14*); a recent selection of s=0.05 that started 100 generations ago; and a sample size of 2n=6,000 individuals. The genealogies (Fig. 1B) were simulated separately for each allele using *ms* (*30*). The null distribution histogram presented in Figure 1F is based on the 2,000 neutral simulations shown in fig. S2.

Simulations of a recent-admixture model for European demography: Consider the scenario that the demography of the studied population includes a recent admixture between populations that, before the admixture, evolved from different initial population sizes and at different growth rates. Their different growth rate increases the rate at which alleles would have changed frequencies in the union of populations before the admixture.

We hypothesized that this would affect the SDS predictions by inflating the raw (unstandardized) SDS variance but without changing the Gaussian shape of this distribution, and therefore without introducing false SDS predictions (since the final SDS scores are standardized by empirical variance).

To test this intuition we simulated a recent-admixture model of a European population. We constructed our model as an extension to the Tennessen model (*14*). The Tennessen model assumes an out-of-Africa bottleneck (2040 generations ago), followed by a second bottleneck for "foundation of Europe" (920 generations ago), after which the population experienced a two epoch exponential growth (rate increased 205 generations ago). We added a population split that formed a second population (pop2) at the "foundation of Europe" bottleneck, and re-admixed the populations 60 generations ago. We parameterized the model with three variables: (1) the bottleneck of pop2; (2) the growth rate of pop2; (3) the proportions of pop2 and pop1 in the admixture. We examined three values for each parameter as follows. Population size of pop2 at the bottleneck relative to pop1: 0.1, 0.5, 1.0; growth rate of pop2 relative to pop1: 0.8, 1.0, 1.2; and the proportion of pop2 to the admixture: 40%, 20%, 0%. Following the admixture we assumed that population size is as in the original Tennessen model. Thus, when proportion of pop2 to the admixture is zero, this model reduces by construction to the original model of Tennessen.

For each of the 3x3x3=27 parameter choices we simulated 1,000 samples of 6,000 10Mb-long haplotypes for alleles that evolved under neutral drift to present derived frequency of 0.7. To this end we used the recently introduced simulator MSPRIME (*33*), the only available method that can perform such large-samples complex-demography simulations in feasible computational time and space. For each simulation we defined the test SNP to be the most central SNP with DAF in 0.70-0.71 (and rejected simulations without such SNPs). Haplotypes were randomly paired into 3,000 diploid genomes, and this sample of diploids was used for the SDS computation.

The 3x3=9 of the total 3x3x3=27 examined parameter combinations that correspond to no admixture served as a control. We examined whether the variance of raw SDS changes across the 27 parameter combinations using the Levene test for equality of variance, and found that indeed admixture can inflate the raw SDS variance ($p=3.3x10^{-28}$). We then tested whether after standardization any of the distributions is different from the standard Normal distribution, using 27 Kolmogorov-Smirnov tests. Accounting for the multiple hypotheses, there was no single parameter choice that significantly changed the normality of the standardized SDS. Illustration of the model, and the SDS distributions before and after standardization for the 27 examined parameter combinations is shown in figure S3. We conclude that recent admixture coupled with different growth rates is unlikely to introduce false SDS predictions under the null of neutral drift.

We may also wonder about a more complex scenario: suppose that selection did occur in one of the pre-admixed populations but before the predicted timeframe of SDS. Would this signal be detected by SDS? This would be a correct inference of selection, but we may underestimate the timeframe of selection. This possibility is difficult to address at the current time. To answer that one would first need to obtain a better estimate of the realistic parameters of the admixture demography; such an understanding is lacking at present. Furthermore, one would need to develop new simulation tools. MSPRIME (*33*) does not simulate selection; meanwhile other methods that allow the simulation of some relevant selection scenarios in such a complex admixture demography, such as MSMS (*34*), do not scale to such large samples and large genomic regions. One may further seek to augment the SDS framework to be able to distinguish pre vs. post-admixture selection events, etc. We believe that the study of more realistic demographic models, in general and with respect to the SDS framework, would be an interesting and important line of future work.

Simulations to quantify the timescale of SDS in comparison with iHS: To quantify the effective timescale of SDS for a sample of n=3,000 individuals, and compare it to the timescale of iHS with the same sample size, we simulated a model of extremely strong

selection that had stopped T generations ago, for varying times T (Fig. 2C). To this end, we used the backward coalescent pipeline as described above, with selection coefficient s=0.10, and ran 100 simulations for each time at which selection had stopped (from T=0 to T=2,000 generations ago). For SDS (with n=3,000 individuals), even super-strong partial hard sweep events are completely undetectable if they stopped more than 100-150 generations before the present. In contrast iHS remains sensitive to events that stopped 1,000-2,000 generations before the present.

As another view of this fundamental difference between SDS and iHS, we compared the signals obtained by the two methods, using ROC analysis, for selection that started 100 generations ago and continued until the present ("recent selection") and selection that ended 100 generations ago and followed by neutral drift ("ancient selection"). This re-presentation of the simulations shown in Fig. 2B is shown and elaborated in figure S5A. While SDS obtains substantially higher scores for the recent selection, iHS typically gets higher scores for the ancient selection. In figure S5B we show that this translates into a unique practice for SDS: for a strong recent selection one can expect with high probability to obtain SDS values that are much higher than expected by chance for drift as well as for any selection with similar strength that did not entered into the very recent history. Thus, given bound estimates on selection strength, one can use SDS to make actual claims with statistical significance that selection persisted into recent times. But this is impossible to claim with iHS, regardless of the selection strength. In figure S14 we show an application of this property for interpreting the strong SDS signal we observed in the UK10K data for the lactase-persistence allele.

Simulation of selection from present variation: We wanted to test the ability of SDS to detect recent selection from standing variation, in a setting that is as close to the real data as possible. Real data reflect much more complicated demography than in our simulations; variation in mutation and recombination rates; and incomplete, noisy detection of singletons. To this end, we simulated instantaneous selection from present variation by subsampling from UK10K data. Specifically, we used the genotype data for the 3,195 individuals from ALSPAC and TWINSUK cohorts that passed our quality

controls (see Methods section: UK10K data preprocessing). For a given present and target frequency pair, we randomly picked 1,000 SNPs with a matching present frequency, and then for each SNP we down-sampled 1,500 genomes, without replacement, to match the target frequency; with Hardy-Weinberg equilibrium genotype proportions. Figures 2D and S9 present the results for target DAF=0.7. Additional target frequencies are shown in figure S8.

UK10K data preprocessing: Variant calling data and sample information for the two whole-genome sequencing cohorts from the UK10K project (*15*), ALSPAC and TwinsUK, were downloaded with permission from the European Genome-phenome Archive (EGA studies EGAD00001000740, EGAD00001000741, EGAD00001000789, EGAD00001000790). We started our quality control by considering only the 1,867 ALSPAC and 1,754 TwinsUK individuals that were reported by the UK10K project to pass their quality control (QC) for inclusion in their association analyses (*15*). Notice that this original QC removed samples with apparent non-European ancestry as detected with PCA analysis, and further excluded samples with extremely high total number of singleton variants as these were suspected to correspond to "ethnic outliers" (*15*).

To further minimize population structure in this data, we analyzed the total number of singleton SNPs per individual and used that to create an additional quality control filter for individuals with substantial non-British ancestry. To this end, we defined the UK10K singletons that were nevertheless reported by the 1,000 Genomes Project (Phase 1) (*35*) to be present outside Europe as "migration singletons". We found that the total number of singletons per individual is notably explained by two factors: the individual sequencing depth, and the proportion of singletons that are likely due to migration (fig. S7). We thus decided to further exclude, separately for the ALSPAC and TwinsUK cohorts, individuals with either highest (10%) proportion of migration singletons, or a worst (5%) fit (i.e., largest residual) to a linear model of total singletons by both sequencing depth and migration proportion (fig. S7). The remaining 3,195 samples (1,647 ALSPAC and 1,548 TwinsUK) were used for SDS predictions.

Computing SDS for UK10K data: We computed SDS for all autosomal bi-allelic variants, for which ancestral/derived annotation was reported in phase 1 of the 1,000 Genomes project (*35*), and that further have derived allele frequency between 0.05-0.95, Hardy-Weinberg equilibrium p-value $>1x10^{-6}$, and at least 5 individuals for each of the three genotypes. Singletons were restricted to single nucleotide variants (i.e., excluding indels). For a given test SNP we identified all nearest upstream and downstream singletons, if such exist, while not allowing opening a gap without any singleton (over the entire sample) longer than 1Mbp. In case more than 5% of singletons in one direction (upstream or downstream) were missing, we did not compute SDS for this test SNP. This criterion tends to exclude SNPs near centromeres and telomeres. Otherwise, for any individual with a missing singleton in either direction, we replaced the missing data value by the nearest boundary, defined as the most distant singleton that is included for this test SNP among other individuals.

The UK10K data set includes 43,826,558 autosomal variants, of which 5,844,996 are bi-allelic, in HW equilibrium ($p>1x10^{-6}$) and with MAF>5%. We excluded 423,574 SNPs because they were close to boundaries, 14,770 SNPs because they had less then 5 individuals for one genotype group, and 955,217 SNPs for a lack of ancestral annotation. Our final data set consists of 4,451,435 SNPs with valid SDS calls. The mean singleton distance (both sides combined) is 1.4 Mb. The number of singletons used for the SDS computation was 16,393,375.

For the "singleton observability" constants in the SDS model, which represent for each individual the probability (up to a scaling factor) that singletons are indeed observed in the data, we took the reported sequencing depths. This is justified by the roughly linear relationship that we observed between total number of singletons and sequencing depth (fig. S7).

The "gamma-shape" constants in the SDS model, which represent an approximate ratio between the mean squared and variance of the ideal tip-branch length distribution under the null, for each allele frequency, were estimated from simulations for sample size of

3,195 individuals using the population size model of (*14*). To this end, we used *ms* (*30*) to simulate 1,000 genealogies for each DAF between 0.05-0.95 in resolution of 0.005, and for each DAF took the average over genealogies, of the tip-branch squared mean over tip-branch variance. For intermediate frequencies "gamma-shape" constants were linearly interpolated. Raw SDS predictions were standardized within bins of 1% DAF (e.g., 0.05-0.06), by subtracting the DAF bin average and dividing by the DAF bin standard deviation. For this purpose, DAF bin average and standard deviation were estimated with chromosomes 2 and 6 excluded, to avoid a bias from the extreme signals and long range LD structures of the *LCT* and MHC regions.

Computing SDS for simulations: SDS was computed on non-phased genotype data, with all singletons observed (and taking "singleton observability" constants equal all to one). The "gamma-shape" constants were estimated similarly to the method we described above (Methods section: Computing SDS for UK10K data), from simulation of genealogies using the population size model of (*14*), but with n=3,000 individuals. Raw SDS predictions were standardized relative to corresponding neutral simulations.

Computing iHS: The integrated Haplotype Score (iHS) (*5*) was calculated using *selscan* (*36*) with default parameters. For UK10K data, predictions were done with the phased haplotypes as provided (*15*), and we excluded singletons as these were not phased. For Figure 2E we computed iHS for the British population from the 1,000-Genomes project (GBR, phase 3) (*37*), using all available phased data, including singletons. For simulations, iHS was calculated on exact haplotypes with all SNPs observed, including singletons, and raw predictions were standardized relative to corresponding neutral simulations.

Validation of SDS for UK10K by a comparison with 1,000 Genomes population frequencies: Our simulations suggest that SDS for 3,195 individuals from UK10K should reflect allele frequency changes in the British population during the past ~2,000 years. Allele frequency differences between populations, on the other hand, give a direct measure of the combined change in frequency since populations effectively separated.

Thus, to validate that SDS looks like allele frequency changes we first tested if we can build a linear model of SDS for UK10K using allele frequencies of all 26 populations in 1000 Genomes Phase 3 (*37*). The model was learned on a training set that excluded chromosomes 2, 6, and 10, as these chromosomes showed the most extreme SDS signal genomes wide (corresponding to the *LCT*, MHC and *WDFY4* regions). Indeed, a notable 25% of the variation in SDS could be explained using population allele frequencies data, and this result cross-validated on the left-out chromosomes (fig. S11).

Moreover, the extreme genome-wide signals are also in general agreement with this population frequencies model of SDS, which is especially important with regard to the MHC region (fig. S11D), providing further support that the extreme signals in this region are not driven by mapping biases (a concern that we further addressed below in Methods section: *Characterization of reference bias in read mapping in the Major Histocompatibility Complex*).

This comparison showed that there are many SNPs with a moderate SDS for which the 1000-Genomes model of SDS predicted them a much higher SDS value (fig. S11A). For example, there are 234 SNPs with $|SDS|<2.5$ but $|predicted\text{-}SDS|>3.5$. However, we found that these outliers are mostly SNPs for which there is a large disagreement between the UK10K and the 1000 Genomes British population (GBR) on both their mapping and frequency estimates: they are highly enriched (>95% vs. 0.4% background) for having a huge difference in allele frequency (>10%) between these datasets, and they are also highly enriched (>21% vs. 0.02% background) for cases in which multiple SNPs mapped by the 1000-Genomes project were mapped later to a single SNP by the UK10K project. Thus these outlier SNPs are most likely not examples of failure of SDS to detect some of the strongest strong selection signals, but rather cases of inaccuracies of the 1000 Genomes estimates.

Last, we set out to validate that the timescale of these estimated allele frequency changes by SDS is indeed much more recent than the timescale of the iHS method. To this end, we computed the Spearman correlation of SDS with allele frequency differences across

pairs of 1000-Genomes phase 3 populations, and compared them to those correlations obtained instead with iHS (Fig. 2E). SDS is mostly correlated with differences between Southern and Northern Europe (e.g., GBR-TSI ρ=0.32), and is only weakly correlated with differences between Southern Europe and Africa (e.g., TSI-YRI ρ=0.057). In comparison, iHS is much less correlated with the within Europe differences (e.g., GBR-TSI ρ=0.15), and is much more correlated with the differences between Southern Europe and Africa (e.g., TSI-YRI ρ=0.32). Thus, iHS predictions correspond to an Out-of-Africa timescale, whereas SDS predictions are specific to the much more recent timescales that had shaped variation within Europe.

Estimating the mean tip-branch length in the UK10K data: The expected tip-branch length of a sample is a function of the demographic history of the sampled population, and a decreasing function of sample size. In this paper we applied SDS to a sample of 3,195 individuals from the British population. As a working model for simulations we have used throughout the demographic model for a central/northern European population presented by Tennessen *et al* (*14*), and considered samples of 3,000 individuals (2n=6,000 haplotypes). The expected tip-branch length under the Tennessen model for a sample of 3,000 individuals is 74.9 generations. This is shown by an analytic computation in fig. S10, and in Fig. 2A by simulations (which further give the variance around this expectation). However, our current confidence in the accuracy of available demographic models is somewhat limited, and considerable uncertainty exists.

Another model, by Nelson *et al* (*38*), predicts a much longer time of 174.8 generations. A third model, by Gazave *et al* (*39*), predicts an even shorter time of 60.1 generations.

Another approach for estimating the related *median* tip-branch length was introduced more recently (*28*). This method is based on modeling an upper bound on the length of the identical-by-descent haplotype segment between "f2" variants (i.e., doubletons—sites that appear exactly twice in the entire sample). Applied to the British sample in 1000-genomes (GBR), the median tip-branch length for this small sample of n=89 individuals was reported to be 90 generations (*28*). Since mean (or median) tip length decreases with

sample size, this gives an even much lower time estimate to the mean tip-length of the 3,195 UK10K individuals. For example, the Tennessen model estimates the mean tip length of a sample of n=100 individuals to be ~275 generations. Alternatively presented, the Tennessen model predicts that the mean tip length of a sample of 3,000 individuals is ~0.27 of the mean tip for 100 individuals. For the Nelson and Gazave models this ratio is 0.4 and 0.29, respectively. By extrapolation, we can give a gross estimate that the 90 generations estimated for the GBR correspond to ~24-36 generations for the 3,195 UK10K individuals. This is an extremely short time estimate, even accounting for the fact that the median is smaller than the mean in the heavy-tailed tip length distribution. This is not a result of using an outdated over-estimate of the mutation rate, as the mutation rate used was 1.2×10^{-8}, consistent with current estimates (*28*).

Most recently, the UK10K consortium reported such an "f2 haplotype-age analysis" for their TwinsUK cohort (*15*). While this analysis focused on median tip length for doubletons shared between different regions in the UK, one can conclude that the estimated median tip length for the TwinsUK sample (n=1,754 individuals) is ~200 generations. By extrapolation from the demographic models, this roughly corresponds to ~148-164 generations for the 3,195 UK10K individuals. However, in our quality control for the UK10K data we have excluded 206 genomes from the TwinsUK cohort with abnormally high number of singletons that likely represented hidden structure due to recent immigration (see fig. S7 and Methods section: *UK10K data preprocessing*). Therefore, for the 3,195 UK10K individuals we have used here, the mean tip length is likely to be substantially shorter than this estimate.

As an alternative, we provide an independent estimate for the mean tip length in our sample (see figure S10). The idea behind this estimate is that although the average sequencing read depth in this sample is low (6.37), for several dozens of individuals the read depth is much higher, and for these individuals the total number of singletons as a function of read depth seems to approach saturation (at ~8,200). (Note that there are off-setting errors here, as some singletons may be missed in the high coverage individuals, but other sites in these individuals may falsely appear to be singletons because they were

missed in other individuals.) For this calculation we assume conservatively that the saturated number of singletons would be even ~5% higher; that only ~80% of the genome could be ideally mapped; and take a recent estimate of the mutation rate of 1.45×10^{-8} (*40*, *41*). Together, we can give a rough estimate that mean sample tip length is ~112 generations.

In conclusion, it is first important to acknowledge that there is a substantial uncertainty on the exact number. At the extremes there are estimates of ~175 and ~25-35 generations, which we believe are both outliers. Then we have on one hand a rough direct estimate for the UK10K data of ~112 generations, and on the other hand demographic model predictions of 60-75 generations. Assuming ~29 years per generation, we conclude that the mean tip length in the 3,195 individuals UK10K sample is about 2,000-3,000 years.

How does the mean sample tip length relate to the timescale of SDS? The mean tip length is just one rough estimate for the characteristic timescale. One may speculate that since the tip length distribution is heavy tailed, then the mean may not be a good indicator of timescale (i.e., that the heavy-tail variance increases the effective timescale way beyond the mean). However, this intuition is wrong. The reason is that ancient changes in allele frequency affect the expected length of only the few tip-branches that reach that age regime; whereas a frequency change very close to the present affects the expected length of *all* tip-branches. Thus the most recent events contribute disproportionally more to the SDS statistic. This is indicated by our simulations (Figures 2B, S6, S14) which show, for example, that when selection stopped ~125 vs. ~150 generations ago the effect on the SDS signal is negligible; whereas if selection had stopped 25 generations ago or alternatively continued until the present, the difference in SDS could be very large (i.e., this difference of 25 generation regarding when selection had stopped contributes more to SDS the later it is). In the section below, we show how one could use SDS to make a statistical claim that the selection on lactase most likely persisted into the past 2,000 years even if the mean tip length is assumed to be as large as 4,000 years.

Timescale of selection on lactase: The specificity of SDS to recent times can be utilized to make statistical claims about whether a strong selection event persisted into recent history, assuming bounds on the selection strength (see fig. S5 and Methods section: *Simulations to quantify the timescale of SDS in comparison with iHS*). Here we applied this principle to interpret the SDS signal we observed in the UK10K data for the lactase-persistent allele (rs4988235, DAF=0.74, SDS=9.58).

The estimated strength of this well-known selection signal in northern Europeans varies considerably. The early study placed it in the wide range of 1.4-15% for the CEPTH population, and 9-19% for the Scandinavian population (*2*). Another work suggested 5.18-15.9% for the subpopulation of dairying farmers (*42*); but this should translate into about a half of that for the entire population (*43*). Other (including the most recent) studies have suggested a lower range of selection coefficients, namely 1-3% (*44*, *45*).

We thus simulated 300 samples of 3,000 individuals for an allele with present derived frequency of 0.7, which undergone a strong selection (s=5%,7.5%,10%) that stopped some time along the past 200 generations (and followed by neutral drift). Simulations were performed as described in Methods section: *Simulations of power and specificity of SDS to recent history*. For each possible stopping time for the simulated selection, we then counted how many simulations resulted in an SDS value greater or equal to the observed SDS signal for lactase in the UK10K data (see figure S14). These simulations suggest that for any selection pressure ≤10%, the observed signal is too high to be the result of selection that stopped ≥37 generations from the present ($p<1/300$). In the Tennessen model 37 generations is half of the mean tip length (~75). We will assume that the true demographic model of the British is similar to the Tennessen model up to a scaling factor, and that the actual mean tip length ≤4,000 years (we estimate it to 2,000-3,000 years; see Methods section above: *Estimating the mean tip branch length in the UK10K data*). Thus, assuming that the strength of selection on the lactase-persistent allele was ≤10%, which is higher than most of current estimates, we can infer that this selection most likely entered into the past 2,000 years ($p<1/300$).

Alternatively, we can give bounds on the selection coefficient. For example, under a more realistic estimate of the mean tip length (≤3,000 years) we can infer that if this selection did not continue into the last 1,000 years (i.e., stopped ≥25 generations ago in the Tennessen model), its selection coefficient was likely strictly greater than 7.5% ($p<1/300$; see figure S14).

Characterization of reference bias in read mapping in the Major Histocompatibility Complex (MHC): Reads carrying more than one alternative allele (i.e., non-reference allele) are difficult to align to the reference genome, so there exists asymmetry in read depth and SNP calling between haplotypes carrying the reference allele and those carrying the alternative allele. We hypothesized that this reference bias might be particularly pronounced in the MHC region, due to its high density of polymorphisms and unusual long-range LD structure. On the one hand, the reference bias could cause difference in the probability of calling nearby singletons among individuals of different genotypes: on average, fewer variants will be called in individuals carrying more alternative alleles, possibly leading to spurious SDS signals for increased frequency of the *alternative* allele at the focal SNP site. On the other hand, the reference bias may lead to underestimation of the frequency of the alternative allele at the focal SNP, possibly producing spurious SDS signals for increased frequency of the *reference* allele (this is basically equivalent to our simulation of instantaneous selection from present data – here the mapping bias is like an instantaneous down-sampling of the alternative allele).

With a concern that the extreme signal observed in the MHC region is a false positive driven by a biased read mapping, we carried out several analyses to characterize the range and extent of reference bias in the MHC region and its effects on SDS. We took an operational definition of the extended MHC region as the region bounded by SNPs rs498548 (chr6: 25,892,529) and rs2772390 (chr6: 33,436,144) plus 2Mb extensions on both sides. We calculated the Spearman correlation between genotype (i.e., the number of alternative alleles) at each common SNP (MAF≥0.05 in the 3,195 individuals included in SDS calculation) and read depth at each of the other SNPs in the MHC region. As expected, the read depth at a SNP is usually negatively correlated with number of

alternative alleles at nearby common SNPs (the diagonal line in fig. S16A1). This read depth asymmetry appears to be particularly strong in a few regions (the vertical patterns in fig. S16A), where the read depths are correlated with genotypes at almost all SNPs across the MHC region, reflecting the extensive LD structure. However, the lack of horizontal patterns shows that no common SNPs are vastly associated with read depths at all sites across the MHC region, indicating that the effect of reference bias in read depths is highly localized.

To quantify the range and extent of the reference bias around a common variant, we combined the read depth at all polymorphic sites within certain distance (5Kb, 50Kb or 500Kb) from the focal SNP, and tested for association between the combined read depth and genotype at the focal SNP. We regressed the combined read depths on the number of alternative alleles across the 3,195 individuals and calculated the ratio of the predicted combined read depth for homozygotes of the alternative allele to that of the reference allele. While the reference bias is evident in certain parts of MHC on short distance scales (in 10Kb window), the effect is weak in general: 95% common SNPs show less than 10% differences in read depth between the two homozygotes. The effect almost diminishes at 1Mb scale (95% common SNPs have less than 3% difference; fig. S16B). Given that the distances of a common variant to the nearest singletons are usually on Megabase scale, the short-scale reference bias in read depth is unlikely to affect SDS results.

Finally, to directly investigate the effect of reference bias on SDS, we quantified the correlation between SDS for the reference allele (positive value meaning selection for) and the log ratio of read depth for the alternative allele to that of the reference allele: the correlation is slightly negative on both short and long distance scales (Spearman's ρ=-0.095 for 10Kb window, and ρ=-0.077 for 1Mb window; fig. S16C). The direction of this correlation is consistent with the expectation that reference bias could lead to underestimation of the alternative allele frequency, thereby falsely increasing the SDS for the reference allele. Nevertheless, the correlation is so weak that very little variation of SDS in the MHC region can be explained by the extent of reference bias in read depth. Moreover, SNPs with relatively high SDS values in this region usually do not show large

asymmetry in read depth (fig. S16C). Additionally, the most extreme SDS values at the MHC regions are signals for increased frequency of the alternative alleles, so the overall weak bias for decreased SDS for alternative alleles is definitely not driving these most extreme SDS signals. Thus, the strongest selection signals in the MHC region are not driven by a reference bias in read mapping.

Analysis of high SDS enrichment for pigmentation variants: To test whether alleles with known phenotypic effects tended to be targets of positive selection in ancestors of the British within the past 2,000 years, we examined the distribution of SDS over NHGRI GWAS catalog SNPs (downloaded October 1, 2015). To this end we considered all SNPs that were reported for phenotypic association with genome-wide significance ($p<5x10^{-8}$), and removed duplicates. We tested for an inflation of squared SDS values using a chi square test, under the assumption that SDS has a Normal distribution. The rationale for this test is that trait-associated SNPs may be more likely than random SNPs to undergo short-term directional selection, however we do not know the direction of change *a priori*.

This set of variants with known effects showed significantly inflated variance of SDS ($p=7.4x10^{-16}$). Although a notable part of this signal comes from the extended HLA and *LCT* (lactase) regions (chr6: 25,892,529-33,436,144 and chr2: 134,608,646-138,608,646; hg19 coordinates), the test was still significant after we excluded those regions ($p=4.9x10^{-7}$).

To further identify relevant categories of SNPs under selection we grouped the significant NHGRI GWAS catalog SNPs by ontology and tested each term separately for increased SDS variance. After removing extended MHC and *LCT* regions, retaining only the most significantly associated SNP per 100Kb window per term and requiring at least 3 SNPs per term, each term was tested for inflated variance by taking the sum of its squared SDS scores and comparing to the chi square distribution with appropriate degrees of freedom. At a false discovery rate of 5%, only three terms were above the threshold: “suntan”, “eye color” and “hair color”.

To follow up on this finding, we compiled all associations from the GWAS catalog that were annotated with pigmentation-associated ontology terms: eye color, freckles, hair color, skin pigmentation, skin sensitivity to sun, sunburn, and suntan. Variants from studies that indirectly measured the pigmentation traits were excluded (e.g., studies which measured clinical response to sun through cancer risk). We then considered all these variants for which we further had SDS predictions (e.g., MAF>5% and available derived/ancestral annotation; see Methods section: Computing SDS for UK10K data). We again excluded variants within the extended MHC and *LCT* regions. For multiple SNPs located in the same locus we kept only the variant with most significant p-value, unless variants were more than 100Kb apart and we further found an experimental support in the literature for their independent functional significance. This resulted in 14 SNPs with available SDS scores that are associated with either pigmentation or freckling. In all 14 cases the derived allele is associated with either lighter pigmentation (i.e., lighter hair, skin, or eyes) or increased freckling. The average SDS for these alleles is significantly larger than expected by chance (average: 1.58, one-sided p-value: $2x10^{-9}$, testing against standard normal with mean=0). The strongest signal is for selection in favor of a blond hair variant at the *KITLG* locus, currently at 12% frequency (p=2x10-6), and a weaker signal in favor of blond hair at *SLC24A4* (p=8x10-3). We also replicate a known signal for blue eyes at the *HERC2*/*OCA2* locus (p=2x10-5) (*12*, *24*).

The case of the *KITLG* blond hair allele (rs12821256) is of special interest (*46*). The *KITLG locus* shows a clear signal of selection in recent human history that includes an extended haplotype at high frequencies outside Africa, in both Europe and East Asia (*47*–*50*). It is thought that this selection is related to the established role of *KITLG* in skin pigmentation (*51*). When rs12821256 was first found to be associated with the blond hair phenotype it was also tested for being a target of selection (*46*). Unlike the extended haplotype, the blond-hair allele is present in Europe but not in East Asia, and at much lower frequencies (~12% in UK10K). Sulem *et al.* showed that rs12821256 is present almost exclusively on the background of the extended haplotype, and that conditioned on the extended haplotype background it does not show a selection signal (of extended

haplotype). They concluded: *"Thus, the rs12821256 C allele is not itself under positive selection, but rather is a hitch-hiker whose frequency is driven up by some selective advantage that is conferred by the extended haplotype."* (*46*). However, it seems likely that the hard sweep method that they used was underpowered for detecting a recent partial sweep that occurred on the background of a much older sweep that had already purged most of the variation from the region.

In contrast, the SDS signal at the *KITLG* locus is peaked almost exactly on the blond-hair allele (fig. S20). To more directly disentangle between the SDS signal and the extended haplotype signal, we computed SDS for the blond-hair allele conditioned on being on the extended haplotype background. To this end we considered three SNPs that were previously used to tag the extended haplotype (list taken from (*52*)): (1) rs642742 (*51*), (2) rs10732643 (*49*), and (3) rs1881227 (*50*). For each of these tag-SNPs we took all UK10K individuals (of the 3,195) that were homozygous for the extended haplotype tag. There were 2,181, 2,836 & 2,881 individuals homozygous for SNPs (1), (2) & (3) above, respectively. The frequency of the blond-hair allele increased in these subsets in line with the reported association with the extended haplotype (from 12.4% in UK10K to 15.2%, 13.1% & 13.1%). We then computed SDS on chromosome 12 for all SNPs within the same DAF-bin to that of the blond hair allele, and used that to standardize the raw scores. In all three cases, the SDS score for the blond hair allele (rs12821256) conditioned on the extended haplotype background is slightly higher (SDS=5.15, 4.87 & 4.91) than the original 3,195-UK10K prediction (SDS=4.60). We conclude that the blond-hair allele was not a hitch-hiker on the extended haplotype but rather a target of more recent selection.

In summary, there is by now evidence that the *KITLG* locus has been subject to at least two independent selection events in recent human history: (1) selection related to light skin color at about the time of migration of humans out of Africa, which left a signature of an extended out-of-Africa haplotype; and (2) more recent selection on the blond-hair allele in northern Europe – likely persisting well into the past 2,000-3,000 years – that

left a signature of lower density of singleton variants in UK10K around the selected allele.

GWAS processing: Summary statistics from published GWAS for 44 traits were downloaded and converted to a common file format. Criteria for inclusion consisted of study subjects of European ancestry, presence of signed effect sizes, estimates for at least 1 million markers, and at least one genome-wide significant hit reported in the paper. For some traits, summary GWAS statistics were reported for men and women separately as well as for the combined cohorts. The analyses presented in Figure 4A,B are based on the sex-combined summary statistics for height from the family-based study of (*20*). For the analysis presented in Figure 4C we only included the sex-specific summaries, where those were available, and excluded the combined GWAS. For consistency with the other studies, Fig. 4C also excludes the height and BMI results from the family-based studies of Robinson et al. After insertion and deletion variants were removed, SNP RSIDs were remapped to the GRCh37 version of dbSNP build 142 and assigned genomic coordinates. Odds ratios were $\log_2$ transformed to report a signed effect size per SNP (to facilitate comparison with fit coefficients per SNP for quantitative traits). To match the same direction as SDS scores, reported effect sizes' signs were polarized such that positive GWAS summary statistics represented a trait-increasing effect of the derived allele. For each GWAS, SNPs for which we had estimated an SDS score were taken forward for significance testing.

GWAS selection testing: For all downstream analysis, strand ambiguous SNPs (i.e. A>T or G>C) were removed to prevent the possibility of strand errors, and the extended MHC and lactase regions were excluded because of the previously known selective pressures exerted on them (chr6: 25,892,529-33,436,144 and chr2: 134,608,646-138,608,646; hg19 coordinates). The pipeline for each GWAS was identical: SDS scores were re-normalized with respect to the derived allele frequency in 1% bins to control for possible ascertainment biases in the GWAS SNPs. Following this adjustment, tSDS scores were generated by switching the sign of SDS scores so that a positive tSDS score represents an increase in the frequency of the trait-increasing allele reported by the GWAS. The

strength of selection acting on a trait was taken as the spearman correlation between tSDS and GWAS log p-value such that positive correlation would reflect selection for increasing the trait and negative for decreasing the trait.

To ascertain the significance of the correlation in the context of LD, a blocked jackknife approach was employed to calculate the standard error in the Spearman correlation estimate (*53*, *54*). For each GWAS, SNPs were assigned to one of 1,500 contiguous blocks based on concatenated genomic coordinates for an average of 2 Mb of sequence per block. Correlated tip lengths should be accounted for by this estimation. Spearman correlation was calculated globally and after removing each block to calculate the jackknife error. Raw p-values were calculated for the point estimates using null normal distributions with mean zero correlation and standard deviation equal to that of the jackknife error. P-values were robust to binning approach and were confirmed by an alternate method in which the genome was partitioned into fixed 1 Mb blocks and the direction of tSDS scores for all SNPs in the block were flipped with one-half probability prior to calculating genome-wide correlations between tSDS and GWAS summary statistics. Log p-values produced by this method were essentially equivalent ($R^2 > 0.99$). We further verified that significance of the GWAS-tSDS correlations is not affected by local variation in G/C content (fig. S24), nor by regional variation in the B-statistic – a proxy for background selection and known correlate of functional density (*55*) (figures S25, S27).

Estimating SDS-GWAS correlations using LD Score regression: Consider a GWAS study of a complex trait. Let z_i be the Z-score of the signal on SNP i. The underlying principle of LD Score regression is that, under a polygenic model (i.e., assuming many causal variants), the expected squared effect size $E[z_i^2]$ scales linearly with the effective number of LD partners tagged by that SNP (referred to as the SNP's LD Score). Thus for a GWAS with true signal (as opposed to spurious signals due to confounding) we expect to see a positive relationship between LD Score and z_i^2. Bulik-Sullivan and colleagues proposed LD Score regression as a tool for assessing the extent of polygenic signal in a GWAS study (*24*). Using theoretical models, and simulations based on real genotype

data, they argued that the main effect of population structure confounding is to push the LD score regression intercept above 1; but that realistic levels of structure are unlikely to have significant impact on the slope.

In a follow-up paper, Bulik-Sullivan and colleagues extended LD Score regression to measure the correlation between pairs of traits (*25*). Let z_{i1} and z_{i2} be the Z-scores at SNP i for two different traits. Then the expected slope of the regression of z_{i1} x z_{i2} is proportional to the genetic covariance between traits 1 and 2. Here we extend this intuition by replacing GWAS z scores by SDS z scores: specifically we measure the correlation between the product of a trait and SDS z scores: i.e., z_i x SDS_i against LD Score. The logic here is similar to that in the GWAS context: we expect trait effect sizes to have larger variance in regions of high LD Score, and hence they should have proportionally greater effects on SDS. Note that this is expected to be a much less powerful analysis than simply testing for correlation between z_i and SDS_i, yet more robust if the GWAS data are confounded by population structure.

To apply LD Score regression, genetic correlation and genetic correlation p-values for SDS Scores and GWAS summary statistics were calculated using code from Bulik-Sullivan et al (*25*) (https://github.com/bulik/ldsc). SDS Scores were processed as genome-wide summary statistics for association study. The sample size option (N) was fixed at the number of informative individuals in the combined TWINSUK and ALSPAC cohort (3,195) for SDS Scores and the number of individuals reported in each GWAS for GWAS traits, though in practice this choice had little impact on the resulting p-values. Other pipeline options were left as the default for all traits. LD Score regression intercepts were fit and largely fell near 1 as expected. Traits showing both a Bonferroni significant tSDS correlation p-value and a LD Score regression one-sided p-value under 0.05 were considered suggestive of selection truly acting on the trait. The LD Score regression p-values were enriched for small p-values and concordant directions with the main analysis, but aside from height the individual significance levels were modest (the smallest was 5×10^{-4}, one-sided, for female hip size). See the Supplemental Excel file for

details of the GWAS studies examined, and the results obtained for them with each of the statistical tests above.

Analysis of family-based height data using ashR. To investigate our observation that mean tSDS is >0 even for nonsignificant p-values, we applied the ashR method (*23*) to the family-based GWAS data from Robinson et al (*20*). The goal was to assess whether it is plausible to find a tSDS signal for SNPs that do not show any significant association in the GWAS analysis. ashR operates under a unimodal assumption whereby the mode of true effect sizes are zero, and null p-values are uniformly distributed across the interval 0 to 1. By maximizing the likelihood of measured effect sizes and standard errors under this assumption and the constraint that the least significant p-values are assigned to the null distribution, ashR can report for each test the probability of the true sign of beta being 0, the same as the estimated sign, or the opposite sign. Our rationale for this analysis is that in order for mean tSDS to be positive, the fraction of SNPs with correct sign has to be substantially greater than the fraction of SNPs with incorrect sign. The context of a GWAS study differs from standard applications of ashR as noncausal SNPs may have large estimated effect sizes because they tag - are in LD with - a causal site. This is appropriate for the present application, since the expected sign of SDS should depend on the summed effects of a SNP and all its LD partners, regardless of whether it, itself, is actually causal.

Results from this analysis are shown in figure S23. As shown in the figure, even for SNPs with p-values >0.5, ashR estimates that a substantial fraction of SNPs have a nonzero effect and, moreover, that the sign is estimated correctly more often than not. This provides independent confirmation of the plausibility of our tSDS results for height at nonsignificant SNPs.

Correspondence between SDS signals of polygenic adaption and phenotypes of modern British: We find it interesting to report that for some traits for which we see a signal of selection there is some evidence that they indeed match phenotypic characteristics of present day British. For height this has been reported before: there is a

notable south to north gradient of increased stature within Europe (*20*, *9*), consistent with our signal for recent selection for tall stature in ancestors of the British. Two GWAS meta-analyses show a signal in our study consistent with selection for delayed sexual maturation in British women: SDS correlates with a lower Tanner stage (*56*) and a later age of menarche (*57*). Perhaps consistent with that, age at menarche was also reported to show a positive south to north gradient in Europe (*58*, *59*). It is also known that head circumference in the British is among the largest in Europe, consistent with the selection signal we observed for increased infant head circumference. For example, the following literature review summarizes reports for children head circumference at two years across 26 countries around the world (*60*).

Supplementary Data File.

Table S2: Summary of GWAS data sets and results. This table is available online as an Excel file.

External Online Resources.

A table of SDS values and R code for computing SDS are available through the Dryad Digital Repository at http://datadryad.org/resource/doi:10.5061/dryad.kd58f and GitHub at https://github.com/yairf/SDS, as well as through the authors' website at http://pritchardlab.stanford.edu.

Supplementary Figures.

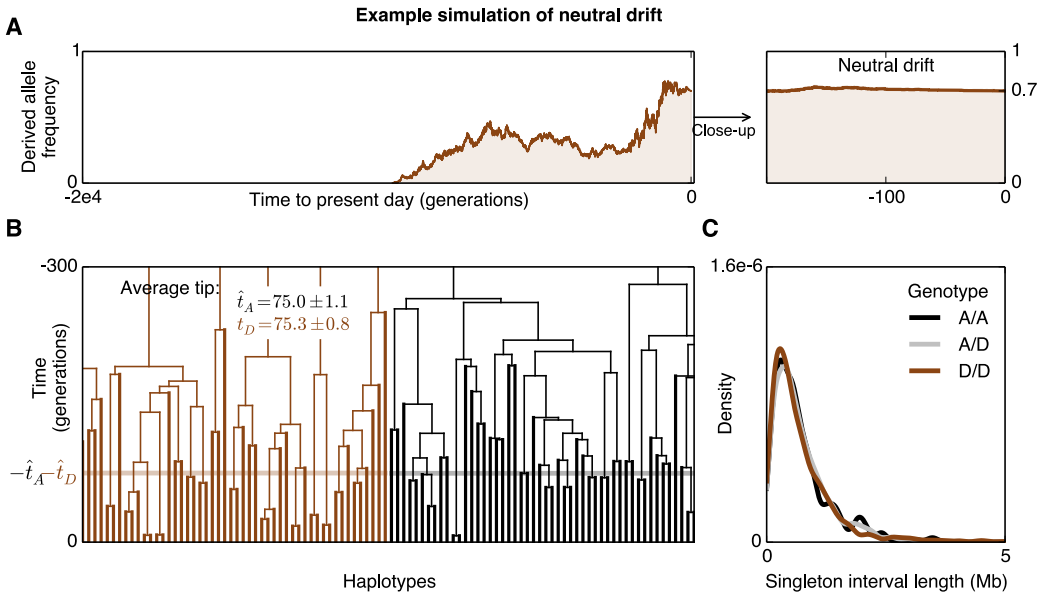

Figure S1: Illustration of SDS in the case of neutral drift. (A) Simulated allele frequency trajectory for a derived allele that drifted (i.e., was selectively neutral) to present day frequency of 0.7. Compare with Figure 1, which shows a simulated example of an allele that followed a strong selection from standing variation starting 100 generations ago to reach the same present day frequency of 0.7. **(B)** Blow-up of a small part of the genealogy for a random sample of 3,000 present day genomes (6,000 haplotypes). The mean lengths of terminal ("tip") branches for the derived and ancestral alleles are highly similar (~75 generations). Compare with the selection example in Figure 1, where recent selection led to favored allele haplotypes having shorter tip-branches than in the neutral case, and to disfavored allele haplotypes having longer tip-branches than in the neutral case. **(C)** The distribution of distances between consecutive singletons over the simulated test-SNP (which captures singleton densities), for the three test-SNP genotypes. As average tip-branch lengths for the ancestral and derived alleles are similar, the singleton densities are similar. SDS is the standardized, inferred log-ratio of mean tip lengths. For this neutral simulated example SDS is not significantly different than zero (p=0.35).

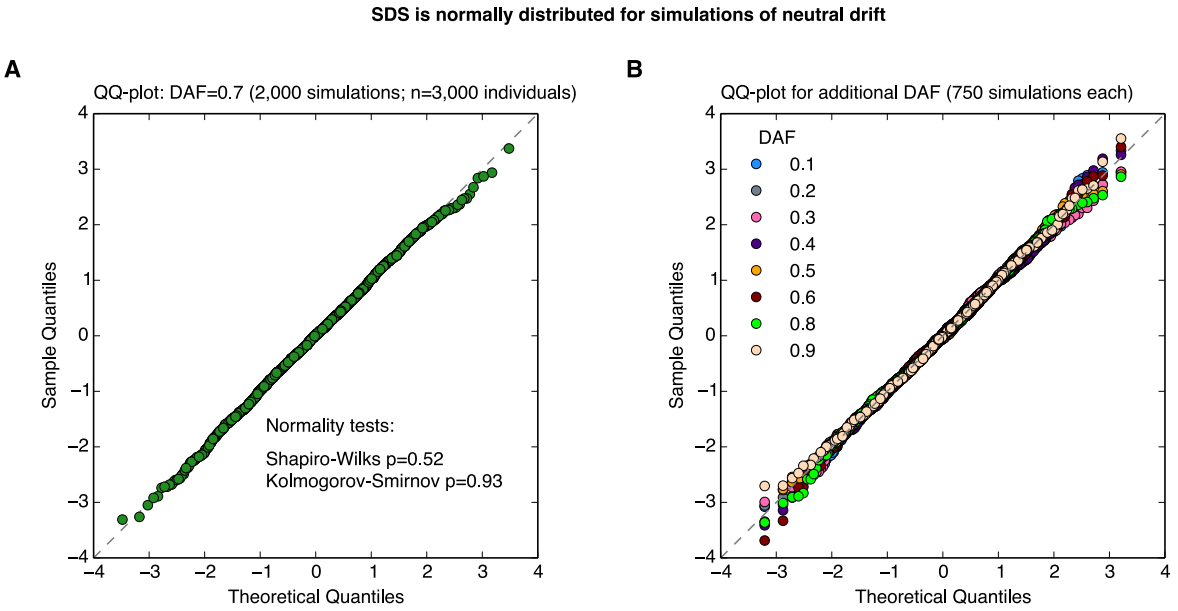

Figure S2: QQ-plot for neutral SDS simulations. (A) Standardized SDS values for 2,000 neutral simulations are shown against the quantiles of the Normal distribution. Simulations are based on a European population size model, for alleles with present day frequency of 0.7, and samples of 3,000 individuals. Formal statistical tests for Normality are not rejected (Shapiro-Wilks p=0.52, and Kolmogorov-Smirnov p=0.93). For more simulation details see Methods section: Simulation of power and specificity of SDS to recent history. **(B)** Standardized SDS values for neutral simulations as in (a), but for additional present-day derived allele frequencies (DAF=0.1-0.9; 750 simulations per DAF). Note that we standardize SDS using the empirical standard deviation. This is essentially an assumption that true selection signals are sparse; it will tend to be conservative if a large fraction of the genome is actually under selection.

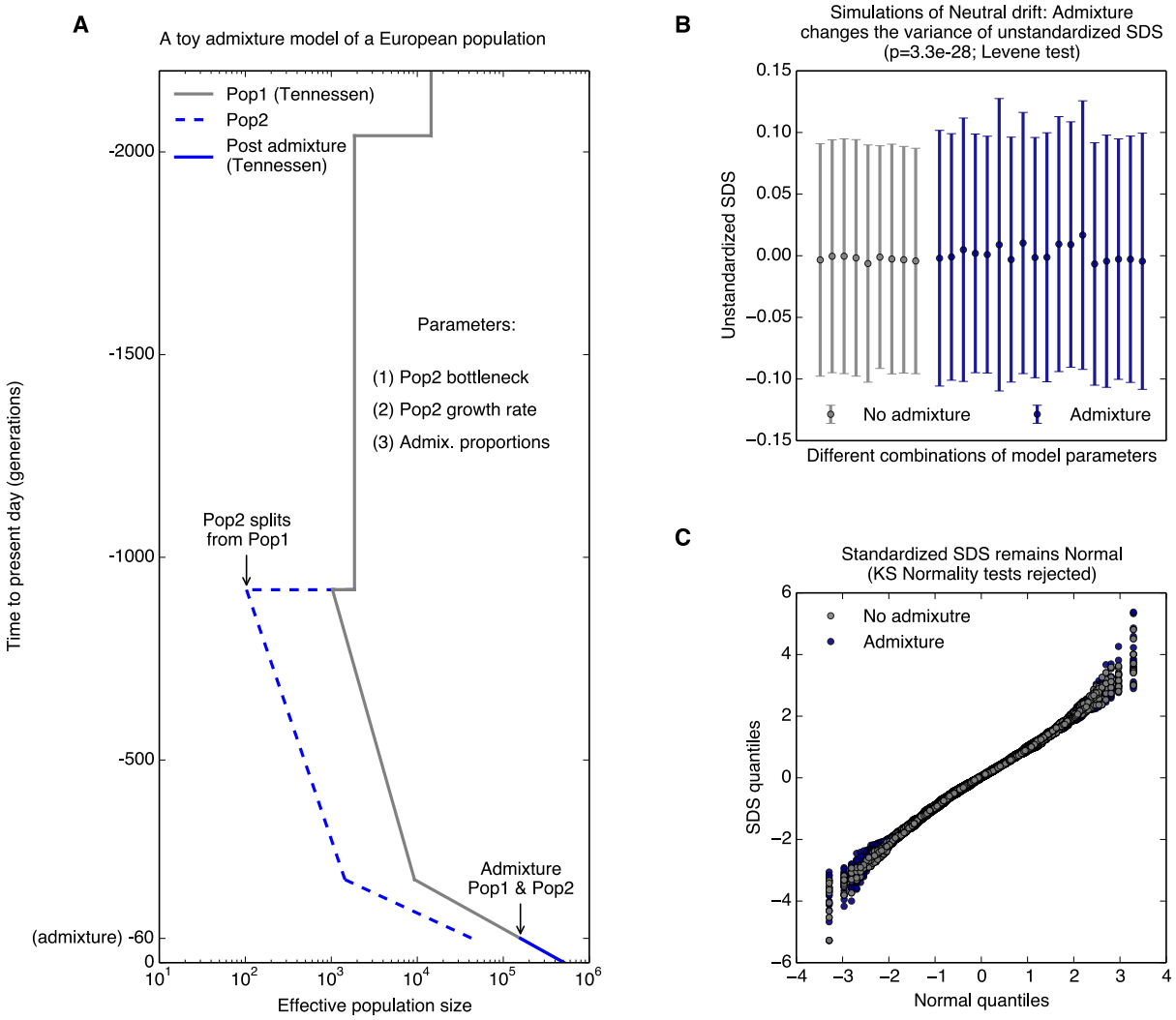

Figure S3: Neutral SDS simulations for a simple admixture model of European demography. (A) Illustration of the demographic model we used to simulate the effect of recent admixture. As a base model we use the Tennessen model (*14*). Tennessen assumes two bottleneck events, one for out-of-Africa and another for foundation of Europe. We assume that a second population (pop2) had split at the Europe-foundation bottleneck. The bottleneck of pop2 relative to pop1 is a parameter, and we considered values: 0.1 (illustrated), 0.5, and 1. Tennessen then models a two-epoch exponential growth. We further parameterized the growth rate of pop2 relative to pop1, and considered values: 0.8, 1, and 1.2 (illustrated). We modeled admixture of the two populations 60 generations ago. After admixture we set the effective population size to that of the Tennessen model. The fraction that pop2 contributed to the admixture is our last parameter, and we considered: 40%, 20%, and 0% (i.e., without admixture). **(B)** Raw (unstandardized) SDS values (mean ± SD) for 1,000 simulations for each of the 3x3x3=27 parameter choices described. Admixture inflates the variance of unstandardized SDS. **(C)** QQ-plot of standardized SDS for the 27 parameters. Thus, admixture (coupled with different growth rates) is unlikely to lead to non-normality of SDS under the null of neutral drift.

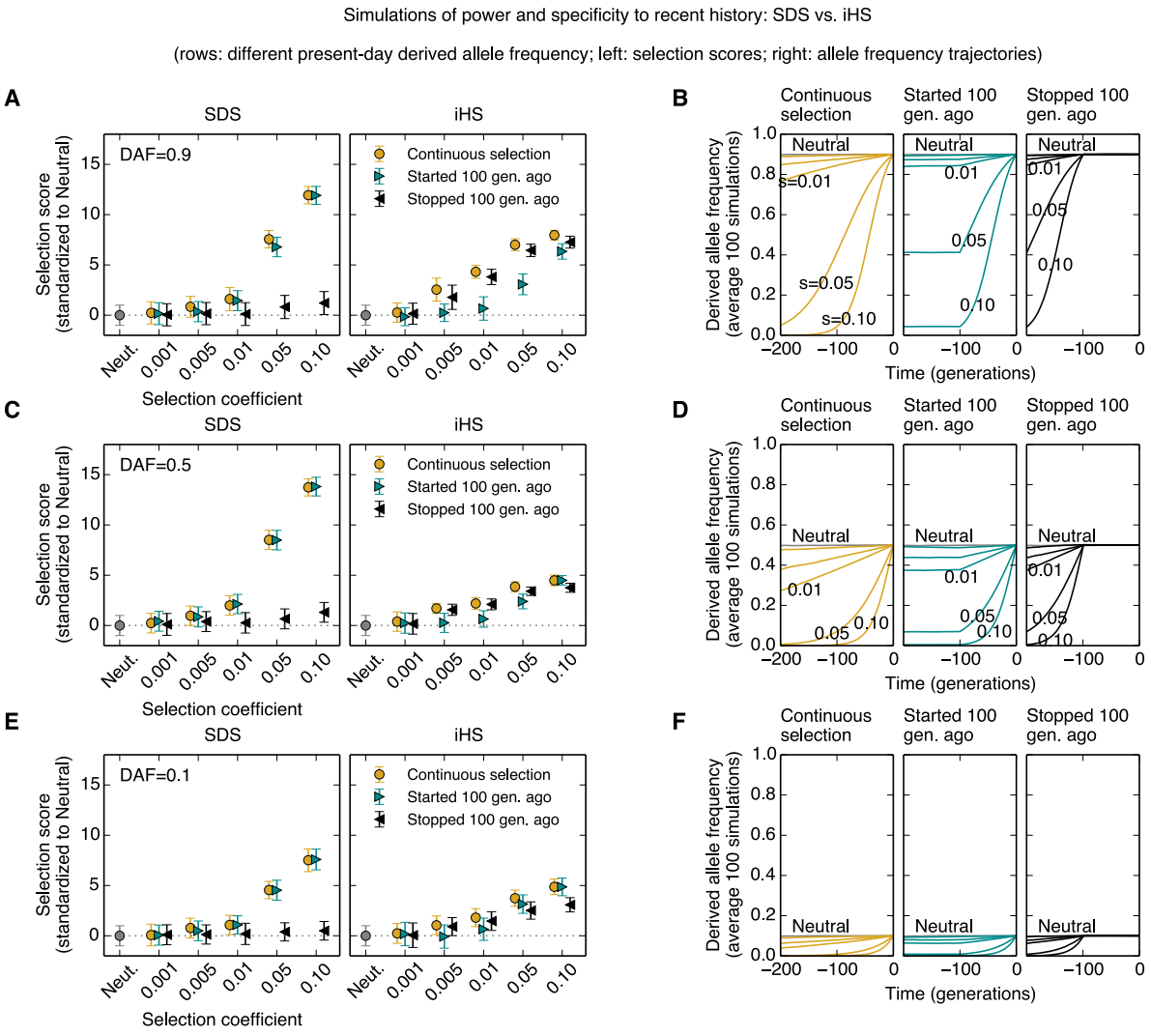

Figure S4: Extended simulations of power and specificity to recent history. (A) The signal (mean ± SD) of SDS (left) and iHS (right) for simulations of sequences for 3,000 individuals under different models of positive selection, for alleles with present derived frequency of 0.9, using a European population size model. Three models of positive selection are presented: orange: continuous selection throughout history (i.e., the classic partial hard sweep model); cyan: selection that started from standing variation 100 generations ago; and black: selection that stopped 100 generations ago. Each model is shown for different selection coefficients, and is standardized relative the Neutral drift model (gray). **(B)** The average frequency trajectories for these simulations are shown for the past 200 generations. **(C,D)** Simulations as in (A,B) for present-day DAF=0.5. **(E,F)** Simulations as in (A,B) for present-day DAF=0.1. This simulation figure thus extends on the simulations presented in Figure 2B by showing results for additional present-day derived allele frequencies, as well as by providing more information on performance of iHS and by showing the actual DAF trajectories. For more simulation details see Methods section: *Simulation of power and specificity of SDS to recent history*.

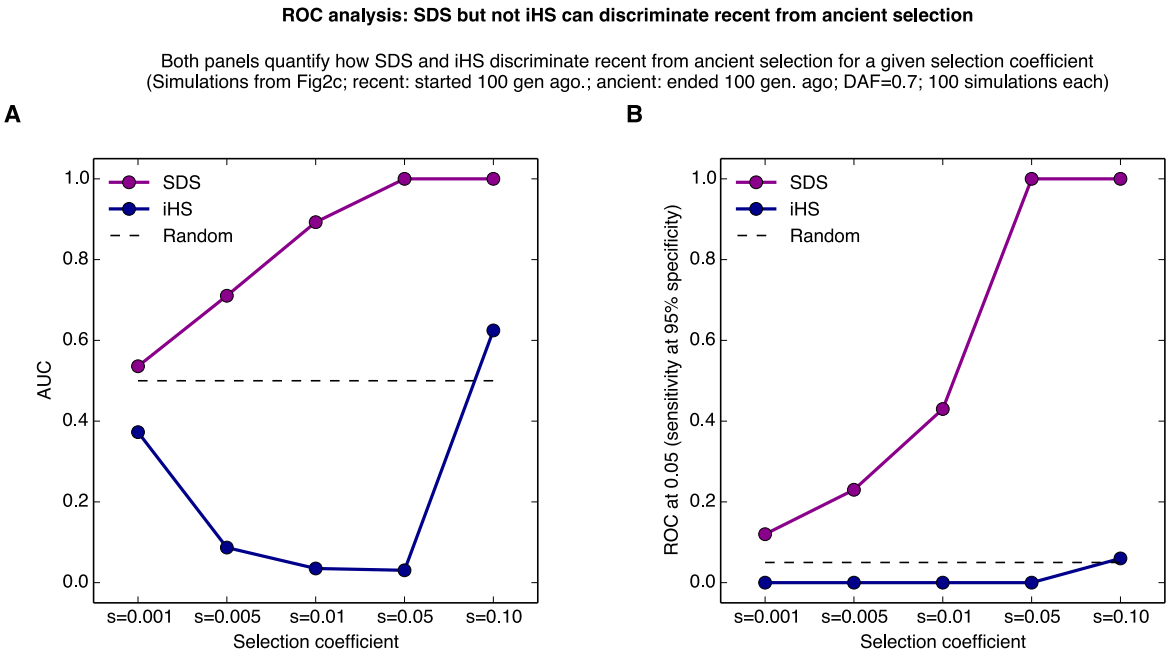

Figure S5: A unique feature of SDS compared with iHS is the ability to show that selection persisted into very recent history. (A) ROC analysis for discriminating 100 simulations of selection that started 100 generations ago (i.e., from standing variation) from 100 simulations of selection at the same strength that stopped 100 generations ago (i.e., followed by drift). For SDS recent selection results with increased score relative to the ancient selection; whereas for iHS it is the ancient selection that typically gets a higher score. Simulations are from Figure 2C (Tennessen model; a sample of n=3,000 individuals) **(B)** Similar to (A) but instead of showing the area under the ROC curve (AUC), we show the point of the ROC at x=0.05. Assume that selection started 100 generations ago (continued until the present), with known selection coefficient. Can we use the observed signal (SDS or iHS) to determine whether selection continued within the past 100 generations? The worst case under the null is a selection at that strength that stopped exactly 100 generations ago. The ROC at x=0.05 is the estimated probability that one observes a signal that is higher than 95% of such worst case null simulations. By chance we expect 5% to be higher than 95% null simulations. With iHS we do not expect to exceed the 5%, which comes from the fact shown in (A) that it usually gets an even higher signal for the older selection scenario. With SDS we can expect to be able to make such claims about selection continuing into the past 100 generations for almost all instances of a strong selection ($s \geq 0.05$). In Figure S14 we use this property to infer that the SDS signal we observe in UK10K for lactase most likely reflects selection that persisted well into the last 2,000 years.

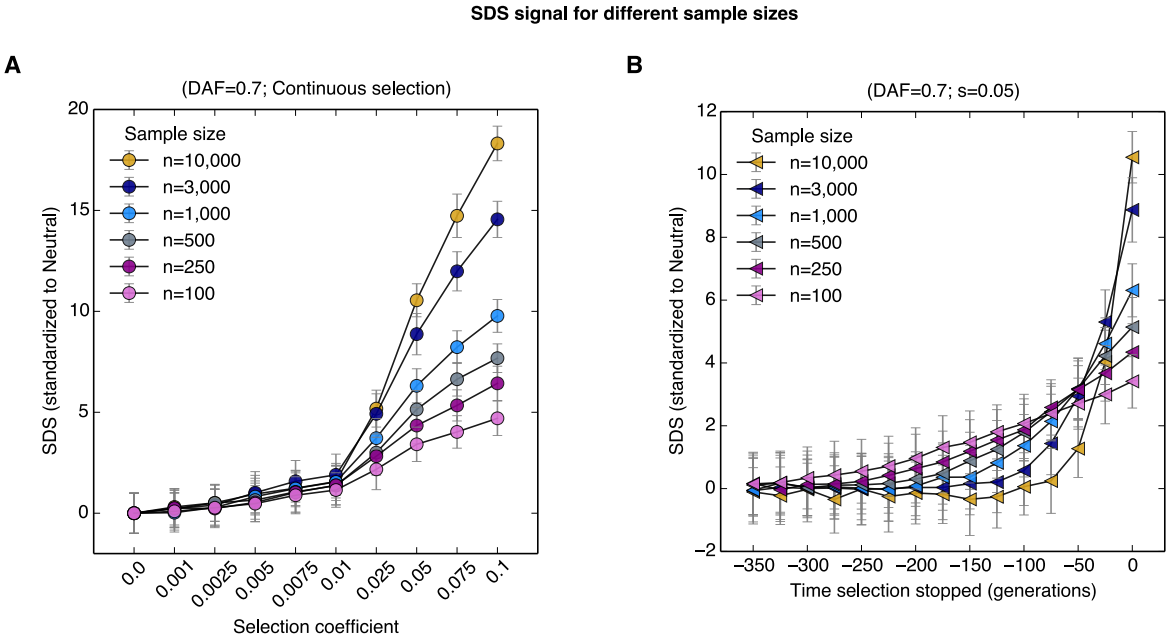

Figure S6: SDS power and timescale depends on sample size. (A) SDS signal (mean ± SD) for simulations of a continuous selection model (i.e., hard sweep) in European demography (*14*), for various selection coefficients (x-axis) and sample sizes (colors). For both panels, each point corresponds to 200 simulations; except for the largest sample size (n=10,000), for which we performed only 50 simulations due to computational burden. **(B)** Simulations of the SDS signal for a strong selection (s=0.05) that stopped in the past and followed by drift, for various stopping times (x-axis) and for various sample sizes (colors). With smaller samples SDS has fewer independent observations to estimate the change in frequency, and so it has less power to detect a selection that persisted to the present. However if selection only occurred at more ancient times, there is a tradeoff between sampling error and effective timescale. The smaller samples have longer tips, and so SDS can integrate these old signals that are out of reach for the large sample SDS; leading to increased signal with decreased sample size.

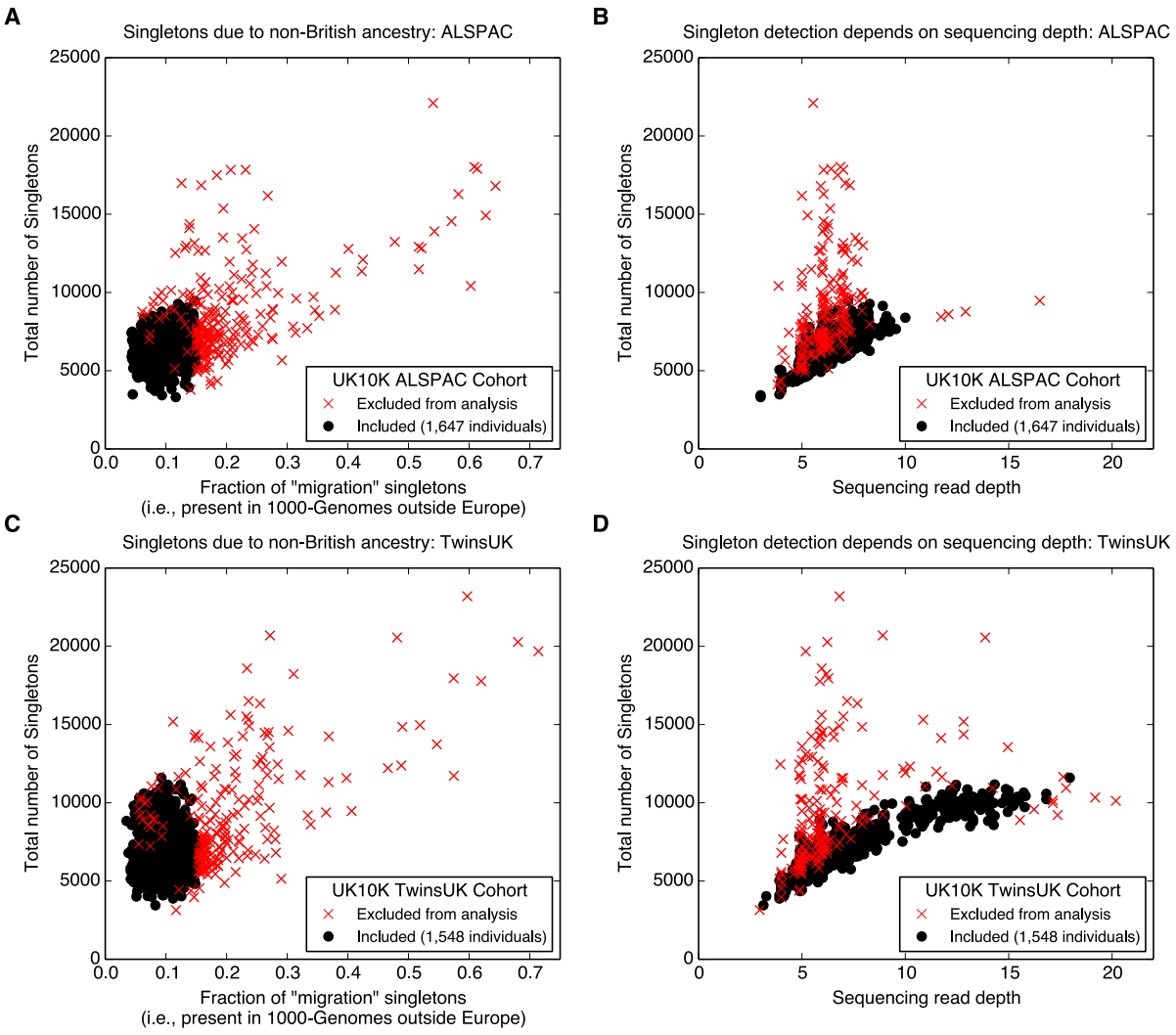

Figure S7: Data quality control to remove UK10K individuals with abnormal numbers of genome-wide singletons. (A,B) For the ALSPAC cohort, shown in (A) is the total number of singletons per individual, by the fraction of these singletons that were detected by the 1000-Genomes project to be also present in non-European populations (termed here "migration" singletons). The total number of singletons by the sequencing depth is shown in (B). Individuals were excluded from downstream analysis (red), if they had either an abnormally high fraction of migration singletons, or an abnormally high number of total singletons accounting for both migration singletons and sequencing depth (see Methods section: *UK10K data preprocessing*). **(C,D)** As in (A,B) but for the TwinsUK cohort.

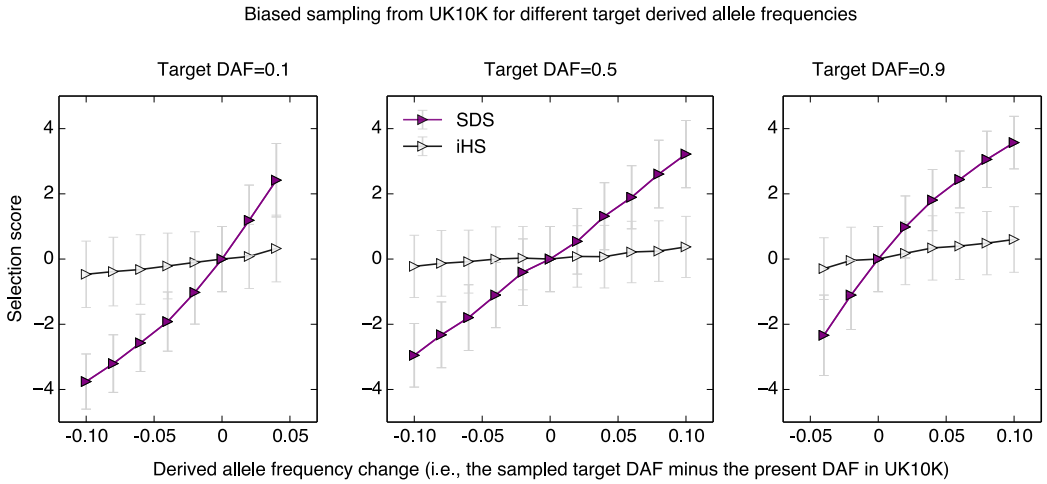

Figure S8: Extended simulations of selection from real present variation. We considered the UK10K genotype data for the 3,195 individuals that passed our quality control. For each pair of present and target derived allele frequencies, we randomly picked 1,000 SNPs with this present DAF in UK10K and subsampled them with 1,500 individuals each to the target DAF. SDS and iHS were ran on these biased subsamples, and the results were standardized relative to the unbiased sampling case (i.e., when present DAF equals the target DAF). Shown here are results (mean ± SD) for three target derived allele frequencies (left to right): 0.1, 0.5, and 0.9. This simulation figure thus extends on the simulation presented in Figure 2D by showing results for additional target frequencies. For more simulation details see Methods section: *Simulation of selection from present variation*.

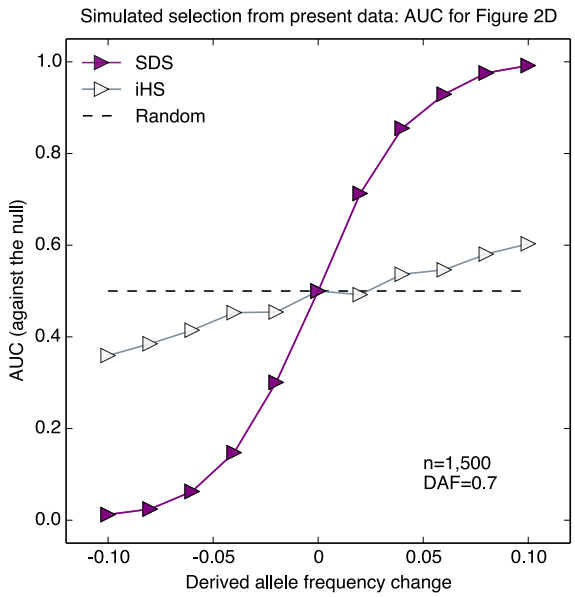

Figure S9: ROC analysis for the simulations of selection from real present variation. We considered the UK10K genotype data for the 3,195 individuals that passed our quality control. For each pair of present and target derived allele frequencies, we randomly picked 1,000 SNPs with this present DAF in UK10K and subsampled them with 1,500 individuals each (without replacement) to the target DAF. SDS and iHS were computed for these biased subsamples. Shown here is a ROC analysis view of the simulations presented in Figure 2D (present DAF=0.7). For each target DAF, shown is the area under the ROC curve (AUC) for discriminating the 1,000 simulations of this target DAF against the 1,000 simulations of unbiased sampling (i.e., when present DAF equals the target DAF). For more simulation details see Methods section: *Simulation of selection from present variation*.

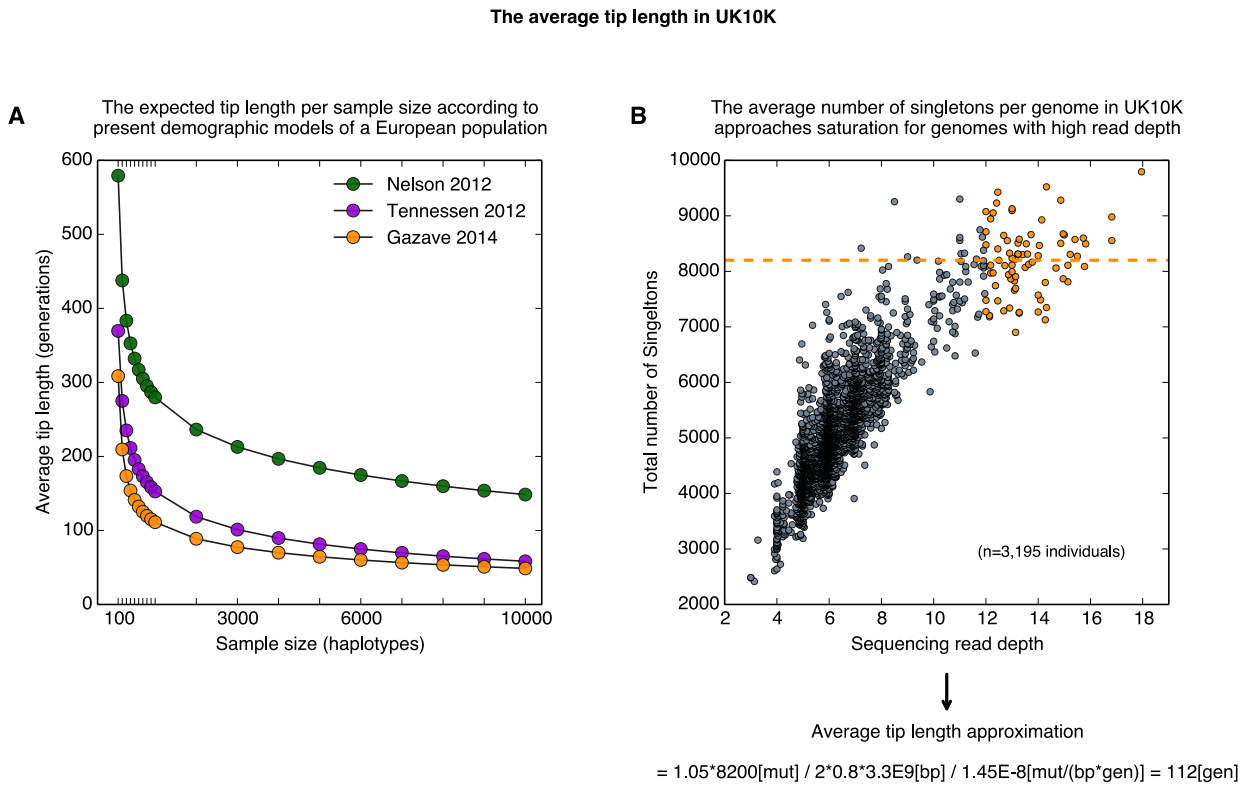

Figure S10: Estimating the mean tip-branch length in the UK10K data. (A) The expected mean tip length under neutral drift for samples of varying sizes, shown for the three recent models of European demography. These expectations were computed using an analytic formula (*61*) (analogous results in Figure 2A are based on simulations). The mean tip length for a sample of 3,000 individuals (2n=6,000 haplotypes) for the Tennessen model (*14*), which was used throughout this paper for simulations, is 74.9 generations. A model by Nelson (*38*) predicts a much longer time of 174.8 generations; but most recent model by Gazave (*39*) gives an even shorter time estimate of 60.1 generations. **(B)** Total number of singletons for the 3,195 UK10K individuals used in this work vs. the individual sequencing read depth. Over all individuals, mean sequencing depth is 6.37±0.03; mean total number of singletons is 5,130±17; and in that regime there is a linear relationship between them. However, the increase in total number of singletons seems to saturate for individuals with highest sequencing depth. The top 78 individuals with highest sequencing depth (≥12; colored orange) have little dependence on sequencing depth, suggesting they are in the regime that approaches saturation. The mean total number of singletons for these 78 individuals is 8,200±70. We use this to give a gross direct estimate of the mean tip length in the UK10K sample. We assume that true saturation of average number of singletons would be ~5% higher; that with saturated read depth, variation could be determined for ~80% of the genome; as well as a mutation rate per generation of 1.45×10^{-8} (*40*, *41*). This gives an estimated mean tip length of ~112 generations. In Methods section *Estimating the mean tip-branch length in the UK10K data* we discuss another two reported estimates, and conclude that the mean tip length is likely on the order of 2,000-3,000 years.

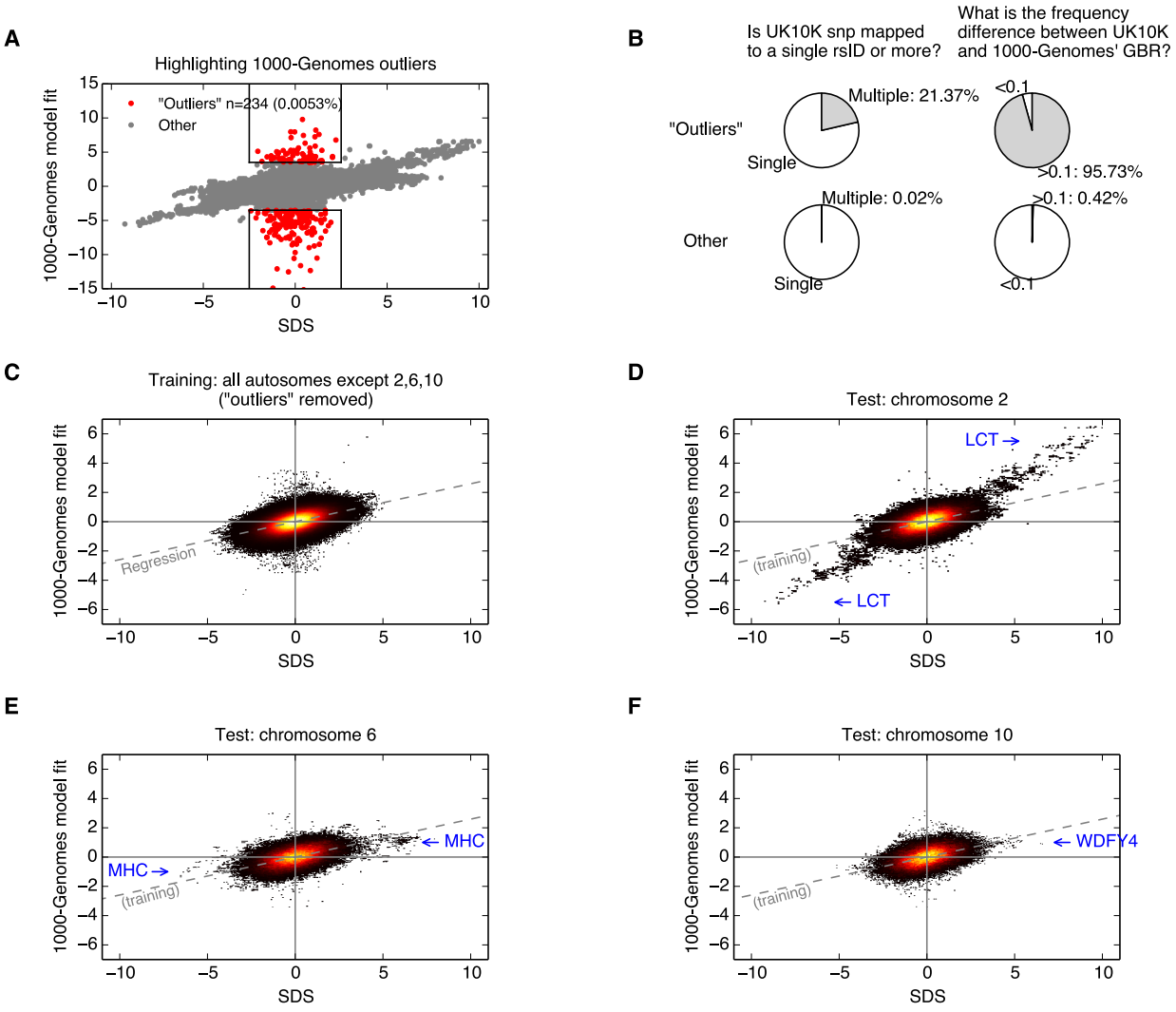

Figure S11: Extended validation of SDS for UK10K using 1,000-Genomes population frequencies. We built a linear regression model to predict SDS in UK10K based on the allele frequencies in the 26 1000-Genomes populations. Most outlier SNPs that have high predicted SDS but low observed SDS are likely due to inaccuracies of the 1000 Genomes estimates as they are highly enriched within the set of SNPs that is in dispute between UK10K and 1000-Genomes GBR, in terms of both mapping **(A)** and frequency estimates **(B)**. We fitted the regression model for SDS given population frequencies without using the three chromosomes with genome-wide significant SDS signals (i.e., 2, 6, and 10) **(C)**. Nevertheless, this model predicts well the linear trend with SDS for those unseen test chromosomes (d-f), and furthermore shows that the extreme SDS signals at *LCT* **(D)**, MHC **(E)**, and *WDFY4* **(F)** are broadly consistent with frequency differences between populations.

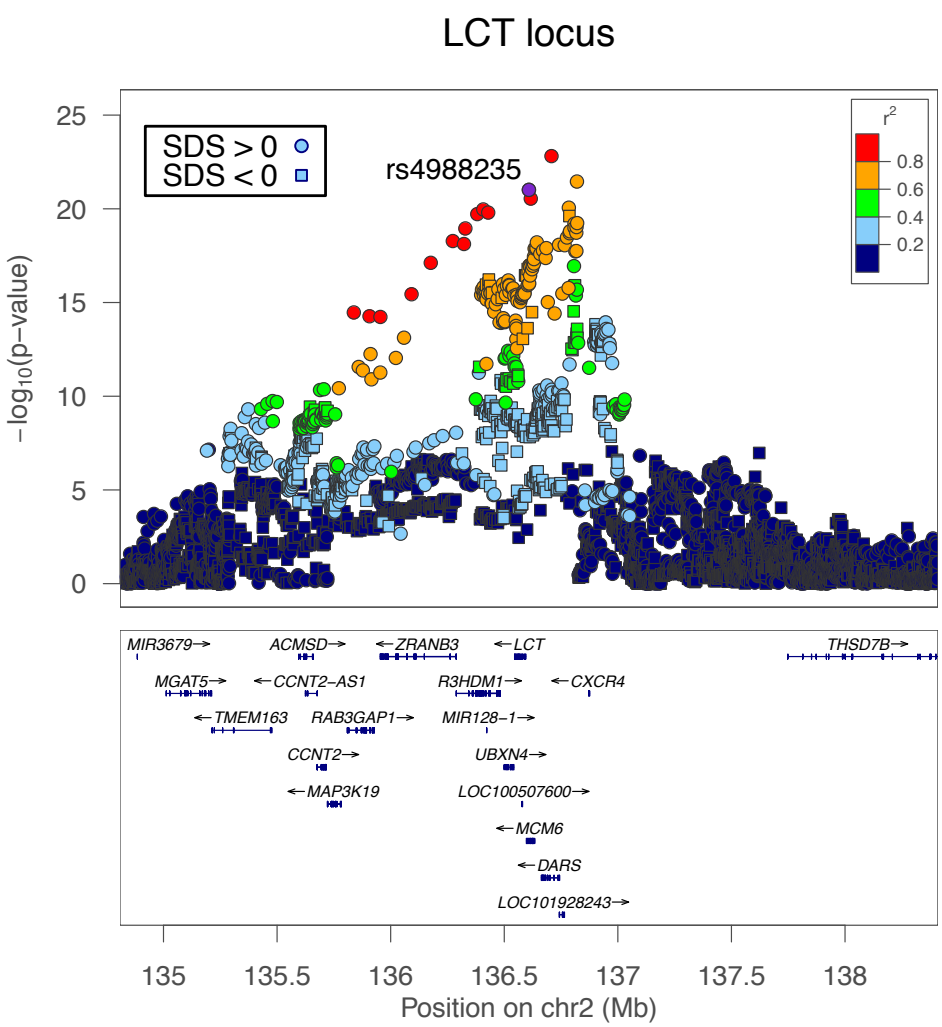

Figure S12: LocusZoom plot of the lactase (*LCT*) gene region. Causal SNP for lactase persistence is the index SNP, in purple. SDS shows massive signal along a multi-Mb region, concentrated towards the 5' UTR of lactase. Derived alleles are largely increasing (SDS > 0) suggesting that many young variants have been pulled up along with the lactase persistence allele.

SDS for UK10K for variants detected as under recent selection in Europe based on ancient DNA

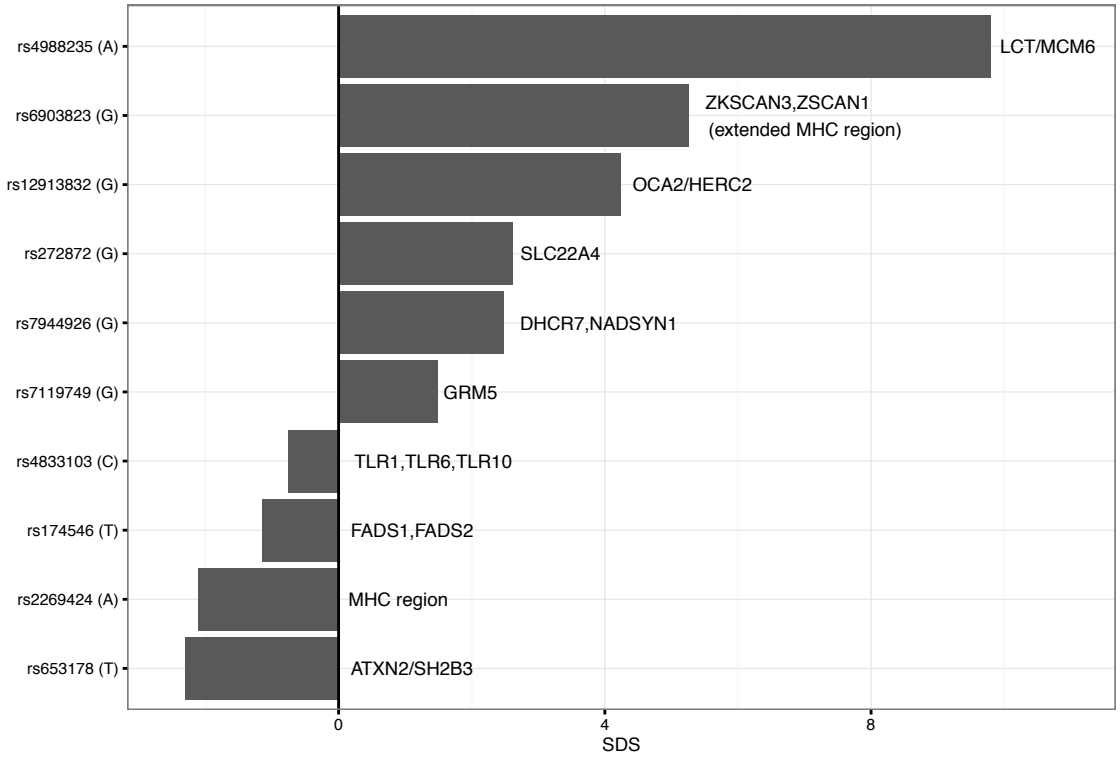

Figure S13: Comparison of SDS for UK10K with selection signals identified in Mathieson et al (2015). Shown are SDS scores for SNPs detected as recently selected in Europe by a study of ancient DNA of 230 Eurasians (*12*). SDS is directed here relative to the reported selected allele, so large positive SDS indicates agreement about selection signals between these studies.

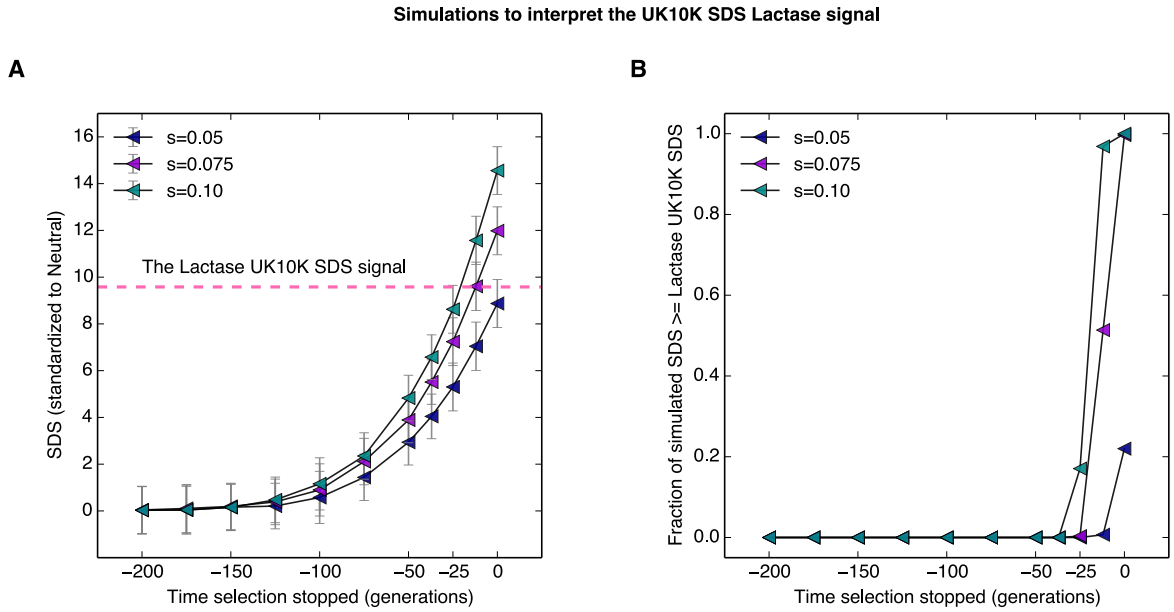

Figure S14: Interpreting the timescale and selection strength of the lactase signal using simulations. (A) SDS signal (mean ± SD) for 300 simulations of a strong selection (s=5%,7.5%,10%) that stopped x generations before the present, and reached allele frequency of 0.7; for a sample of 3,000 individuals and the Tennessen demographic model of Europe (*14*). The SDS signal observed for the known lactase-persistence allele is shown in dashed pink line. **(B)** For the simulations in (A), shown is the fraction of the 300 simulations that result in SDS value greater or equal to the observed lactase signal in the UK10K data. For selection that stopped any time ≥37 generation from the present, and for any selective pressure of ≤10%, all 300 simulations gave SDS signal strictly lower than we observed for the lactase allele in the UK10K data. As we elaborate in Methods section: *Timescale of selection on lactase*, this suggests that under conservative assumptions (s≤10% and mean tip length in UK10K ≤4,000 years), we can estimate that it is highly unlikely ($p<1/300$) that selection on the lactase-persistent allele did not enter the past 2,000 years.

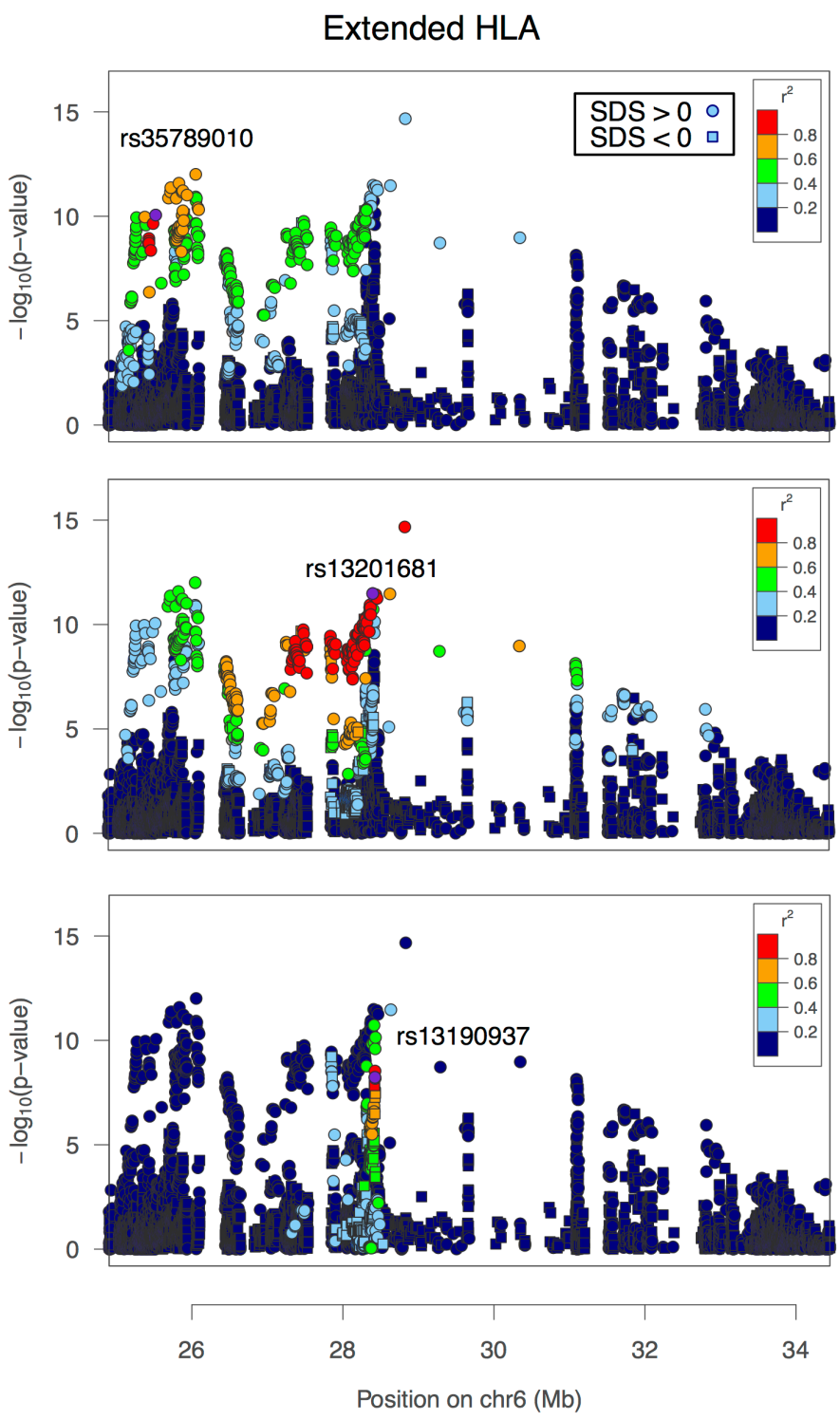

Figure S15: LocusZoom plot of the extended MHC region. Analysis of different index SNPs suggests at least three independent targets of selection within the extended MHC.

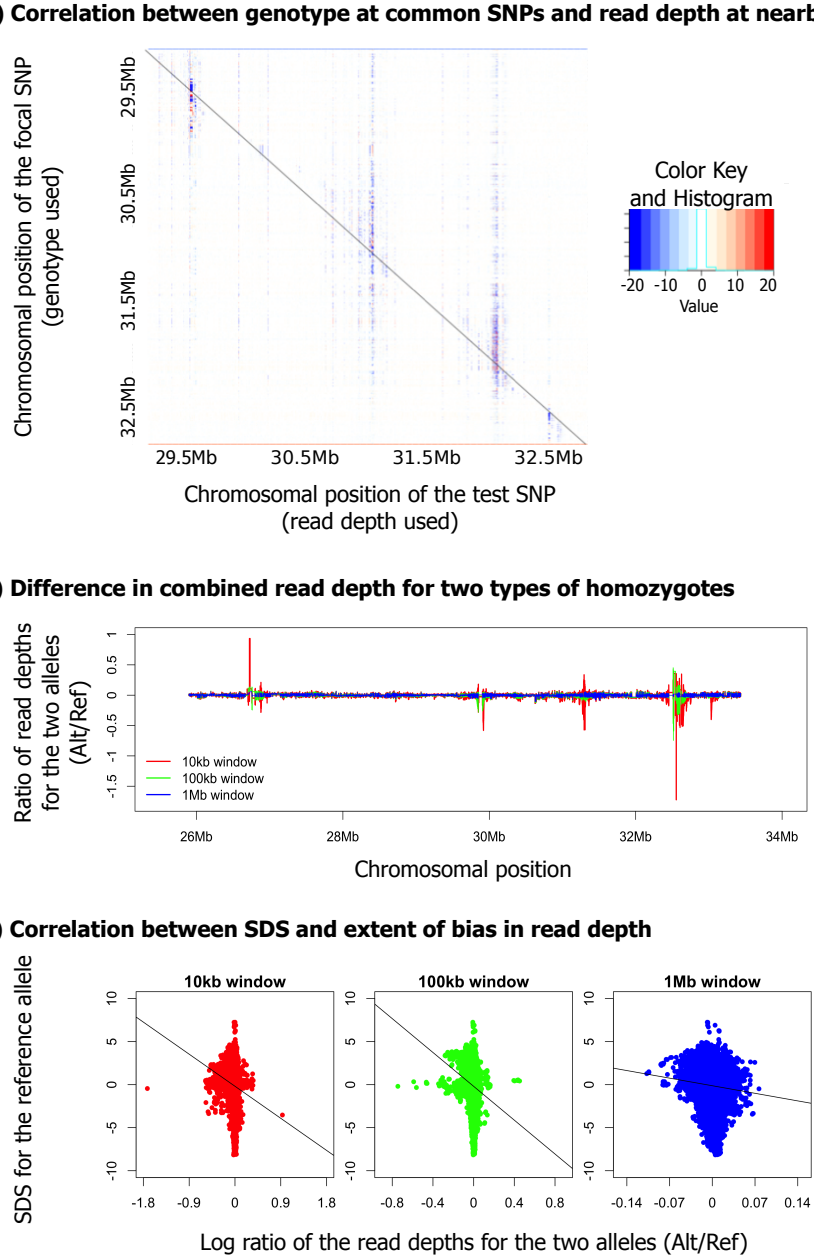

Figure S16: Reference bias in read depth in the MHC region and its effects on SDS. (A) Correlation between genotype at common SNPs (MAF>0.05) and read depth at nearby SNPs. Red and blue represent positive and negative correlations respectively, and the shade of color represents the -log(P-value) of Spearman's rank correlation. For illustration purpose, only part of the MHC region is shown (chr6: 29,500,000-33,000,000). **(B)** Difference in combined read depth between two types of homozygotes in 10kb, 100kb and 1Mb windows. **(C)** Correlation between SDS for the reference allele and log ratio of combined read depths for the two alleles (alternative over reference). Notice that although some positions have biased read mapping, there is not a strong effect that some genotypes are associated with much higher read depths across a large region (A). Moreover, the largest SDS scores are at SNPs with little or no difference in read depths between genotypes, arguing that reference bias does not drive the biggest signals.

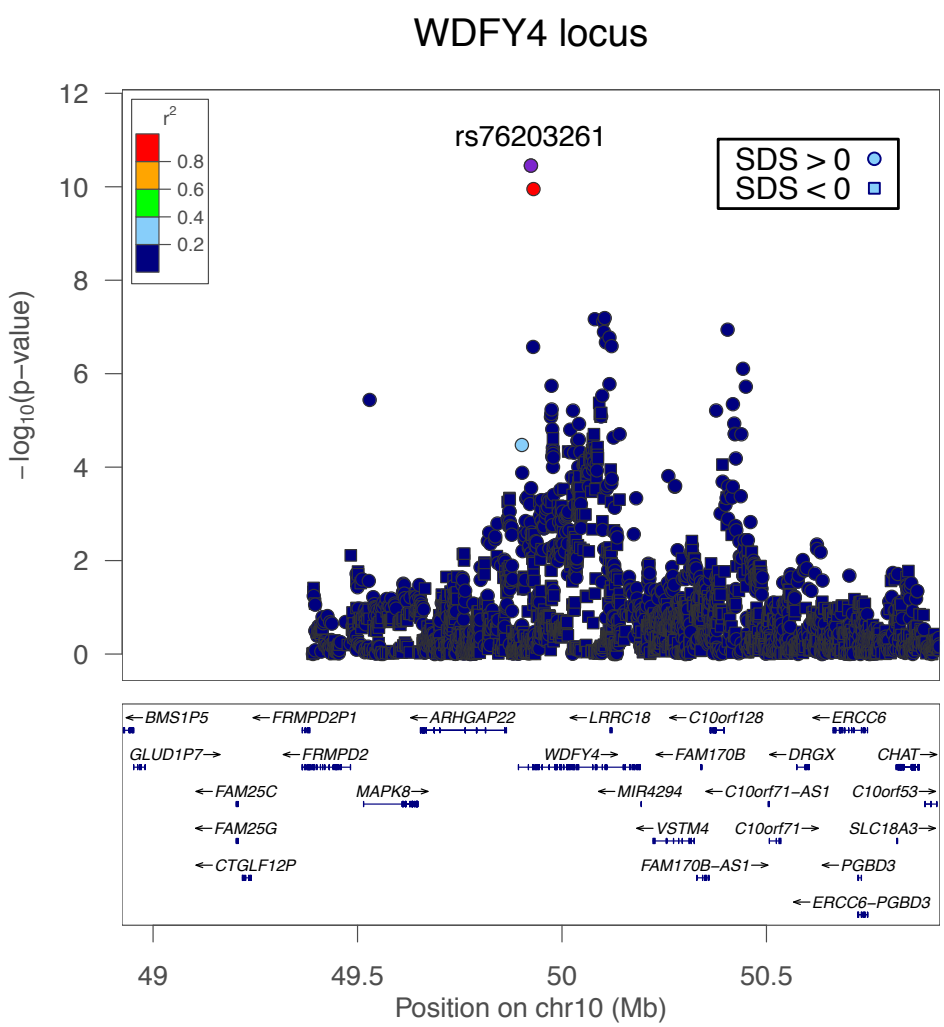

Figure S17: LocusZoom plot of the *WDFY4* gene region. The causal gene for the locus is unknown; the greatest SDS SNP was chosen as the index SNP. The two genome-wide significant hits are tightly linked to each other. Interestingly, additional SNPs in the region approach genome-wide significance despite being unlinked.

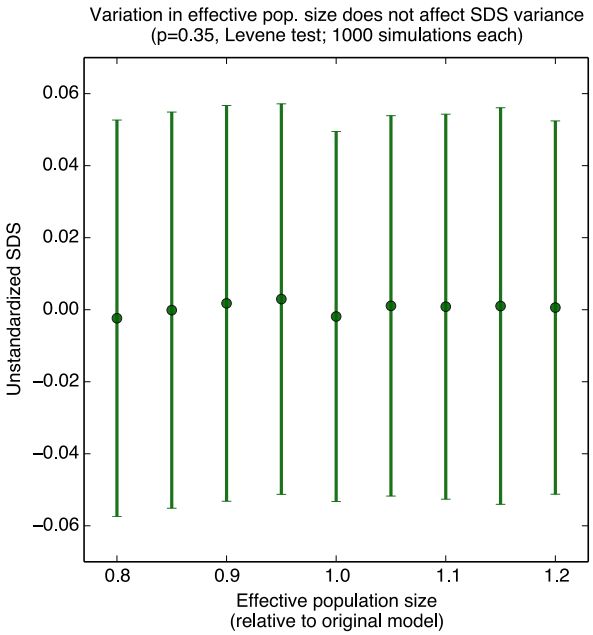

Figure S18: Background selection does not inflate SDS variance. Background selection varies along the genome and was shown to be associated with GWAS-catalog variants (*62*). One might wonder whether our finding that GWAS-catalog variants have inflated SDS variance is an indirect result of some dependence of SDS variance on background selection. Variation in background selection is commonly modeled as a regional variation in the effective population size (*55*, *63*). (This is an approximate model of background selection; however more explicit models are computationally infeasible at the scale required for SDS simulations.) Shown here are the unstandardized SDS values (mean ± SD) for 1,000 neutral simulations of present DAF 0.7, using the Tennessen demographic model (*14*) whose effective population size was scaled by up to a 20% increase or 20% decrease. Equality of variance over the different effective population sizes is not rejected (p=0.35; Leven test). Thus, background selection is unlikely to be the cause for the association of GWAS-catalog variants with increased SDS variance.

Further information for the pigmentation variants shown in Figure 3C

ID	GENE	CHR	POS	SDS	FIRST AUTHOR	PUBMED ID	PVALUE	DAF (UK10K)
rs12821256	*KITLG*	12	89328335	4.6001293	Sulem P	17952075	4e-30	0.119
rs12913832	*OCA2/HERC2*	15	28365618	4.0750304	Eriksson N	20585627	1e-300	0.770
rs619865	*ASIP*	20	33867697	2.8031028	Eriksson N	20585627	5e-14	0.886
rs12896399	*SLC24A4*	14	92773663	2.4172635	Sulem P	17952075	1e-48	0.446
rs1393350	*TYR*	11	89011046	1.4666560	Jin Y	20410501	2e-18	0.273
rs1408799	*TYRP1*	9	12672097	1.4084269	Sulem P	18488028	6e-17	0.695
rs975739	*EDNRB*	13	78381146	1.4204976	Zhang M	23548203	2e-14	0.603
rs2153271	*BNC2*	9	16864521	1.1036386	Eriksson N	20585627	4e-10	0.606
rs916977	*OCA2/HERC2*	15	28513364	1.1088416	Kayser M	18252221	1e-43	0.850
rs1003719	*TTC3*	21	38491095	1.0790324	Liu F	20463881	2e-10	0.559
rs1667394	*OCA2/HERC2*	15	28530182	1.0656963	Sulem P	17952075	2e-53	0.841
rs35264875	*TPCN2*	11	68846399	0.2078374	Sulem P	18488028	4e-30	0.152
rs1015362	*ASIP*	20	32738612	-0.4427474	Sulem P	18488028	6e-37	0.282
rs13289810	*TYRP1*	9	12396731	-0.5271060	Kenny EE	22556244	1e-19	0.350

Figure S19: Analysis of SDS signal at pigmentation variants. Information about the 14 pigmentation variants presented in Figure 3C. Genome coordinates are in hg19. The referenced manuscript and p-value indicate studies that reported the association of the variant to the pigmentation trait. For more details about literature curation see Methods section: *Analysis of high SDS enrichment for pigmentation variants*.

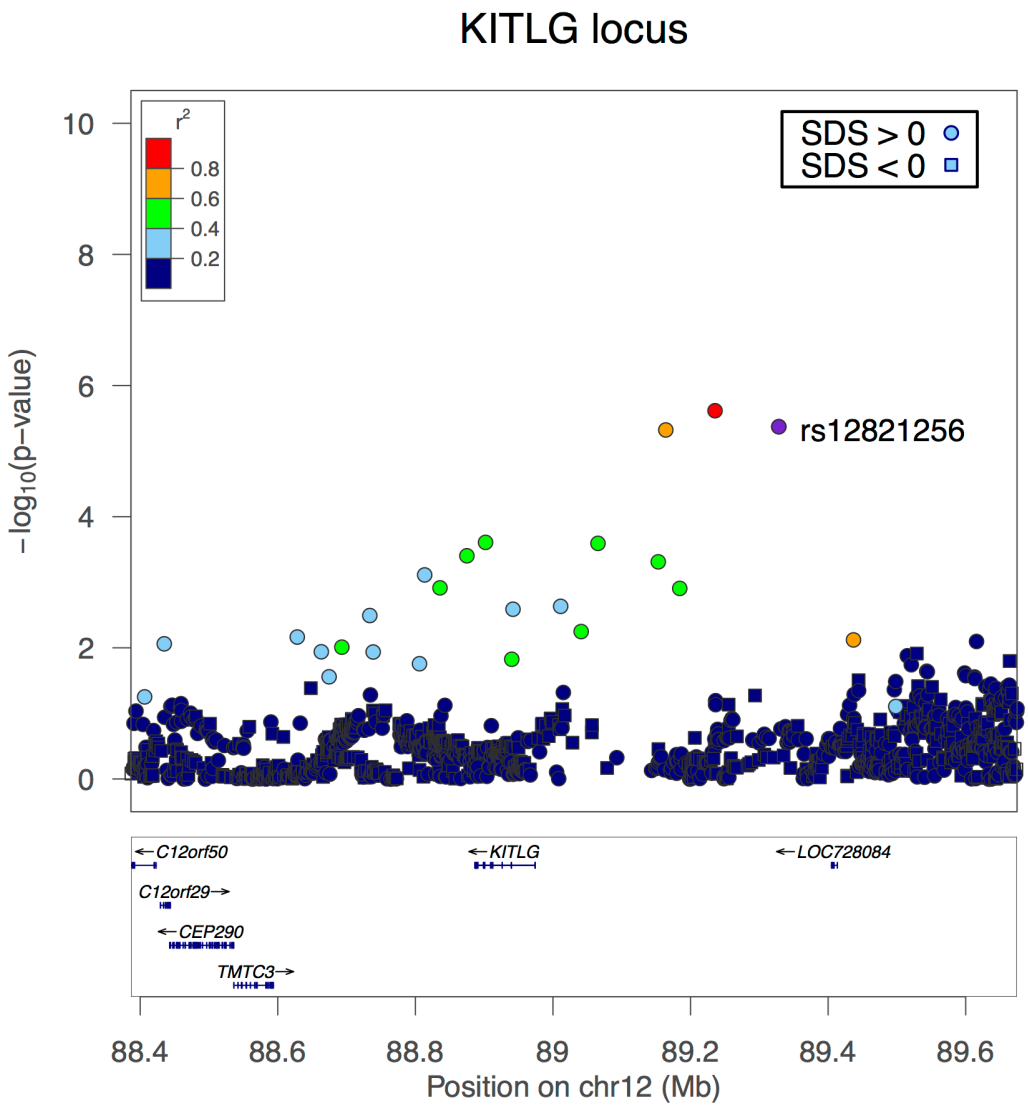

Figure S20: LocusZoom plot of the *KITLG* gene region. The blond-hair allele rs12821256 has been chosen as the index SNP. SDS scores suggest increase in haplotype frequency without achieving genome-wide significance.

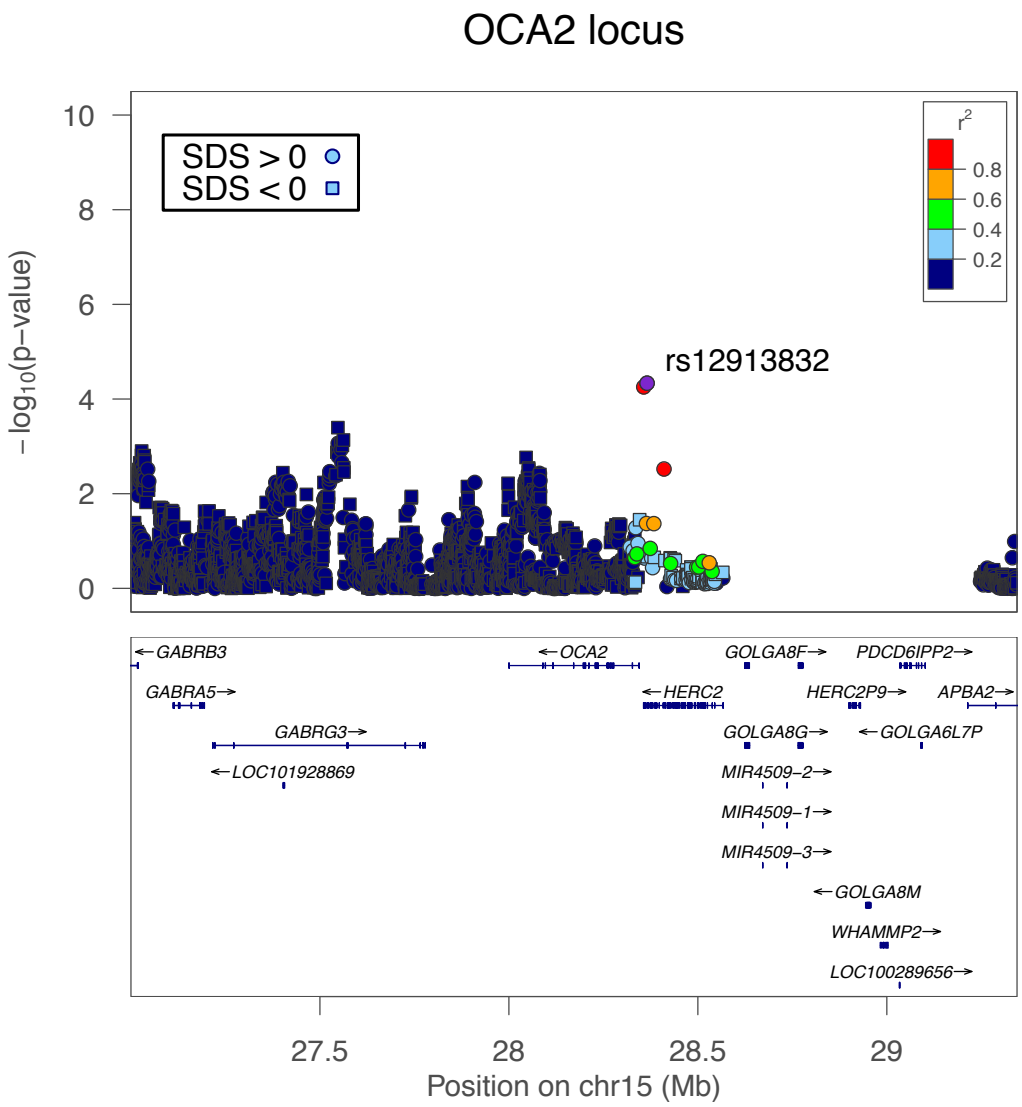

Figure S21: LocusZoom plot of the *OCA2* gene region. rs12913832, implicated in eye color, has been chosen as the index SNP. SDS scores suggest increase in haplotype frequency without achieving genome-wide significance.

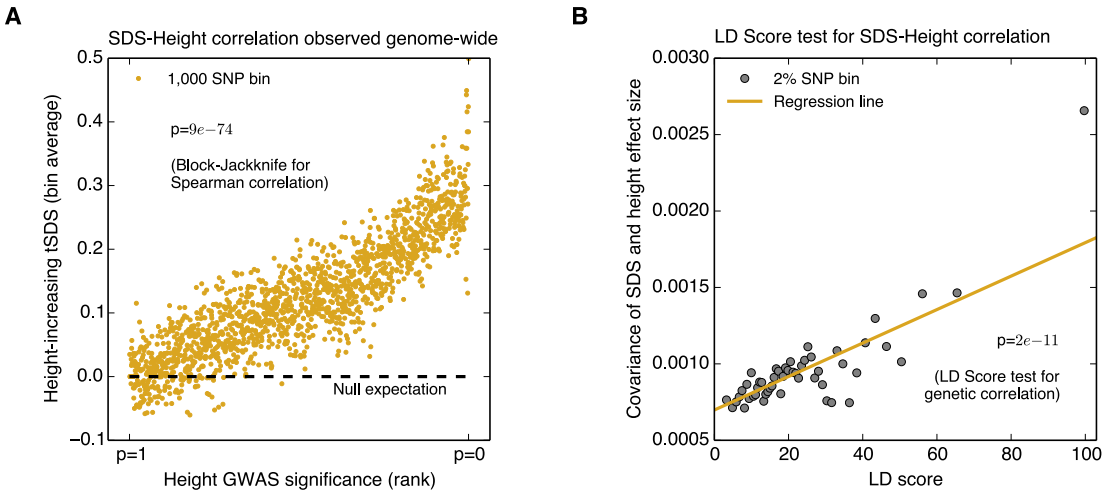

Figure S22: SDS signal for polygenic adaptation for increased height using GWAS meta analysis data. (A) Mean tSDS of SNPs, where tSDS is polarized according to the estimated direction of effect of each SNP on height in a recent GWAS meta analysis (*21*). The x-axis is ordered from least significant SNPs (p~1) to most significant (p~0) and SNPs are placed into bins of 1,000 consecutive SNPs (ρ=0.078, p=9x10^{-74}). **(B)** Covariance of height effect size (Z-scores) and SDS, as a function of LD score (p=3x10^{-17} using the LD score test for genetic correlation (*25*)). These plots are analogous to Figure 4A,B, but using the GWAS meta analysis instead of the family-based data.

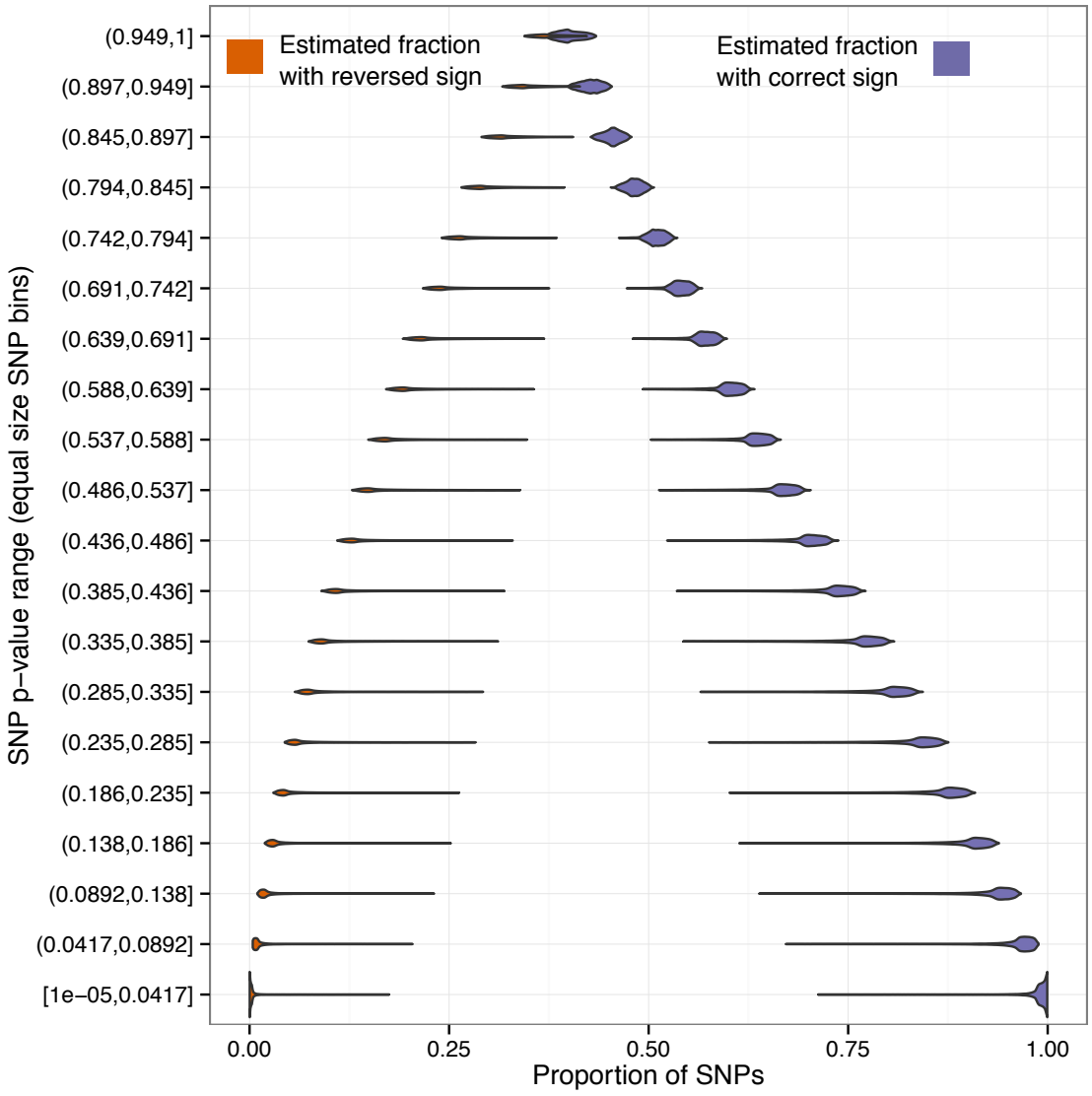

Figure S23: Estimated false sign rates in family-based height GWAS. The rows show 5% bins of SNPs (sorted by p-value). For each SNP, ashR reports a probability that the estimated sign is reversed (e.g., positive when it should be negative) [left-hand violins]; correct [right-hand violins]; or zero [not shown]. For the most significant SNPs (bottom row), ashR estimates that nearly all SNPs reflect true effects with correctly estimated sign. SNPs with p-values as high as 0.75 appear to reflect true signal more often than not.

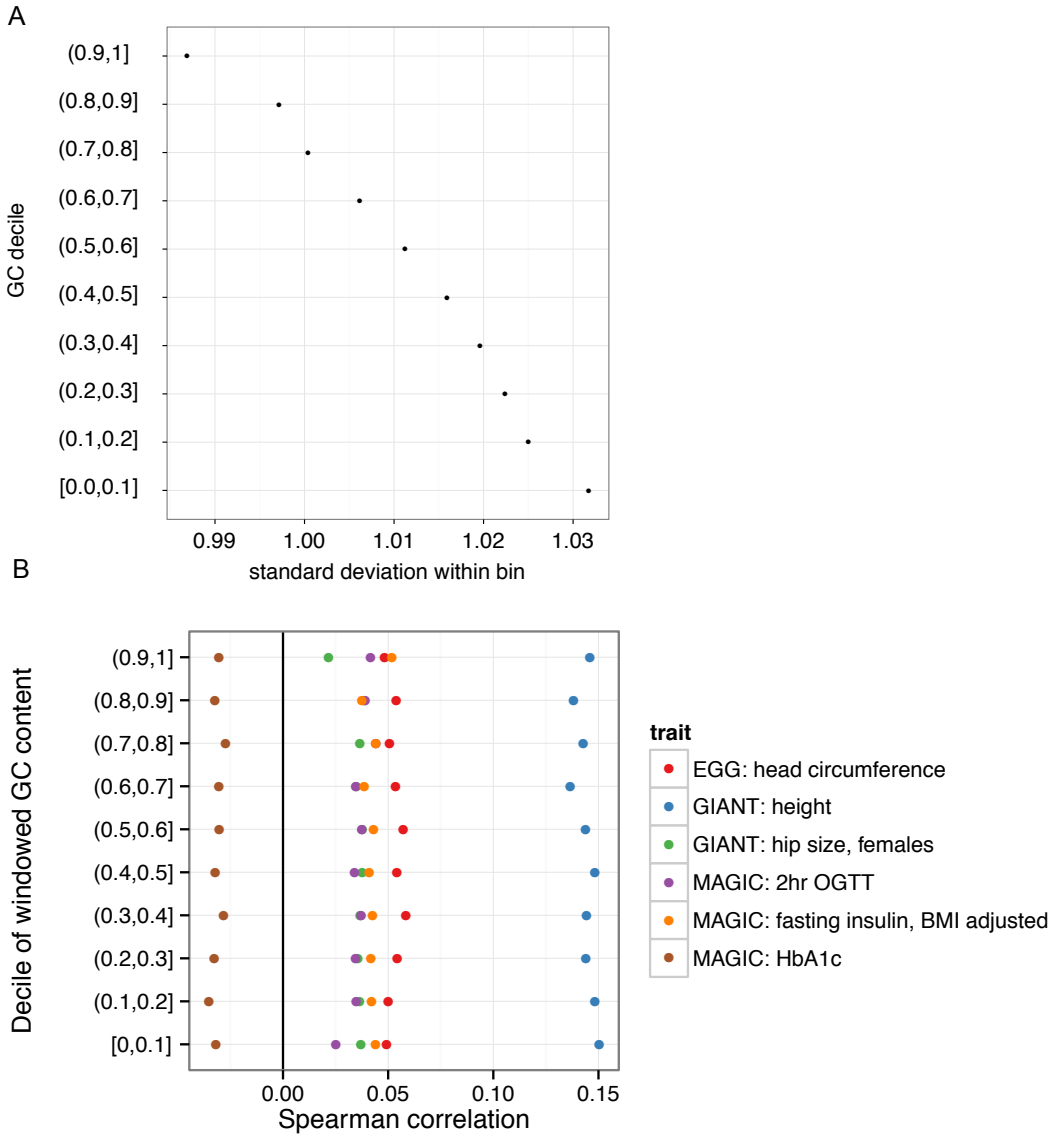

Figure S24: SDS associations as a function of GC content. (A) Lower GC content regions (measured in 150bp bins around the test SNP) possess systematically greater SDS variance, although the effect of this is small. **(B)** GC content-associated differences appear to have little effect on GWAS Z-score correlation testing. The SDS vs. GWAS Z-score correlations vary little across the different partitions of GC content.

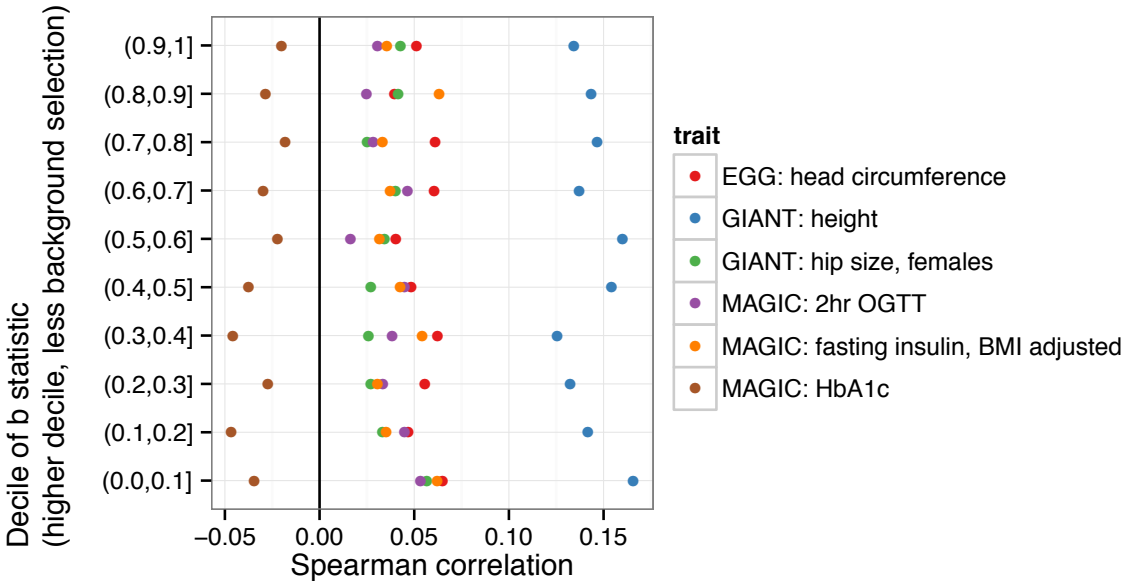

Figure S25: SDS associations partitioned by background selection. SDS signal is apparent across different partitions of susceptibility to background selection. SDS correlations with GWAS Z-scores for a selection of traits are only marginally stronger for functionally dense regions as determined by the B-statistic than functionally poor regions.

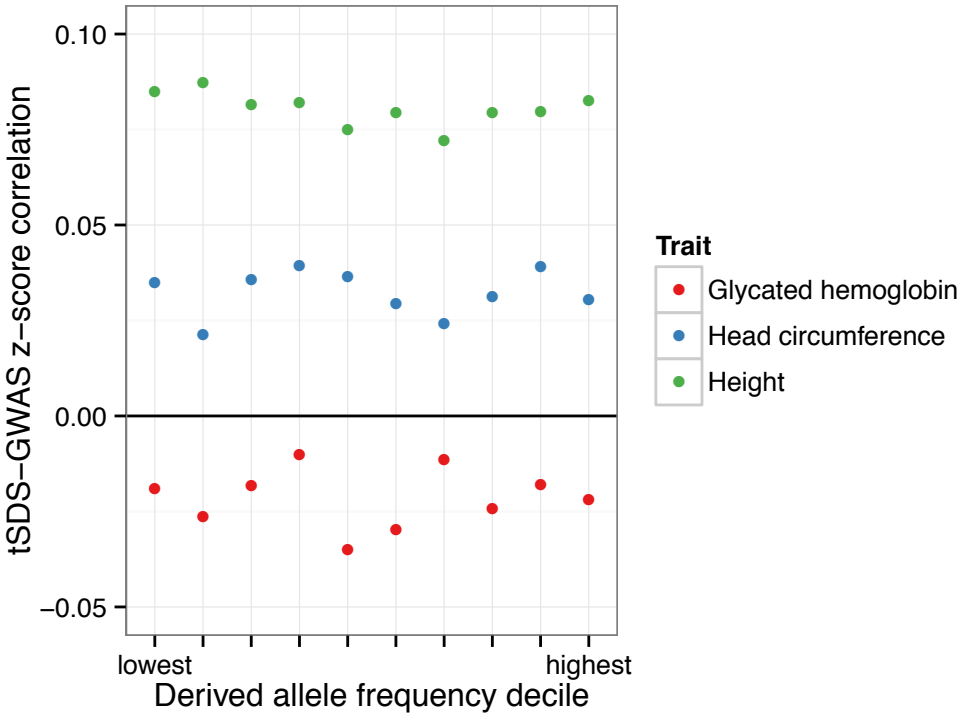

Figure S26: SDS associations partitioned by derived allele frequency. SDS correlations with GWAS Z-scores are stable across the spectrum of derived allele frequencies.

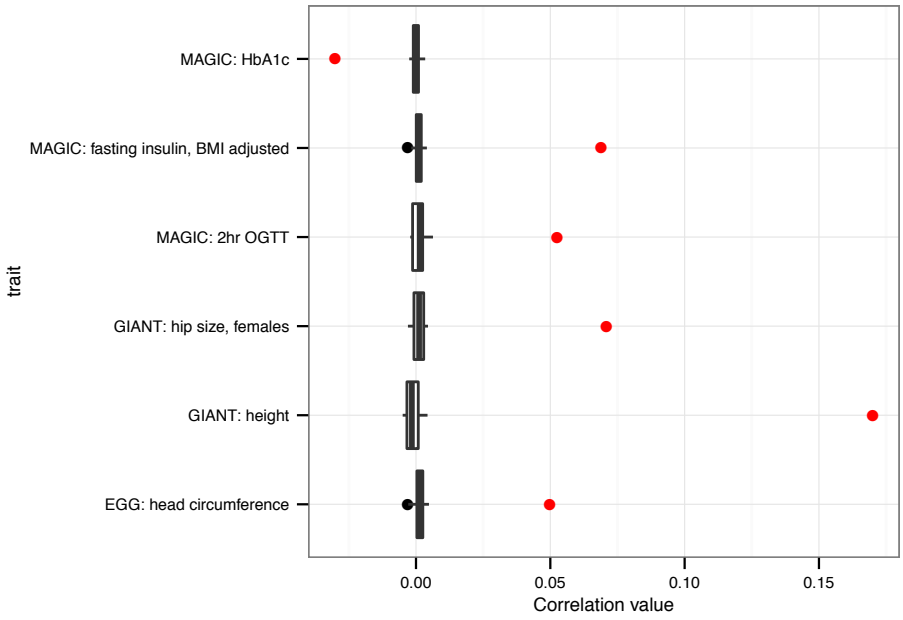

Figure S27: SDS associations vs. permutations that preserve derived allele frequency and B-statistic. The correlations between GWAS Z-scores and tSDS scores were compared to correlations obtained for SNP permutations that preserve derived allele frequency and B-statistic – a proxy for background selection and a known correlate of functional density (*55*). SNPs were placed into bins on a two-dimensional grid defined by 20 equally sized DAF intervals and 10 equally sized B-statistic intervals. SDS values within each of the 200 grid bins were permuted such that the SNPs retained their contiguity. The observed GWAS-tSDS correlations are shown in red. Correlations obtained for 19 SNP permutations, in black (box-plot and outer quartile outliers). Permuted values lead to correlations tightly distributed around zero, and the observed GWAS-tSDS associations remain highly significant by this test.

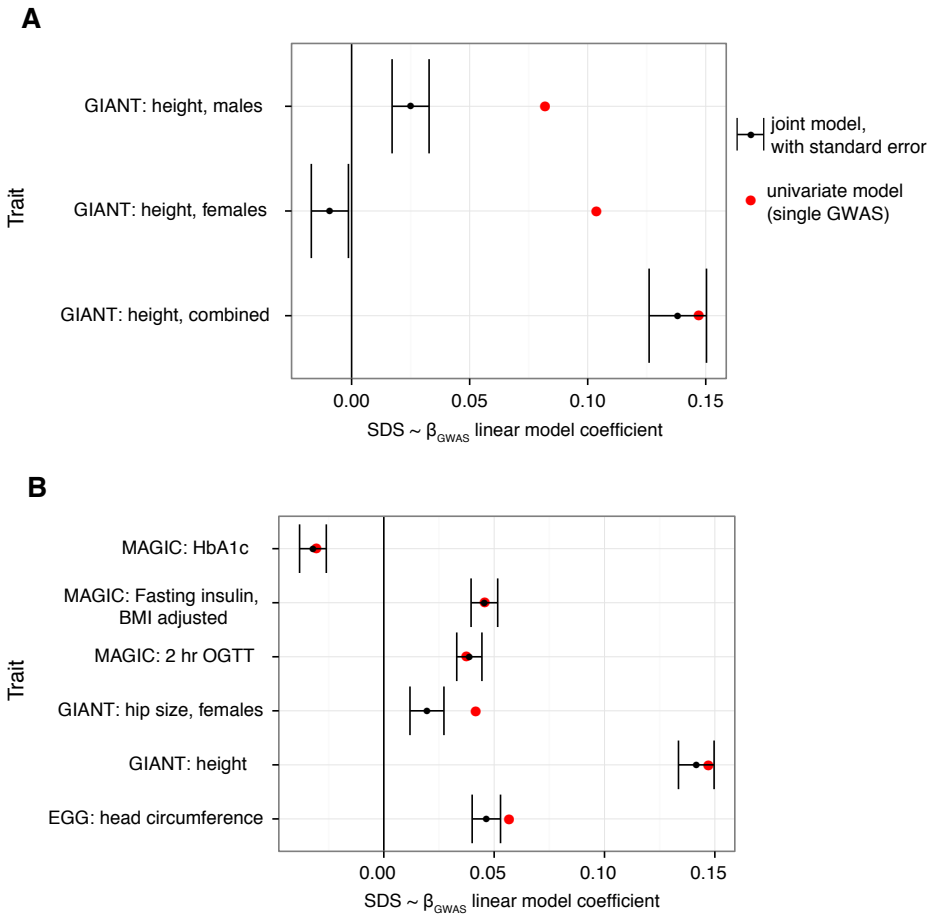

Figure S28: Height does not underlie other top signals in SDS associations. (A) SDS suggests selective pressures have been increasing height polygenic score in both men and women in Britain when each trait is tested individually (red dots). When all three height GWAS summary statistics are used to predict GWAS, most of the signal evaporates and only the meta-analysis holds close to the original effect size (black dots with standard error bars). **(B)** When different candidate traits for selection are used to predict SDS score jointly, the contributions of each of the top candidate traits largely persist.

Trait	Spearman rho	Jackknife p-value	LD score p-value
Height	0.078	9.59E-74	1.02E-11
Height (M)	0.059	1.74E-58	9.80E-11
Height (F)	0.053	7.92E-46	3.26E-06
BMI (M)	-0.022	2.74E-08	2.60E-02
BMI (F)	0.002	5.84E-01	1.45E-02
Hip size (M)	0.009	2.19E-02	2.43E-01
Hip size (F)	0.022	2.43E-09	5.00E-04
Waist circumference (M)	0.010	1.07E-02	8.70E-02
Waist circumference (F)	0.021	9.13E-08	6.50E-03
Waist to hip ratio (M)	0.013	9.28E-04	2.07E-01
Waist to hip ratio (F)	0.008	4.40E-02	2.23E-01
Infant head circumference	0.031	2.15E-17	8.50E-03
Birth weight	0.014	4.87E-05	2.85E-02
Birth length	0.010	2.20E-03	2.75E-02
Childhood obesity	-0.004	2.74E-01	9.21E-01
Age at menarche (F)	0.018	1.35E-06	7.50E-03
Growth spurt (M)	0.011	3.36E-03	7.50E-03
Growth spurt (F)	0.001	7.02E-01	5.66E-01
Tanner stage (M)	0.005	1.98E-01	3.81E-01
Tanner stage (F)	-0.008	2.37E-02	4.85E-02
Pubertal growth (M)	0.005	1.49E-01	3.49E-01
Pubertal growth (F)	0.004	2.31E-01	5.15E-01
Fasting insulin (BMI adj)	0.026	1.98E-14	1.30E-02
Two-hour OGTT	0.023	4.34E-10	3.00E-02
Type 2 Diabetes	-0.023	1.12E-10	2.08E-01
HbA1c	-0.023	8.67E-12	4.00E-03
Corrected insulin response	-0.012	1.16E-03	1.85E-02
HOMA beta cell function	0.011	2.32E-03	9.95E-02
Fasting proinsulin levels	-0.009	1.56E-02	9.79E-01
HOMA insulin resistance	0.008	2.56E-02	2.19E-01
Fasting glucose (BMI adj)	-0.008	2.76E-02	1.01E-01
LDL	-0.018	7.50E-06	6.40E-02
Total cholesterol	-0.014	5.75E-04	7.50E-03
HDL	-0.004	3.24E-01	4.05E-02
Triglycerides	-0.001	7.82E-01	1.88E-01
Crohn's disease	0.014	1.12E-04	3.33E-01
Ulcerative colitis	-0.002	4.92E-01	5.07E-01
Rheumatoid arthritis	0.002	5.58E-01	5.01E-01
Epilepsy	0.012	1.69E-04	2.29E-01
Schizophrenia	-0.013	1.84E-03	4.27E-01
Bipolar disorder	-0.007	9.80E-02	3.53E-01
Alzheimer's Disease	0.002	5.41E-01	4.30E-01
Years of education	0.009	1.79E-02	7.95E-02
Coronary artery disease	0.001	7.74E-01	2.87E-01

Table S1: Summary of GWAS correlation testing with SDS. Positive values suggest trait-increasing alleles are associated with recent selection. LD score genetic correlation p-values reported are one-sided for confirming the direction of effect in the raw correlation.

References and Notes

1. J. J. Vitti, S. R. Grossman, P. C. Sabeti, Detecting natural selection in genomic data. *Annu. Rev. Genet.* **47**, 97–120 (2013). doi:10.1146/annurev-genet-111212-133526 Medline

2. T. Bersaglieri, P. C. Sabeti, N. Patterson, T. Vanderploeg, S. F. Schaffner, J. A. Drake, M. Rhodes, D. E. Reich, J. N. Hirschhorn, Genetic signatures of strong recent positive selection at the lactase gene. *Am. J. Hum. Genet.* **74**, 1111–1120 (2004). doi:10.1086/421051 Medline

3. R. L. Lamason, M. A. Mohideen, J. R. Mest, A. C. Wong, H. L. Norton, M. C. Aros, M. J. Jurynec, X. Mao, V. R. Humphreville, J. E. Humbert, S. Sinha, J. L. Moore, P. Jagadeeswaran, W. Zhao, G. Ning, I. Makalowska, P. M. McKeigue, D. O'donnell, R. Kittles, E. J. Parra, N. J. Mangini, D. J. Grunwald, M. D. Shriver, V. A. Canfield, K. C. Cheng, SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science* **310**, 1782–1786 (2005). doi:10.1126/science.1116238 Medline

4. X. Yi, Y. Liang, E. Huerta-Sanchez, X. Jin, Z. X. P. Cuo, J. E. Pool, X. Xu, H. Jiang, N. Vinckenbosch, T. S. Korneliussen, H. Zheng, T. Liu, W. He, K. Li, R. Luo, X. Nie, H. Wu, M. Zhao, H. Cao, J. Zou, Y. Shan, S. Li, Q. Yang, P. Asan, P. Ni, G. Tian, J. Xu, X. Liu, T. Jiang, R. Wu, G. Zhou, M. Tang, J. Qin, T. Wang, S. Feng, G. Li, J. Huasang, J. Luosang, W. Wang, F. Chen, Y. Wang, X. Zheng, Z. Li, Z. Bianba, G. Yang, X. Wang, S. Tang, G. Gao, Y. Chen, Z. Luo, L. Gusang, Z. Cao, Q. Zhang, W. Ouyang, X. Ren, H. Liang, H. Zheng, Y. Huang, J. Li, L. Bolund, K. Kristiansen, Y. Li, Y. Zhang, X. Zhang, R. Li, S. Li, H. Yang, R. Nielsen, J. Wang, J. Wang, Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* **329**, 75–78 (2010). doi:10.1126/science.1190371 Medline

5. B. F. Voight, S. Kudaravalli, X. Wen, J. K. Pritchard, A map of recent positive selection in the human genome. *PLOS Biol.* **4**, e72 (2006). doi:10.1371/journal.pbio.0040072 Medline

6. J. Hermisson, P. S. Pennings, Soft sweeps: Molecular population genetics of adaptation from standing genetic variation. *Genetics* **169**, 2335–2352 (2005). doi:10.1534/genetics.104.036947 Medline

7. M. Przeworski, G. Coop, J. D. Wall, The signature of positive selection on standing genetic variation. *Evolution* **59**, 2312–2323 (2005). doi:10.1554/05-273.1 Medline

8. J. K. Pritchard, J. K. Pickrell, G. Coop, The genetics of human adaptation: Hard sweeps, soft sweeps, and polygenic adaptation. *Curr. Biol.* **20**, R208–R215 (2010). doi:10.1016/j.cub.2009.11.055 Medline

9. M. C. Turchin, C. W. K. Chiang, C. D. Palmer, S. Sankararaman, D. Reich, J. N. Hirschhorn; Genetic Investigation of ANthropometric Traits (GIANT) Consortium, Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nat. Genet.* **44**, 1015–1019 (2012). doi:10.1038/ng.2368 Medline

10. G. Bhatia, N. Patterson, B. Pasaniuc, N. Zaitlen, G. Genovese, S. Pollack, S. Mallick, S. Myers, A. Tandon, C. Spencer, C. D. Palmer, A. A. Adeyemo, E. L. Akylbekova, L. A. Cupples, J. Divers, M. Fornage, W. H. L. Kao, L. Lange, M. Li, S. Musani, J. C. Mychaleckyj, A. Ogunniyi, G. Papanicolaou, C. N. Rotimi, J. I. Rotter, I. Ruczinski, B. Salako, D. S. Siscovick, B. O. Tayo, Q. Yang, S. McCarroll, P. Sabeti, G. Lettre, P. De

Jager, J. Hirschhorn, X. Zhu, R. Cooper, D. Reich, J. G. Wilson, A. L. Price, Genome-wide comparison of African-ancestry populations from CARe and other cohorts reveals signals of natural selection. *Am. J. Hum. Genet.* **89**, 368–381 (2011). doi:10.1016/j.ajhg.2011.07.025 Medline

11. S. Wilde, A. Timpson, K. Kirsanow, E. Kaiser, M. Kayser, M. Unterländer, N. Hollfelder, I. D. Potekhina, W. Schier, M. G. Thomas, J. Burger, Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 4832–4837 (2014). doi:10.1073/pnas.1316513111 Medline

12. I. Mathieson, I. Lazaridis, N. Rohland, S. Mallick, N. Patterson, S. A. Roodenberg, E. Harney, K. Stewardson, D. Fernandes, M. Novak, K. Sirak, C. Gamba, E. R. Jones, B. Llamas, S. Dryomov, J. Pickrell, J. L. Arsuaga, J. M. B. de Castro, E. Carbonell, F. Gerritsen, A. Khokhlov, P. Kuznetsov, M. Lozano, H. Meller, O. Mochalov, V. Moiseyev, M. A. R. Guerra, J. Roodenberg, J. M. Vergès, J. Krause, A. Cooper, K. W. Alt, D. Brown, D. Anthony, C. Lalueza-Fox, W. Haak, R. Pinhasi, D. Reich, Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015). doi:10.1038/nature16152 Medline

13. Information on materials and methods is available at *Science* Online.

14. J. A. Tennessen, A. W. Bigham, T. D. O'Connor, W. Fu, E. E. Kenny, S. Gravel, S. McGee, R. Do, X. Liu, G. Jun, H. M. Kang, D. Jordan, S. M. Leal, S. Gabriel, M. J. Rieder, G. Abecasis, D. Altshuler, D. A. Nickerson, E. Boerwinkle, S. Sunyaev, C. D. Bustamante, M. J. Bamshad, J. M. Akey; Broad GO; Seattle GO; NHLBI Exome Sequencing Project, Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337**, 64–69 (2012). doi:10.1126/science.1219240 Medline

15. K. Walter, J. L. Min, J. Huang, L. Crooks, Y. Memari, S. McCarthy, J. R. B. Perry, C. Xu, M. Futema, D. Lawson, V. Iotchkova, S. Schiffels, A. E. Hendricks, P. Danecek, R. Li, J. Floyd, L. V. Wain, I. Barroso, S. E. Humphries, M. E. Hurles, E. Zeggini, J. C. Barrett, V. Plagnol, J. B. Richards, C. M. T. Greenwood, N. J. Timpson, R. Durbin, N. Soranzo, S. Bala, P. Clapham, G. Coates, T. Cox, A. Daly, P. Danecek, Y. Du, R. Durbin, S. Edkins, P. Ellis, P. Flicek, X. Guo, X. Guo, L. Huang, D. K. Jackson, C. Joyce, T. Keane, A. Kolb-Kokocinski, C. Langford, Y. Li, J. Liang, H. Lin, R. Liu, J. Maslen, S. McCarthy, D. Muddyman, M. A. Quail, J. Stalker, J. Sun, J. Tian, G. Wang, J. Wang, Y. Wang, K. Wong, P. Zhang, I. Barroso, E. Birney, C. Boustred, L. Chen, G. Clement, M. Cocca, P. Danecek, G. Davey Smith, I. N. M. Day, A. Day-Williams, T. Down, I. Dunham, R. Durbin, D. M. Evans, T. R. Gaunt, M. Geihs, C. M. T. Greenwood, D. Hart, A. E. Hendricks, B. Howie, J. Huang, T. Hubbard, P. Hysi, V. Iotchkova, Y. Jamshidi, K. J. Karczewski, J. P. Kemp, G. Lachance, D. Lawson, M. Lek, M. Lopes, D. G. MacArthur, J. Marchini, M. Mangino, I. Mathieson, S. McCarthy, Y. Memari, S. Metrustry, J. L. Min, A. Moayyeri, D. Muddyman, K. Northstone, K. Panoutsopoulou, L. Paternoster, J. R. B. Perry, L. Quaye, J. Brent Richards, S. Ring, G. R. S. Ritchie, S. Schiffels, H. A. Shihab, S.-Y. Shin, K. S. Small, M. Soler Artigas, N. Soranzo, L. Southam, T. D. Spector, B. St Pourcain, G. Surdulescu, I. Tachmazidou, N. J. Timpson, M. D. Tobin, A. M. Valdes, P. M. Visscher, L. V. Wain, K. Walter, K. Ward, S. G. Wilson, K. Wong, J. Yang, E. Zeggini, F. Zhang, H.-F. Zheng, R. Anney, M. Ayub, J. C. Barrett, D. Blackwood, P. F. Bolton, G. Breen, D. A. Collier, N. Craddock, L. Crooks, S.

Curran, D. Curtis, R. Durbin, L. Gallagher, D. Geschwind, H. Gurling, P. Holmans, I. Lee, J. Lönnqvist, S. McCarthy, P. McGuffin, A. M. McIntosh, A. G. McKechanie, A. McQuillin, J. Morris, D. Muddyman, M. C. O'Donovan, M. J. Owen, A. Palotie, J. R. Parr, T. Paunio, O. Pietilainen, K. Rehnström, S. I. Sharp, D. Skuse, D. St Clair, J. Suvisaari, J. T. R. Walters, H. J. Williams, I. Barroso, E. Bochukova, R. Bounds, A. Dominiczak, R. Durbin, I. S. Farooqi, A. E. Hendricks, J. Keogh, G. Marenne, S. McCarthy, A. Morris, D. Muddyman, S. O'Rahilly, D. J. Porteous, B. H. Smith, I. Tachmazidou, E. Wheeler, E. Zeggini, S. Al Turki, C. A. Anderson, D. Antony, I. Barroso, P. Beales, J. Bentham, S. Bhattacharya, M. Calissano, K. Carss, K. Chatterjee, S. Cirak, C. Cosgrove, R. Durbin, D. R. Fitzpatrick, J. Floyd, A. Reghan Foley, C. S. Franklin, M. Futema, D. Grozeva, S. E. Humphries, M. E. Hurles, S. McCarthy, H. M. Mitchison, D. Muddyman, F. Muntoni, S. O'Rahilly, A. Onoufriadis, V. Parker, F. Payne, V. Plagnol, F. Lucy Raymond, N. Roberts, D. B. Savage, P. Scambler, M. Schmidts, N. Schoenmakers, R. K. Semple, E. Serra, O. Spasic-Boskovic, E. Stevens, M. van Kogelenberg, P. Vijayarangakannan, K. Walter, K. A. Williamson, C. Wilson, T. Whyte, A. Ciampi, C. M. T. Greenwood, A. E. Hendricks, R. Li, S. Metrustry, K. Oualkacha, I. Tachmazidou, C. J. Xu, E. Zeggini, M. Bobrow, P. F. Bolton, R. Durbin, D. R. Fitzpatrick, H. Griffin, M. E. Hurles, J. Kaye, K. Kennedy, A. Kent, D. Muddyman, F. Muntoni, F. Lucy Raymond, R. K. Semple, C. Smee, T. D. Spector, N. J. Timpson, R. Charlton, R. Ekong, M. Futema, S. E. Humphries, F. Khawaja, L. R. Lopes, N. Migone, S. J. Payne, V. Plagnol, R. C. Pollitt, S. Povey, C. K. Ridout, R. L. Robinson, R. H. Scott, A. Shaw, P. Syrris, R. Taylor, A. M. Vandersteen, J. C. Barrett, I. Barroso, G. Davey Smith, R. Durbin, I. S. Farooqi, D. R. Fitzpatrick, M. E. Hurles, J. Kaye, K. Kennedy, C. Langford, S. McCarthy, D. Muddyman, M. J. Owen, A. Palotie, J. Brent Richards, N. Soranzo, T. D. Spector, J. Stalker, N. J. Timpson, E. Zeggini, A. Amuzu, J. Pablo Casas, J. C. Chambers, M. Cocca, G. Dedoussis, G. Gambaro, P. Gasparini, T. R. Gaunt, J. Huang, V. Iotchkova, A. Isaacs, J. Johnson, M. E. Kleber, J. S. Kooner, C. Langenberg, J. Luan, G. Malerba, W. März, A. Matchan, J. L. Min, R. Morris, B. G. Nordestgaard, M. Benn, S. Ring, R. A. Scott, N. Soranzo, L. Southam, N. J. Timpson, D. Toniolo, M. Traglia, A. Tybjaerg-Hansen, C. M. van Duijn, E. M. van Leeuwen, A. Varbo, P. Whincup, G. Zaza, E. Zeggini, W. Zhang; UK10K Consortium, The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82–90 (2015). doi:10.1038/nature14962 Medline

16. P. I. W. de Bakker, S. Raychaudhuri, Interrogating the major histocompatibility complex with high-throughput genomics. *Hum. Mol. Genet.* **21** (R1), R29–R36 (2012). doi:10.1093/hmg/dds384 Medline

17. E. M. Leffler, Z. Gao, S. Pfeifer, L. Ségurel, A. Auton, O. Venn, R. Bowden, R. Bontrop, J. D. Wall, G. Sella, P. Donnelly, G. McVean, M. Przeworski, Multiple instances of ancient balancing selection shared between humans and chimpanzees. *Science* **339**, 1578–1582 (2013). doi:10.1126/science.1234070 Medline

18. D. Welter, J. MacArthur, J. Morales, T. Burdett, P. Hall, H. Junkins, A. Klemm, P. Flicek, T. Manolio, L. Hindorff, H. Parkinson, The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **42**, D1001–D1006 (2014). doi:10.1093/nar/gkt1229 Medline

19. J. J. Berg, G. Coop, A population genetic signal of polygenic adaptation. *PLOS Genet.* **10**, e1004412 (2014). doi:10.1371/journal.pgen.1004412 Medline

20. M. R. Robinson, G. Hemani, C. Medina-Gomez, M. Mezzavilla, T. Esko, K. Shakhbazov, J. E. Powell, A. Vinkhuyzen, S. I. Berndt, S. Gustafsson, A. E. Justice, B. Kahali, A. E. Locke, T. H. Pers, S. Vedantam, A. R. Wood, W. van Rheenen, O. A. Andreassen, P. Gasparini, A. Metspalu, L. H. Berg, J. H. Veldink, F. Rivadeneira, T. M. Werge, G. R. Abecasis, D. I. Boomsma, D. I. Chasman, E. J. C. de Geus, T. M. Frayling, J. N. Hirschhorn, J. J. Hottenga, E. Ingelsson, R. J. F. Loos, P. K. E. Magnusson, N. G. Martin, G. W. Montgomery, K. E. North, N. L. Pedersen, T. D. Spector, E. K. Speliotes, M. E. Goddard, J. Yang, P. M. Visscher, Population genetic differentiation of height and body mass index across Europe. *Nat. Genet.* **47**, 1357–1362 (2015). doi:10.1038/ng.3401 Medline

21. A. R. Wood, T. Esko, J. Yang, S. Vedantam, T. H. Pers, S. Gustafsson, A. Y. Chu, K. Estrada, J. Luan, Z. Kutalik, N. Amin, M. L. Buchkovich, D. C. Croteau-Chonka, F. R. Day, Y. Duan, T. Fall, R. Fehrmann, T. Ferreira, A. U. Jackson, J. Karjalainen, K. S. Lo, A. E. Locke, R. Mägi, E. Mihailov, E. Porcu, J. C. Randall, A. Scherag, A. A. E. Vinkhuyzen, H.-J. Westra, T. W. Winkler, T. Workalemahu, J. H. Zhao, D. Absher, E. Albrecht, D. Anderson, J. Baron, M. Beekman, A. Demirkan, G. B. Ehret, B. Feenstra, M. F. Feitosa, K. Fischer, R. M. Fraser, A. Goel, J. Gong, A. E. Justice, S. Kanoni, M. E. Kleber, K. Kristiansson, U. Lim, V. Lotay, J. C. Lui, M. Mangino, I. Mateo Leach, C. Medina-Gomez, M. A. Nalls, D. R. Nyholt, C. D. Palmer, D. Pasko, S. Pechlivanis, I. Prokopenko, J. S. Ried, S. Ripke, D. Shungin, A. Stancáková, R. J. Strawbridge, Y. J. Sung, T. Tanaka, A. Teumer, S. Trompet, S. W. van der Laan, J. van Setten, J. V. Van Vliet-Ostaptchouk, Z. Wang, L. Yengo, W. Zhang, U. Afzal, J. Arnlöv, G. M. Arscott, S. Bandinelli, A. Barrett, C. Bellis, A. J. Bennett, C. Berne, M. Blüher, J. L. Bolton, Y. Böttcher, H. A. Boyd, M. Bruinenberg, B. M. Buckley, S. Buyske, I. H. Caspersen, P. S. Chines, R. Clarke, S. Claudi-Boehm, M. Cooper, E. W. Daw, P. A. De Jong, J. Deelen, G. Delgado, J. C. Denny, R. Dhonukshe-Rutten, M. Dimitriou, A. S. F. Doney, M. Dörr, N. Eklund, E. Eury, L. Folkersen, M. E. Garcia, F. Geller, V. Giedraitis, A. S. Go, H. Grallert, T. B. Grammer, J. Gräßler, H. Grönberg, L. C. P. G. M. de Groot, C. J. Groves, J. Haessler, P. Hall, T. Haller, G. Hallmans, A. Hannemann, C. A. Hartman, M. Hassinen, C. Hayward, N. L. Heard-Costa, Q. Helmer, G. Hemani, A. K. Henders, H. L. Hillege, M. A. Hlatky, W. Hoffmann, P. Hoffmann, O. Holmen, J. J. Houwing-Duistermaat, T. Illig, A. Isaacs, A. L. James, J. Jeff, B. Johansen, Å. Johansson, J. Jolley, T. Juliusdottir, J. Junttila, A. N. Kho, L. Kinnunen, N. Klopp, T. Kocher, W. Kratzer, P. Lichtner, L. Lind, J. Lindström, S. Lobbens, M. Lorentzon, Y. Lu, V. Lyssenko, P. K. E. Magnusson, A. Mahajan, M. Maillard, W. L. McArdle, C. A. McKenzie, S. McLachlan, P. J. McLaren, C. Menni, S. Merger, L. Milani, A. Moayyeri, K. L. Monda, M. A. Morken, G. Müller, M. Müller-Nurasyid, A. W. Musk, N. Narisu, M. Nauck, I. M. Nolte, M. M. Nöthen, L. Oozageer, S. Pilz, N. W. Rayner, F. Renstrom, N. R. Robertson, L. M. Rose, R. Roussel, S. Sanna, H. Scharnagl, S. Scholtens, F. R. Schumacher, H. Schunkert, R. A. Scott, J. Sehmi, T. Seufferlein, J. Shi, K. Silventoinen, J. H. Smit, A. V. Smith, J. Smolonska, A. V. Stanton, K. Stirrups, D. J. Stott, H. M. Stringham, J. Sundström, M. A. Swertz, A.-C. Syvänen, B. O. Tayo, G. Thorleifsson, J. P. Tyrer, S. van Dijk, N. M. van Schoor, N. van der Velde, D. van Heemst, F. V. A. van Oort, S. H. Vermeulen, N. Verweij, J. M. Vonk, L. L. Waite, M. Waldenberger, R. Wennauer, L. R. Wilkens, C.

Willenborg, T. Wilsgaard, M. K. Wojczynski, A. Wong, A. F. Wright, Q. Zhang, D. Arveiler, S. J. L. Bakker, J. Beilby, R. N. Bergman, S. Bergmann, R. Biffar, J. Blangero, D. I. Boomsma, S. R. Bornstein, P. Bovet, P. Brambilla, M. J. Brown, H. Campbell, M. J. Caulfield, A. Chakravarti, R. Collins, F. S. Collins, D. C. Crawford, L. A. Cupples, J. Danesh, U. de Faire, H. M. den Ruijter, R. Erbel, J. Erdmann, J. G. Eriksson, M. Farrall, E. Ferrannini, J. Ferrières, I. Ford, N. G. Forouhi, T. Forrester, R. T. Gansevoort, P. V. Gejman, C. Gieger, A. Golay, O. Gottesman, V. Gudnason, U. Gyllensten, D. W. Haas, A. S. Hall, T. B. Harris, A. T. Hattersley, A. C. Heath, C. Hengstenberg, A. A. Hicks, L. A. Hindorff, A. D. Hingorani, A. Hofman, G. K. Hovingh, S. E. Humphries, S. C. Hunt, E. Hypponen, K. B. Jacobs, M.-R. Jarvelin, P. Jousilahti, A. M. Jula, J. Kaprio, J. J. P. Kastelein, M. Kayser, F. Kee, S. M. Keinanen-Kiukaanniemi, L. A. Kiemeney, J. S. Kooner, C. Kooperberg, S. Koskinen, P. Kovacs, A. T. Kraja, M. Kumari, J. Kuusisto, T. A. Lakka, C. Langenberg, L. Le Marchand, T. Lehtimäki, S. Lupoli, P. A. F. Madden, S. Männistö, P. Manunta, A. Marette, T. C. Matise, B. McKnight, T. Meitinger, F. L. Moll, G. W. Montgomery, A. D. Morris, A. P. Morris, J. C. Murray, M. Nelis, C. Ohlsson, A. J. Oldehinkel, K. K. Ong, W. H. Ouwehand, G. Pasterkamp, A. Peters, P. P. Pramstaller, J. F. Price, L. Qi, O. T. Raitakari, T. Rankinen, D. C. Rao, T. K. Rice, M. Ritchie, I. Rudan, V. Salomaa, N. J. Samani, J. Saramies, M. A. Sarzynski, P. E. H. Schwarz, S. Sebert, P. Sever, A. R. Shuldiner, J. Sinisalo, V. Steinthorsdottir, R. P. Stolk, J.-C. Tardif, A. Tönjes, A. Tremblay, E. Tremoli, J. Virtamo, M.-C. Vohl, P. Amouyel, F. W. Asselbergs, T. L. Assimes, M. Bochud, B. O. Boehm, E. Boerwinkle, E. P. Bottinger, C. Bouchard, S. Cauchi, J. C. Chambers, S. J. Chanock, R. S. Cooper, P. I. W. de Bakker, G. Dedoussis, L. Ferrucci, P. W. Franks, P. Froguel, L. C. Groop, C. A. Haiman, A. Hamsten, M. G. Hayes, J. Hui, D. J. Hunter, K. Hveem, J. W. Jukema, R. C. Kaplan, M. Kivimaki, D. Kuh, M. Laakso, Y. Liu, N. G. Martin, W. März, M. Melbye, S. Moebus, P. B. Munroe, I. Njølstad, B. A. Oostra, C. N. A. Palmer, N. L. Pedersen, M. Perola, L. Pérusse, U. Peters, J. E. Powell, C. Power, T. Quertermous, R. Rauramaa, E. Reinmaa, P. M. Ridker, F. Rivadeneira, J. I. Rotter, T. E. Saaristo, D. Saleheen, D. Schlessinger, P. E. Slagboom, H. Snieder, T. D. Spector, K. Strauch, M. Stumvoll, J. Tuomilehto, M. Uusitupa, P. van der Harst, H. Völzke, M. Walker, N. J. Wareham, H. Watkins, H.-E. Wichmann, J. F. Wilson, P. Zanen, P. Deloukas, I. M. Heid, C. M. Lindgren, K. L. Mohlke, E. K. Speliotes, U. Thorsteinsdottir, I. Barroso, C. S. Fox, K. E. North, D. P. Strachan, J. S. Beckmann, S. I. Berndt, M. Boehnke, I. B. Borecki, M. I. McCarthy, A. Metspalu, K. Stefansson, A. G. Uitterlinden, C. M. van Duijn, L. Franke, C. J. Willer, A. L. Price, G. Lettre, R. J. F. Loos, M. N. Weedon, E. Ingelsson, J. R. O'Connell, G. R. Abecasis, D. I. Chasman, M. E. Goddard, P. M. Visscher, J. N. Hirschhorn, T. M. Frayling; Electronic Medical Records and Genomics (eMEMERGEGE) Consortium; MIGen Consortium; PAGEGE Consortium; LifeLines Cohort Study, Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014). doi:10.1038/ng.3097 Medline

22. J. Yang, B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders, D. R. Nyholt, P. A. Madden, A. C. Heath, N. G. Martin, G. W. Montgomery, M. E. Goddard, P. M. Visscher, Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010). doi:10.1038/ng.608 Medline

23. M. Stephens, False Discovery Rates: A New Deal. *bioRxiv* (2016). doi:10.1101/038216

24. B. K. Bulik-Sullivan, P.-R. Loh, H. K. Finucane, S. Ripke, J. Yang, N. Patterson, M. J. Daly, A. L. Price, B. M. Neale; Schizophrenia Working Group of the Psychiatric Genomics Consortium, LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015). doi:10.1038/ng.3211 Medline

25. B. Bulik-Sullivan, H. K. Finucane, V. Anttila, A. Gusev, F. R. Day, P.-R. Loh, L. Duncan, J. R. B. Perry, N. Patterson, E. B. Robinson, M. J. Daly, A. L. Price, B. M. Neale; ReproGen Consortium; Psychiatric Genomics Consortium; Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3, An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015). doi:10.1038/ng.3406 Medline

26. J. Wakeley, *Coalescent Theory: An Introduction* (Freeman, 2008); https://books.google.com/books?id=x30RAgAACAAJ).

27. P. C. Sabeti, D. E. Reich, J. M. Higgins, H. Z. P. Levine, D. J. Richter, S. F. Schaffner, S. B. Gabriel, J. V. Platko, N. J. Patterson, G. J. McDonald, H. C. Ackerman, S. J. Campbell, D. Altshuler, R. Cooper, D. Kwiatkowski, R. Ward, E. S. Lander, Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**, 832–837 (2002). doi:10.1038/nature01140 Medline

28. I. Mathieson, G. McVean, Demography and the age of rare variants. *PLOS Genet.* **10**, e1004528 (2014). doi:10.1371/journal.pgen.1004528 Medline

29. K. Harris, Evidence for recent, population-specific evolution of the human mutation rate. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 3439–3444 (2015). doi:10.1073/pnas.1418652112 Medline

30. R. R. Hudson, Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* **18**, 337–338 (2002). doi:10.1093/bioinformatics/18.2.337 Medline

31. B. Peng, M. Kimmel, simuPOP: A forward-time population genetics simulation environment. *Bioinformatics* **21**, 3686–3687 (2005). doi:10.1093/bioinformatics/bti584 Medline

32. K. M. Teshima, H. Innan, mbs: Modifying Hudson's ms software to generate samples of DNA sequences with a biallelic site under selection. *BMC Bioinformatics* **10**, 166 (2009). doi:10.1186/1471-2105-10-166 Medline

33. J. Kelleher, A. M. Etheridge, G. McVean, Efficient Coalescent Simulation and Genealogical Analysis for Large Sample Sizes. *PLOS Comput. Biol.* **12**, e1004842 (2016). doi:10.1371/journal.pcbi.1004842 Medline

34. G. Ewing, J. Hermisson, MSMS: A coalescent simulation program including recombination, demographic structure and selection at a single locus. *Bioinformatics* **26**, 2064–2065 (2010). doi:10.1093/bioinformatics/btq322 Medline

35. G. R. Abecasis, A. Auton, L. D. Brooks, M. A. DePristo, R. M. Durbin, R. E. Handsaker, H. M. Kang, G. T. Marth, G. A. McVean; 1000 Genomes Project Consortium, An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012). Medline

36. Z. A. Szpiech, R. D. Hernandez, selscan: An efficient multithreaded program to perform EHH-based scans for positive selection. *Mol. Biol. Evol.* **31**, 2824–2827 (2014). doi:10.1093/molbev/msu211 Medline

37. A. Auton, L. D. Brooks, R. M. Durbin, E. P. Garrison, H. M. Kang, J. O. Korbel, J. L. Marchini, S. McCarthy, G. A. McVean, G. R. Abecasis; 1000 Genomes Project Consortium, A global reference for human genetic variation. *Nature* **526**, 68–74 (2015). doi:10.1038/nature15393 Medline

38. M. R. Nelson, D. Wegmann, M. G. Ehm, D. Kessner, P. St Jean, C. Verzilli, J. Shen, Z. Tang, S. A. Bacanu, D. Fraser, L. Warren, J. Aponte, M. Zawistowski, X. Liu, H. Zhang, Y. Zhang, J. Li, Y. Li, L. Li, P. Woollard, S. Topp, M. D. Hall, K. Nangle, J. Wang, G. Abecasis, L. R. Cardon, S. Zöllner, J. C. Whittaker, S. L. Chissoe, J. Novembre, V. Mooser, An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science* **337**, 100–104 (2012). doi:10.1126/science.1217876 Medline

39. E. Gazave, L. Ma, D. Chang, A. Coventry, F. Gao, D. Muzny, E. Boerwinkle, R. A. Gibbs, C. F. Sing, A. G. Clark, A. Keinan, Neutral genomic regions refine models of recent rapid human population growth. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 757–762 (2014). doi:10.1073/pnas.1310398110 Medline

40. P. Moorjani, Z. Gao, M. Przeworski, Human germline mutation and the erratic evolutionary clock. *bioRxiv* (2016); http://biorxiv.org/content/early/2016/08/05/058024.

41. V. M. Narasimhan *et al*., A direct multi-generational estimate of the human mutation rate from autozygous segments seen in thousands of parentally related individuals. *bioRxiv* (2016); http://biorxiv.org/content/early/2016/06/17/059436.

42. Y. Itan, A. Powell, M. A. Beaumont, J. Burger, M. G. Thomas, The origins of lactase persistence in Europe. *PLOS Comput. Biol.* **5**, e1000491 (2009). doi:10.1371/journal.pcbi.1000491 Medline

43. P. Gerbault, A. Liebert, Y. Itan, A. Powell, M. Currat, J. Burger, D. M. Swallow, M. G. Thomas, Evolution of lactase persistence: An example of human niche construction. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **366**, 863–877 (2011). doi:10.1098/rstb.2010.0268 Medline

44. P. Gerbault, C. Moret, M. Currat, A. Sanchez-Mazas, Impact of selection and demography on the diffusion of lactase persistence. *PLOS ONE* **4**, e6369 (2009). doi:10.1371/journal.pone.0006369 Medline

45. O. Ó. Sverrisdóttir, A. Timpson, J. Toombs, C. Lecoeur, P. Froguel, J. M. Carretero, J. L. Arsuaga Ferreras, A. Götherström, M. G. Thomas, Direct estimates of natural selection in Iberia indicate calcium absorption was not the only driver of lactase persistence in Europe. *Mol. Biol. Evol.* **31**, 975–983 (2014). doi:10.1093/molbev/msu049 Medline

46. P. Sulem, D. F. Gudbjartsson, S. N. Stacey, A. Helgason, T. Rafnar, K. P. Magnusson, A. Manolescu, A. Karason, A. Palsson, G. Thorleifsson, M. Jakobsdottir, S. Steinberg, S. Pálsson, F. Jonasson, B. Sigurgeirsson, K. Thorisdottir, R. Ragnarsson, K. R. Benediktsdottir, K. K. Aben, L. A. Kiemeney, J. H. Olafsson, J. Gulcher, A. Kong, U. Thorsteinsdottir, K. Stefansson, Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat. Genet.* **39**, 1443–1452 (2007). doi:10.1038/ng.2007.13 Medline

47. T. C. S. and Analysis Consortium; Chimpanzee Sequencing and Analysis Consortium, Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69–87 (2005). doi:10.1038/nature04072 Medline

48. O. Lao, J. M. de Gruijter, K. van Duijn, A. Navarro, M. Kayser, Signatures of positive selection in genes associated with human skin pigmentation as revealed from analyses of single nucleotide polymorphisms. *Ann. Hum. Genet.* **71**, 354–369 (2007). doi:10.1111/j.1469-1809.2006.00341.x Medline

49. S. H. Williamson, M. J. Hubisz, A. G. Clark, B. A. Payseur, C. D. Bustamante, R. Nielsen, Localizing recent adaptive evolution in the human genome. *PLOS Genet.* **3**, e90 (2007). doi:10.1371/journal.pgen.0030090 Medline

50. G. Coop, J. K. Pickrell, J. Novembre, S. Kudaravalli, J. Li, D. Absher, R. M. Myers, L. L. Cavalli-Sforza, M. W. Feldman, J. K. Pritchard, The role of geography in human adaptation. *PLOS Genet.* **5**, e1000500 (2009). doi:10.1371/journal.pgen.1000500 Medline

51. C. T. Miller, S. Beleza, A. A. Pollen, D. Schluter, R. A. Kittles, M. D. Shriver, D. M. Kingsley, cis-Regulatory changes in Kit ligand expression and parallel evolution of pigmentation in sticklebacks and humans. *Cell* **131**, 1179–1189 (2007). doi:10.1016/j.cell.2007.10.055 Medline

52. S. Beleza, A. M. Santos, B. McEvoy, I. Alves, C. Martinho, E. Cameron, M. D. Shriver, E. J. Parra, J. Rocha, The timing of pigmentation lightening in Europeans. *Mol. Biol. Evol.* **30**, 24–35 (2013). doi:10.1093/molbev/mss207 Medline

53. H. R. Kunsch, The Jackknife and the Bootstrap for General Stationary Observations. *Ann. Stat.* **17**, 1217–1241 (1989). doi:10.1214/aos/1176347265

54. F. M. T. A. Busing, E. Meijer, R. van der Leeden, Delete- m Jackknife for Unequal m. *Stat. Comput.* **9**, 3–8 (1999). doi:10.1023/A:1008800423698

55. G. McVicker, D. Gordon, C. Davis, P. Green, Widespread genomic signatures of natural selection in hominid evolution. *PLOS Genet.* **5**, e1000471 (2009). doi:10.1371/journal.pgen.1000471 Medline

56. D. L. Cousminer, E. Stergiakouli, D. J. Berry, W. Ang, M. M. Groen-Blokhuis, A. Körner, N. Siitonen, I. Ntalla, M. Marinelli, J. R. B. Perry, J. Kettunen, R. Jansen, I. Surakka, N. J. Timpson, S. Ring, G. Mcmahon, C. Power, C. Wang, M. Kähönen, J. Viikari, T. Lehtimäki, C. M. Middeldorp, H. E. Hulshoff Pol, M. Neef, S. Weise, K. Pahkala, H. Niinikoski, E. Zeggini, K. Panoutsopoulou, M. Bustamante, B. W. J. H. Penninx, J. Murabito, M. Torrent, G. V. Dedoussis, W. Kiess, D. I. Boomsma, C. E. Pennell, O. T. Raitakari, E. Hyppönen, G. Davey Smith, S. Ripatti, M. I. McCarthy, E. Widén; ReproGen Consortium; Early Growth Genetics Consortium, Genome-wide association study of sexual maturation in males and females highlights a role for body mass and menarche loci in male puberty. *Hum. Mol. Genet.* **23**, 4452–4464 (2014). doi:10.1093/hmg/ddu150 Medline

57. J. R. B. Perry, L. Stolk, N. Franceschini, K. L. Lunetta, G. Zhai, P. F. McArdle, A. V. Smith, T. Aspelund, S. Bandinelli, E. Boerwinkle, L. Cherkas, G. Eiriksdottir, K. Estrada, L. Ferrucci, A. R. Folsom, M. Garcia, V. Gudnason, A. Hofman, D. Karasik, D. P. Kiel, L. J. Launer, J. van Meurs, M. A. Nalls, F. Rivadeneira, A. R. Shuldiner, A. Singleton, N.

Soranzo, T. Tanaka, J. A. Visser, M. N. Weedon, S. G. Wilson, V. Zhuang, E. A. Streeten, T. B. Harris, A. Murray, T. D. Spector, E. W. Demerath, A. G. Uitterlinden, J. M. Murabito, Meta-analysis of genome-wide association data identifies two loci influencing age at menarche. *Nat. Genet.* **41**, 648–650 (2009). doi:10.1038/ng.386 Medline

58. A.-S. Parent, G. Teilmann, A. Juul, N. E. Skakkebaek, J. Toppari, J.-P. Bourguignon, The timing of normal puberty and the age limits of sexual precocity: Variations around the world, secular trends, and changes after migration. *Endocr. Rev.* **24**, 668–693 (2003). doi:10.1210/er.2002-0019 Medline

59. O. Karapanou, A. Papadimitriou, Determinants of menarche. *Reprod. Biol. Endocrinol.* **8**, 115 (2010). doi:10.1186/1477-7827-8-115 Medline

60. V. Natale, A. Rajagopalan, Worldwide variation in human growth and the World Health Organization growth standards: A systematic review. *BMJ Open* **4**, e003735–e003735 (2014). doi:10.1136/bmjopen-2013-003735 Medline

61. A. Bhaskar, Y. X. R. Wang, Y. S. Song, Efficient inference of population size histories and locus-specific mutation rates from large-sample genomic variation data. *Genome Res.* **25**, 268–279 (2015). doi:10.1101/gr.178756.114 Medline

62. M. C. Maher, L. H. Uricchio, D. G. Torgerson, R. D. Hernandez, Population genetics of rare variants and complex diseases. *Hum. Hered.* **74**, 118–128 (2012). doi:10.1159/000346826 Medline

63. B. Charlesworth, M. T. Morgan, D. Charlesworth, The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289–1303 (1993). Medline