Widespread signatures of negative selection in the genetic architecture of human complex traits

Jian Zeng[1], Ronald de Vlaming[2,3], Yang Wu[1], Matthew R Robinson[1,4], Luke Lloyd-Jones[1], Loic Yengo[1], Chloe Yap[1], Angli Xue[1], Julia Sidorenko[1], Allan F McRae[1], Joseph E Powell[1], Grant W Montgomery[1], Andres Metspalu[5], Tonu Esko[5], Greg Gibson[6], Naomi R Wray[1,7], Peter M Visscher[1,7], Jian Yang[1,7]

[1] Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland 4072, Australia
[2] Department of Complex Trait Genetics, VU University, Center for Neurogenomics and Cognitive Research, Amsterdam, 1081 HV, The Netherlands
[3] Erasmus University Rotterdam Institute for Behavior and Biology, Rotterdam, 3062 PA, The Netherlands
[4] Department of Computational Biology, University of Lausanne, 1011 Lausanne, Switzerland
[5] Estonian Genome Center, University of Tartu, Tartu, Estonia
[6] School of Biology and Center for Integrative Genomics, Georgia Institute of Technology, Atlanta, GA 30332, USA
[7] Queensland Brain Institute, The University of Queensland, Brisbane, Queensland 4072, Australia
Correspondence: Jian Yang <jian.yang@uq.eud.au>

Abstract

Estimation of the joint distribution of effect size and minor allele frequency (MAF) for genetic variants is important for understanding the genetic basis of complex trait variation and can be used to detect signature of natural selection. We develop a Bayesian mixed linear model that simultaneously estimates SNP-based heritability, polygenicity (i.e. the proportion of SNPs with nonzero effects) and the relationship between effect size and MAF for complex traits in conventionally unrelated individuals using genome-wide SNP data. We apply the method to 28 complex traits in the UK Biobank data (N = 126,752), and show that on average across 28 traits, 6% of SNPs have nonzero effects, which in total explain 22% of phenotypic variance. We detect significant ($p < 0.05/28 = 1.8\times10^{-3}$) signatures of natural selection for 23 out of 28 traits including reproductive, cardiovascular, and anthropometric traits, as well as educational attainment. We further apply the method to 27,869 gene expression traits (N = 1,748), and identify 30 genes that show significant ($p < 2.3\times10^{-6}$) evidence of natural selection. All the significant estimates of the relationship between effect size and MAF in either complex traits or gene expression traits are consistent with a model of negative selection, as confirmed by

forward simulation. We conclude that natural selection acts pervasively on human complex traits shaping genetic variation in the form of negative selection.

Introduction

Dissecting the genetic architecture of complex traits is important for understanding the genetic basis of phenotypic variation and evolution. For a complex trait that influences fitness, natural selection plays an important role in shaping its genetic architecture[1], which in turn provides information to infer the action of natural selection. Given most traits are polygenic, natural selection is likely to act simultaneously on many trait-associated variants that have pleiotropic effects on fitness (known as polygenic selection[2-4]). Unlike a selective sweep model[5] where there are often a limited number of mutations under relatively strong selection, it is difficult to detect the signals of polygenic selection due to the selection pressure being spread over many variants of small effect. However, evidence for natural selection can be inferred from the relationship between effect size and minor allele frequency (MAF) at the genome-wide variants. For example, mutations that are deleterious to fitness are selected against and thus kept at low frequencies by negative selection, resulting in a correlation between effect sizes and MAF[6,7]. The estimation of the joint distribution of effect size and MAF can be used to detect signature of natural selection and thereby to infer the relationship between a complex trait and fitness.

Genome-wide association studies (GWAS) have detected thousands of SNPs associated with complex traits, which have helped to characterize the genetic architecture of these traits[8-13]. However, the genome-wide significant SNPs discovered in GWAS jointly tend to explain only a fraction of the heritability as many SNPs with small effects yet to be detected[14]. Furthermore, a proportion is missed due to the incomplete linkage disequilibrium (LD) between causal variants and SNP markers[14]. To address the "missing heritability" problem[14,15] in GWAS, mixed linear model (MLM) approaches have been used to estimate the genetic variance explained by all SNPs used in a GWAS. GREML is a prevailing class of MLM-based approaches where all SNP effects are fitted together as random effects[16]. GREML analyses using common SNPs (MAF > 0.01) have uncovered a large proportion of the "missing heritability" for height[17], BMI[17] and psychiatric disorders[18]. The GREML method assumes that all SNPs have an effect on the trait[16] and thus does not allow us to estimate the degree of polygenicity (i.e. the proportion of SNPs with nonzero effects). Bayesian multiple regression is another class of MLM-based methods that enable us to make posterior inference about polygenicity by assuming SNP effects are drawn from a mixture distribution of zero and nonzero components[19-21]. Bayesian MLM methods have been widely used in livestock and plant breeding[22] and have attracted increasing attention in humans for characterizing the genetic architecture of complex traits and diseases[20,23,24].

However, neither GREML nor Bayesian MLM approaches explicitly model the relationship between effect size and MAF, an important characteristic of the genetic architecture for complex traits. This relationship can be used to detect signatures of natural selection[7,25] and inform the design of future genetic mapping studies.

In this study, we developed an MLM-based Bayesian method that can simultaneously estimate SNP-based heritability, polygenicity and the joint distribution of effect size and MAF in conventionally unrelated individuals using GWAS data. We applied the method to 28 complex traits in the UK Biobank (UKB) data[26], and 27,869 gene expression traits in the Consortium for the Architecture of Gene Expression (CAGE) dataset[27], and identified a number of complex traits and gene expression traits for which there is significant evidence of natural selection on the associated SNPs.

Results

Method overview

Under the Bayesian MLM framework, we propose to model the relationship between effect size and MAF using the following mixture distribution as prior for each SNP effect

$$\beta_j \sim N\left(0, \left[2p_j(1-p_j)\right]^S \sigma_\beta^2\right)\pi + \phi(1-\pi)$$

where β_j is the allelic substitution effect of a SNP j, p_j is the MAF of the SNP, σ_β^2 is a constant factor (i.e. variance of SNP effects under a neutral model), ϕ is a point mass at zero, and π is the proportion of SNPs with nonzero effects (polygenicity). The variance of the effect size of SNP j is $\sigma_j^2 = \left[2p_j(1-p_j)\right]^S \sigma_\beta^2$, which is a function of MAF of the SNP. Thus, the parameter S measures the relationship between effect size and MAF. If $S = 0$, the effect size is independent of MAF (neutral model). If there is selection, the effect size can be positively ($S > 0$) or negatively ($S < 0$) related to MAF. All these parameters are treated as unknown with appropriate priors (**Online Methods**). Our model (referred to as BayesS) allows simultaneous estimation of multiple characteristics of the genetic architecture: SNP-based heritability (h_{SNP}^2), polygenicity (π) and the relationship between SNP effect and MAF (S). We use a gradient-based sampling algorithm, Hamiltonian Monte Carlo[28], to sample S from the posterior distribution, and use Gibbs sampling for other parameters in the model by assuming conjugate priors. Furthermore, we use a parallel computing strategy following Fernando *et al.*[29] to allow the analysis to be scalable to very large samples sizes ($N > 100{,}000$). Details of sampling scheme and parallel computing strategies are given in the **Supplementary Note**.

In the hypothesis test against $S = 0$, we used two approaches to control false positives. The first approach is to control the family-wise type I error rate (FWER) using the theory that the posterior mode standardized by the posterior standard error (s.e.) asymptotically follows a standard normal distribution under the null[30]. The asymptotic normality of the posterior distribution was justified by simulation with the UKB cohort (**Supplementary Fig. 1**). The second approach is to control the proportion of false positives[31] (PFP) among rejections (also known as the marginal false discovery rate or mFDR[32]) based on the posterior probability given the data, *e.g.* $\Pr(S < 0|\mathcal{D})$ (**Supplementary Note**). We show by simulation that if the true distribution of S is used as the prior, then rejecting $S = 0$ with $\Pr(S < 0|\mathcal{D}) \geq \gamma$ guarantees PFP or mFDR to be less than $1 - \gamma$ (**Supplementary Fig. 2**). The former approach is more stringent but the advantage of the latter approach is that the power is not inversely related to the number of traits (tests)[31].

Assessing the robustness of parameter estimation through simulations based on real genotype data

We used simulations based on real GWAS genotype data from the Atherosclerosis Risk in Communities (ARIC) study[33] to assess our method in estimating the parameters $\boldsymbol{\theta} = [S, h^2_{SNP}, \pi]$. The ARIC data consist of 12,942 unrelated individuals and 564,959 Affymetrix SNPs with MAF > 1% after quality control (**Online Methods**). In our simulation, 1,000 SNPs were chosen at random to be causal variants, with their effects related to MAF through an S value ranging from -1 to 1 in different scenarios (**Online Methods**). Since the number of causal variants was known, polygenicity was assessed by the number of SNPs with nonzero effects (m_{NZ}). Based on the Markov chain Monte Carlo (MCMC) samples, the point estimate ($\hat{\theta}$), standard error (s.e.) or credible interval for each parameter was given by the mode, standard deviation (s.d.) or highest probability density (HPD) of its posterior distribution, respectively.

Results (**Fig. 1**) show that when both causal variants and SNP markers were fitted in the analysis, $\widehat{\boldsymbol{\theta}}$ from BayesS was unbiased with respect to the true parameters. When the causal variants were not included in the analysis, both $\hat{h}^2_{SNP}$ and the absolute value of $\hat{S}$ were slightly underestimated, due to imperfect tagging, a similar issue as discussed in Yang *et al.*[16]. For polygenicity, however, m_{NZ} estimate tended to be larger than the number of causal variants, probably because some causal variants could be better tagged by multiple SNPs. Thus, in practice, $\hat{\pi}$ should be interpreted as the proportion of non-null SNPs, which is likely to be larger than the proportion of causal variants. Results also show that the s.e. for $\hat{S}$, $\hat{h}^2_{SNP}$ and $\hat{\pi}$ is

consistent with the s.d. of the estimates from 100 simulation replicates (**Supplementary Table 1**).

Analysis of 28 complex traits in the UK Biobank data

We applied the BayesS method to 36 complex traits on 126,545 unrelated individuals of European ancestry in the UKB[26] with 483,634 Affymetrix SNPs (MAF > 1%) after quality controls (**Online Methods**). Out of the 36 traits 21 have $N > 100{,}000$. Two commonly used long-chain diagnostic tests were adopted to assess the convergence of the MCMC algorithm (**Supplementary Note**). Traits with results that did not pass our convergence tests were those with the smallest sample sizes, $\hat{h}^2_{SNP}$ close to zero, or both (**Supplementary Fig. 3**). We focus on the results of 28 traits that passed both convergence tests for all of the three genetic architecture parameters. These traits include 24 quantitative traits, 2 diseases: major depressive disorder (MDD) and type 2 diabetes (T2D), and 2 categorical traits: male pattern baldness (MPB) and years of schooling (educational attainment). **Supplementary Fig. 4** shows the distributions of the estimates across these traits for the three genetic architecture parameters.

Comparison of the genetic architecture between Height and BMI

The genetic architectures of height and BMI have been relatively well studied compared to other complex traits[34-40]. Thus, it is interesting to compare our results for height and BMI (**Fig. 2**) with the previous findings. Both traits have a large sample size in the UKB: N = 126,545 for height and N = 126,389 for BMI. For both traits, a negative S was detected with extremely high significance level ($P = 1.8\times10^{-106}$ for height and $P = 2.8\times10^{-14}$ for BMI), meaning that lower-MAF variants tend to have larger effect size (absolute values). These results suggest that both height- and BMI-associated SNPs have been under selection, in line with the conclusions drawn from two recent studies[36,39]. The posterior mode of S was -0.422 (s.e. = 0.019) for height, remarkably lower than that of -0.295 (s.e. = 0.039) for BMI, suggesting that the proportion of genetic variation attributable to SNPs with low MAF for height is larger than that for BMI, consistent with the result from a previous study[36]. These results also imply that overall height-associated SNPs are under stronger selection than BMI-associated SNPs. The phenotypic variance explained by common SNPs (MAF > 0.01) was 52.8% (s.e. = 0.3%) for height and 27.7% (s.e. = 0.4%) for BMI, consistent with the estimates of h^2_{SNP} for height and BMI based on common SNPs reported previously[34-36]. The posterior distribution of π provides an estimate of 4.8% (s.e. = 0.1%) of SNPs having nonzero effects on height, significantly lower than that of 9.4% (s.e. = 0.5%) for BMI. These results suggest that BMI is more polygenic but less heritable than height, consistent with the results from a recent study using BayesR, a Bayesian multi-

component mixture model[41]. As a consequence, a BMI-associated SNP on average would explain a relatively smaller proportion of h^2_{SNP}, compared with a height-associated SNP, which may explain the higher uncertainty in the estimates of the hyperparameters such as S and π for BMI. These results also explain why the number of genome-wide significant SNPs (m_{GWS}) identified from the GIANT meta-analysis for BMI ($m_{GWS} = 97$) was smaller than that for height ($m_{GWS} = 697$) despite the fact that the sample size for BMI (N =~ 340,000)[35] was considerably larger than that for height (N = ~250,000)[34].

Inference on natural selection

Of the 28 traits that passed our convergence tests, 23 traits (including reproductive, cardiovascular and anthropometric traits and educational attainment) had significant negative S estimates with $\Pr(S < 0|\mathcal{D}) = 1$ and $P < 0.05/28$ (**Supplementary Table 2**), providing strong evidence that these traits have been under selection. The estimates of S over traits ranged from -0.601 (age at menopause) to 0.016 (fluid intelligence score) with mean -0.348, median -0.364 and s.d. 0.112. Interestingly, all the significant estimates of S were negative (see below for forward simulation to infer the type of selection from the sign of S). The magnitudes of $\hat{S}$, *i.e.* $|\hat{S}|$, reflects the strength of selection on the trait-associated SNPs. Traits with the largest $|\hat{S}|$ are related to fertility and heart function (**Fig. 3**), including age at menopause ($\hat{S} = -0.601$, s.e. = 0.073), pulse rate ($\hat{S} = -0.481$, s.e. = 0.048), waist circumference adjusted for BMI (WCadjBMI, $\hat{S} = -0.436$, s.e. = 0.036) and waist-hip ratio adjusted for BMI (WHRadjBMI, $\hat{S} = -0.436$, s.e. = 0.049). It has been reported that WCadjBMI and WHRadjBMI are associated with cardiovascular events[42], and WHRadjBMI is strongly correlated with pregnancy rate[43]. Other reproductive and cardiovascular traits, such as age at first live birth, age at menarche and blood pressure, had relatively high $|\hat{S}|$ as well. Thus, our results suggest that reproductive and cardiovascular traits are closely related to fitness and the SNPs that are associated with these traits have been under relatively stronger selection than SNPs associated with other traits.

Height ($\hat{S} = -0.422$), handgrip strength (right: -0.404, left: -0.374), lung function related traits (-0.405 – -0.362), heel bone mineral density (-0.394) and basal metabolic rate (-0.367) had a moderate to high $|\hat{S}|$ (**Fig. 3** and **Supplementary Table 2**). Evidence of selection for height has been reported from multiple studies using different approaches[36-40]. The two diseases, MDD and T2D, had negative $\hat{S}$ but the P-values did not reach FWER significance threshold, although the posterior probability of $S < 0$ for T2D was as high as 0.983. However, the power to detect a significant $\hat{S}$ may not be comparable to those quantitative traits, given the number of cases was less than 10,000 for each. A recent large-scale GWAS based on whole-genome sequencing data

also did not detect a signal of selection on T2D-associated variants[8]. Fluid intelligence score is the only trait with $\hat{S}$ at almost zero ($\hat{S} = 0.016$, s.e. = 0.096), which seems to suggest that fluid intelligence (FI) is not pertinent to fitness. However, there is strong evidence of negative selection on the SNPs associated with educational attainment (EA, $\hat{S} = -0.350$, s.e. = 0.055), which is thought to be a proxy of intelligence. Indeed, the genetic correlation between EA and FI was as high as 0.665 (s.e. = 0.052) estimated from a bivariate LD score regression[44]. Thus, it may be due to the limited statistical power that we did not detect the signal of selection for FI.

For traits with a significant estimate of S, we demonstrated the relationship between effect size and MAF by a plot of the cumulative genetic variance explained by SNPs (V_g) against MAF (**Fig. 4**), where MCMC samples of SNP effects were used to compute V_g for SNPs with MAF smaller than a threshold on the *x*-axis (**Supplementary Note**). Under an evolutionarily neutral model, V_g is linearly proportional to MAF[45] (diagonal line), therefore the area under the curve (AUC) is 0.5. All traits with significant estimates of S had the curve of cumulative genetic variance above the diagonal line, with $|\hat{S}|$ highly correlated with the AUC (r = 0.902), an alternative way of illustrating the evidence of natural selection.

Inference on SNP-based heritability

The 28 traits had low to moderate estimates of h^2_{SNP} with mean 0.221, median 0.212, and s.d. 0.093, and were all significantly above zero (**Supplementary Table 3**). Note that traits with $\hat{h}^2_{SNP}$ close to zero had failed in MCMC convergence tests, therefore the mean h^2_{SNP} estimate across traits is likely to be inflated. For MDD ($\hat{h}^2_{SNP} = 0.111$, s.e. = 0.021) and T2D ($\hat{h}^2_{SNP} = 0.222$, s.e. = 0.015), the estimates were on the liability scale and were converted from the observed scale [15], assuming a population prevalence of 15%[46] and 3%[47], respectively. The sorted estimates across traits are shown in **Supplementary Fig. 5**. Besides height ($\hat{h}^2_{SNP} = 0.528$), traits with the highest $\hat{h}^2_{SNP}$ include basal metabolic rate (0.336), which has been reported to be 0.2—0.4 in model animals[48], and MPB (0.335), which has been reported to be a highly heritable trait in both pedigree[49] and genomic[50] analysis. Traits with the lowest $\hat{h}^2_{SNP}$ include mean time to correctly identify matches (0.081), MDD (0.111), birth weight (0.114) and neuroticism score (0.125), in line with the low estimates of h^2_{SNP} from previous studies in MDD[51] and neuroticism score[52]. Given that most published estimates were obtained using whole-genome imputed SNPs, they are likely to be slightly higher than our estimates that are only based on the SNPs on Affymetrix Axiom Genotyping Arrays. For example, a recent study[53] on educational attainment in UKB gave an estimate of 0.21 (s.e. = 0.006), slightly higher than our estimate of 0.182 (s.e. = 0.004). Our estimate of 0.528 (s.e. = 0.003) for height is slightly but not significantly lower than

that of 0.56 (s.e. = 0.023) in Yang *et al.*[36]. For BMI, our estimate of 0.277 (s.e. = 0.004) is highly consistent with that of 27% (s.e. = 2.5%) in Yang *et al.*[36]. Across traits, $\hat{h}^2_{SNP}$ seems to be independent of either $\hat{S}$ or $\hat{\pi}$ but the s.e. of $\hat{S}$ and $\hat{\pi}$ decrease as $\hat{h}^2_{SNP}$ increases (**Supplementary Fig. 6**).

Inference on polygenicity

The distribution of $\hat{\pi}$ had mean 5.9%, median 5.5% and s.d. 3.6% across traits, and ranged from 0.6% (s.e. = 0.1%) to 13.6% (s.e. = 1.3%) (**Supplementary Table 4**). This suggests that all the 28 complex traits are polygenic with ~30,000 common SNPs with nonzero effects on average. Note that our simulation above suggests that this is likely to be an overestimation of the number of causal variants (**Fig. 1**). Interestingly, age at menopause, the trait with highest magnitude of $\hat{S}$ (-0.601), had the lowest estimate of polygenicity $\hat{\pi}$ (0.6%, s.e. = 0.1%) (**Supplementary Fig. 5**). Educational attainment had the highest $\hat{\pi}$ (13.6%, s.e. = 1.3%), which is reasonable because it is a compound trait of several sub-phenotypes so that many SNPs have an effect. It is followed by age at first live birth ($\hat{\pi}$ = 13.3%, s.e. = 2.5%), body fat percentage ($\hat{\pi}$ = 11.1%, s.e. = 0.8%) and BMI ($\hat{\pi}$ = 9.4%, s.e. = 0.5%). On the contrary, these traits had low to moderate magnitude of $\hat{S}$.

Analysis of gene expression traits in the CAGE data

Analysing expression levels of all probes in the CAGE dataset[27] (1,748 unrelated individuals of European ancestry) using the standard BayesS approach would be computationally challenging, as it would require us to perform 36,778 distinct BayesS analyses. However, given that many probes have a very limited number of associated SNPs, we developed a nested version of the BayesS model. This nested approach speeds up the analyses by considering SNPs in proximity collectively as a window, which allows for fast "jumping" over windows with zero effect (**Online Methods**). We showed by simulation that the nested model produces similar results as the standard BayesS approach in the analyses of both simulated (**Supplementary Fig. 7**) and UKB data (**Supplementary Fig. 8**) while being six times as fast as the standard BayesS approach for traits with low polygenicity (**Supplementary Fig. 9**). Using the nested BayesS model, we were able to fit 1,066,738 imputed SNPs (MAF > 1% and in common with those on HapMap3[54]) for the gene expression traits by partitioning the genome into 12,937 non-overlapping 200-Kb segments. Thus, the polygenicity (π) is interpreted as the proportion of segments with nonzero effects in nested BayesS.

After convergence tests, 27,869 probes remained, most of which had low $\hat{h}^2_{SNP}$ (mean = 0.147, median = 0.122 and s.d. = 0.088) and polygenic architecture (mean π = median π = 5.2% and s.d. = 2.3%) (**Supplementary Fig. 10**). With unrelated individuals only, our $\hat{h}^2_{SNP}$ were moderately

correlated with the GREML estimates (r = 0.568, **Supplementary Fig. 11**) despite the relatively small sample size. The estimates of polygenicity $\hat{\pi}$ suggest widespread trans-regulatory effects on gene expression in humans. To identify genes under selection, we mapped 21,303 out of the 27,869 probes to the genome with at least "good" probe annotation quality[55], which tagged 15,615 genes. Applying a Bonferroni correction for the number of probes mapped to the genome, we identified 32 probes that had $\hat{S}$ significantly different from zero ($P < 0.05/21{,}303 = 2.3\times10^{-6}$; **Fig. 5** and **Supplementary Table 5**). These probes were mapped to 30 unique genes (**Fig. 6**) and all had negative $\hat{S}$ (mean = -1.259, s.d. = 0.185), moderate $\hat{h}^2_{SNP}$ (mean = 0.412, s.d. = 0.075) and small $\hat{\pi}$ (mean = 0.0268, s.d. = 0.011). The alternative approach to control false positives is to limit PFP, which is less stringent but more powerful compared with limiting FWER. With this approach, a number of additional probes were identified with $\Pr(S < 0|\mathcal{D}) \geq 0.95$, giving a significant set of 266 probes for which 67 probes had $\Pr(S < 0|\mathcal{D}) = 1$ (**Fig. 5**). After mapping these probes to genes, a total of 252 genes were identified with the proportion of false positives < 5%. The results of gene ontology (GO) over-representation tests showed that these genes were enriched in the molecular function of IgG binding (P = 0.032 after Bonferroni correction). Moreover, we detected 45 genes that had $\Pr(S > 0|\mathcal{D}) \geq 0.95$ (**Fig. 5**), which were enriched in the molecular function of α_1-adrenergic receptor activity (P = 0.048) and potassium channel activity (P = 0.016). These results are consistent with a previous review[56] that a proportion of genes showing evidence of selection were significantly enriched in the function of immunity, receptor and potassium channel activity.

The directions of parameter *S* under different types of natural selection

Besides detecting selection and quantifying its strength on the trait-associated SNPs, the sign of S allows us to further infer the type of selection. To demonstrate this, we used forward simulations (**Online Methods**) to simulate common types of natural selection for a quantitative trait by relating the normally distributed phenotype to fitness through a hypothetical function (**Fig. 7**, top row). In the last generation of selection, the relationship between the variance σ_j^2 in the effect of coded allele and its frequency showed different patterns across different types of selection (**Fig. 7**, bottom row). As expected, when all the variants were selectively neutral, σ_j^2 was uniformly distributed across MAF (S = 0). Under stabilizing selection, σ_j^2 was negatively related to MAF (S < 0), a result of purifying trait-associated variants with large effect size which was deleterious to fitness through pleiotropy (also known as negative selection). Both directional (in either direction) and disruptive selection led to a positive relationship between σ_j^2 and MAF (S > 0). This is because in both cases, alleles with favourable effects increased in

frequency due to positive selection, so that high MAF bins were enriched with derived alleles of large effect. The difference is that disruptive selection kept the alleles with large effects at the intermediate frequencies, while directional selection persistently drove them toward fixation, resulting in a sigmoidal or convex shape of the relationship between σ_j^2 and MAF (**Supplementary Fig. 12**). In conclusion, estimate of S is informative to detect the signature of natural selection and is able to distinguish stabilising selection from directional and disruptive selection for a trait. At the level of genetic variants, a negative (positive) value of S is indicative of negative (positive) selection on the variants associated with the trait.

Discussion

We infer the action of natural selection on a complex trait from the signature left in the genetic architecture – the relationship between effect size and MAF. We introduced a method to simultaneously estimate the SNP-based heritability, polygenicity and the relationship between effect size and MAF using all genome-wide SNPs. In contrast to the contemporary methods that use independent SNPs that are significantly associated with traits[3,37,57,58], our method accounts for genome-wide SNP effects jointly and therefore has more statistical power to detect the signature of selection for polygenic traits. Results of the simulations using real genotypes showed that our estimate of the relationship (S) is unbiased when the causal variants are observed; otherwise, the estimate tends to be conservative depending on the LD between SNPs and the causal variants (**Fig. 1**). We detected significant signatures of natural selection ($S \neq 0$) for 23 out of 28 complex traits in the UKB data, with the strongest selection signals from the reproductive and cardiovascular trait-associated SNPs, followed by those associated with height, handgrip strength, lung function and other anthropometric traits as well as educational attainment (**Fig. 3**). Our findings are in line with an increasing body of literature supporting the hypothesis of widespread polygenic selection on standing variants in complex traits[4,39,40,59,60]. Together with the high prevalence of selection signals across traits (23/28 = 82%), our observation of high degree of polygenicity (~6% on average) underlines the role of pleiotropy in the action of natural selection.

In the analysis of the UKB data, all the significant estimates of S for 23 traits were negative (**Fig. 3**), consistent with a model of negative selection (**Fig. 7**). The evidence of negative selection against the trait-associated variants has been previously reported in some of these traits, such as height and BMI[36]. A recent study on ~110,000 Icelanders also detected negative selection on EA-increasing variants over recent generations, as a result of delayed reproduction and fewer children for the people with higher EA[61]. To support our results on some of the other traits, we

used the imputed SNPs based on a reference panel constructed by Haplotype Reference Consortium[62] to estimate the genetic variance across SNPs that are stratified by MAF and LD scores (GREML-LDMS[36]). We found that the genetic contribution of rare SNPs (MAF < 0.01) is disproportional to that of common SNPs (MAF > 0.01) in height, BMI, WHR, and diastolic blood pressure (**Supplementary Table 6**). These results also suggest that negative selection is pervasive across traits, in line with the conclusion drawn from the BayesS analysis.

In the analyses of CAGE data, we identified 30 genes showing significant signatures, all of which had negative $\hat{S}$ (**Fig. 6**). With a less stringent criterion for hypothesis testing, we identified additional 267 genes but only 45 of them had positive $\hat{S}$ (**Fig. 5**). These results again suggest the predominant role of negative selection in the human genome[59,63,64] and support the hypothesis that gene expression evolves primarily under stabilizing selection[65,66]. The genes that showed evidence of negative selection in our analysis may be functionally important and may deserve a downstream study. There are gene-level metrics, such as Residual Variation Intolerance Score (RVIS)[67] and Gene Damage Index (GDI)[68], aiming to prioritize genes for disease involvement based on the functional variation within a gene, which can be used to infer the strength of natural selection on the (coding) sequence of the gene. In contrast, our method interrogates genome-wide SNPs to detect signals of selection on the SNPs associated with the expression level of a gene, which largely depend on the trans-effects for polygenic transcripts. We found that $\hat{S}$ is not correlated with either RVIS or GDI (**Supplementary Fig. 13**). However, some genes indeed showed strong evidence of selection in both lines. For example, *HERC2* ($\hat{S}$ = -1.16, RVIS = -5.99) is a pigmentation-related gene which has been suggested a target of recent selection[69,70]. In addition, genes with significantly negative $\hat{S}$ generally also had low GDI, which is considered to be an indicator of the relative biological indispensability[68]. We do not expect to detect all genes that are known to be under selection, such as the lactase gene *LCT* ($\hat{S}$ = 0.221, s.e. = 0.317). One possible reason is that the signatures of selection for these genes are concentrated in the cis-regions and therefore might be diluted when we use all genome-wide SNPs to estimate S.

We conclude with several caveats. First, the polygenicity estimate ($\hat{\pi}$) only partially reflects the actual fraction of causal variants since SNPs can possess nonzero effects through LD with unobserved causal variants. Nevertheless, $\hat{\pi}$ can be used to compare the levels of polygenicity across traits. Second, the power of detecting a signal of natural selection (i.e. testing against S = 0) may improve if whole-genome sequence (WGS) or imputed sequence data, which include a large number of rare variants, are used in the analysis. However, it is computationally challenging to run BayesS on all the WGS variants in a large cohort like UKB, a common problem in the analysis of individual-level data with Bayesian methods. Depending on the sparsity of the

391 genetic signals, the nested BayesS provides a possibility to run the analysis in a manageable
392 amount of time but would still require a huge amount of memory to store the genotype matrix
393 for the WGS variants. A more practical approach is to model the genetic architecture using
394 summary statistics. Finally, in the simulation we observed a small inflation in estimating S using
395 the nested BayesS model, when all causal variants were genotyped and the true S is positive
396 (**Supplementary Fig. 7**). This suggests that the positive $\hat{S}$ may be slightly overestimated in the
397 CAGE data analysis, but it would not change our conclusions since there was no significant
398 positive $\hat{S}$. Given that most complex traits have negative estimates of the relationship between
399 effect size and MAF, we expect to discover more rare variants of large effect in future GWAS
400 using WGS or imputed data.
401

402 **Online Methods**

403 **The BayesS model.** BayesS is a Bayesian multiple regression that simultaneously fits all the
404 SNP effects as random:

405 $$\boldsymbol{y} = \mathbf{1}\mu + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{e}$$

406 where $\boldsymbol{y}$ is the vector of phenotypes, μ is the fixed effect, $\mathbf{X}$ is the matrix of SNP genotype scores
407 centred by the column means, $\boldsymbol{\beta}$ is the vector of SNP effects, and $\boldsymbol{e}$ is the residuals. The fixed
408 effect has a flat prior: $\mu \propto constant$. In practice, we fitted principal components and other
409 covariates as well in the model as fixed effects. It is common to standardize the SNP genotypes
410 such that each column of $\mathbf{X}$ has variance one. But we do not standardize the SNP genotypes, as
411 the standardization implicitly assumes a strong negative relationship between SNP effect size
412 and MAF (S = -1)[36,71-73]. As shown in Method Overview, we assume that the SNP effect β_j has a
413 hierarchical mixture prior

414 $$\beta_j \sim N\left(0, \left[2p_j(1-p_j)\right]^S \sigma_\beta^2\right)\pi + \phi(1-\pi)$$

415 where ϕ is a point mass at zero and π, the proportion of SNPs with nonzero effects, is the
416 polygenicity. We allow data to dominate the inference on polygenicity by assuming a uniform
417 prior

418 $$\pi \sim U(0,1).$$

419 The variance of SNP effects, which quantifies our prior belief on the effect size, is modelled to be
420 related to MAF p_j through S, which is assumed to have a normal prior

421 $$S \sim N(0, \sigma_S^2).$$

422 Namely, we *a priori* believe a selectively neutral model with some certainty (quantified by the
423 given variance) to allow the detection of selection to be driven by the data. We set $\sigma_S^2 = 1$ as the
424 prior in the analysis of UK Biobank traits, but a more informative prior $\sigma_S^2 = 0.1$ was used in the

analysis of CAGE gene expressions to shrink noise heavier toward zero given the much smaller sample size. The prior for the common variance factor is

$$\sigma_\beta^2 \sim \nu_\beta \tau_\beta^2 \chi_{\nu_\beta}^{-2}$$

where $\nu_\beta = 4$ and τ_β^2 is computed utilizing the characteristic of the distribution: if $\sigma^2 \sim \nu\tau^2\chi_\nu^{-2}$, then $E(\sigma^2) = \nu\tau^2/(\nu - 2)$. Rearranging the equation gives

$$\tau_\beta^2 = \frac{\nu_\beta - 2}{\nu_\beta} E\left(\sigma_\beta^2\right)$$

where

$$E\left(\sigma_\beta^2\right) = \frac{V_g}{\tilde{\pi} \sum_j [2p_j(1-p_j)]^{1+S}}$$

with $V_g, \tilde{\pi}$ and $\tilde{S}$ are the prior knowledge of the genetic variance, π and S. To remove the dependence of the hyperparameter τ_β^2 on the prior values of the genetic variance, π and S, we compute τ_β^2 deterministically using the sampled values of these parameters for the first 2,000 MCMC cycles, and then set τ_β^2 to the average value across these cycles. Likewise, the prior for the residual variance is

$$\sigma_e^2 \sim \nu_e \tau_e^2 \chi_{\nu_e}^{-2}$$

where $\nu_e = 4$ and $\tau_e^2 = \frac{\nu_e - 2}{\nu_e} V_e$ with V_e a prior knowledge of the residual variance. Note that when S = 0, our model becomes BayesCπ [19], a method that has been widely used for genomic prediction in agriculture, or Bayesian Variable Selection Regression (BVSR) in statistics literature[74]. The sampling scheme of the parameters is given in the **Supplementary Note**.

The nested BayesS model is developed based on a previously published method, BayesN[75], to speed up computation when a large number of SNPs is included in the analysis. In the nested BayesS, the genome is partitioned into *W*-Kb non-overlapping segments. Each window *a priori* has *k* SNPs with nonzero effects, where *W* and *k* are some given numbers. SNPs in the same window are individually modelled as in BayesS as well as collectively considered as a window effect with a normal-zero mixture prior. Remarkable speedups are obtained by "jumping" fast over the windows with zero effect, focusing solely on the windows that harbour genetic signals. Thus, the reduction in computing time is inversely related to the polygenicity, which is defined here as the proportion of segments with nonzero effects. When the causal variants are not observed, choosing $k > 1$ may lead to better performance in parameter estimation than BayesS, as it refines the signal of causal variant by allowing the flanking SNPs to jointly capture its effect. Details on the nested BayesS and the comparison with the standard BayesS are given in the **Supplementary Note**.

Estimation of heritability. We estimate the SNP-based heritability using the sampled values of SNP effects in MCMC. By definition, the genetic variance is the variance of the genetic values across individuals. In each MCMC cycle, we calculate the genetic values for each individual ($\tilde{g}_i$) using SNPs with sampled nonzero effects ($\tilde{\beta}_j$):

$$\tilde{g}_i = \sum_j X_{ij}\tilde{\beta}_j$$

Then, the genetic variance in the current cycle is

$$\tilde{\sigma}_g^2 = \frac{\sum_i \tilde{g}_i^2}{N} - \left(\frac{\sum_i \tilde{g}_i}{N}\right)^2$$

Conditional on the sampled value of residual variance ($\tilde{\sigma}_e^2$), the SNP-based heritability is

$$\tilde{h}_{SNP}^2 = \frac{\tilde{\sigma}_g^2}{\tilde{\sigma}_g^2 + \tilde{\sigma}_e^2}$$

The mean over all cycles after burn-in is the estimate of heritability

$$\hat{h}_{SNP}^2 = E(\tilde{h}_{SNP}^2)$$

The standard deviation of the MCMC samples gives the standard error of the estimate and the highest probability density gives the credible interval for posterior inference.

ARIC simulation analysis. The simulation based on Atherosclerosis Risk in Communities (ARIC) cohort[33] was used for testing the methods. We used PLINK 1.9[76] to carry out standard quality control (QC) procedures on the dataset, including removal of SNPs with missingness > 5%, Hardy-Weinberg equilibrium test $P < 10^{-6}$, or MAF < 1%, and removal of individuals with missing genotypes < 1% and genetic relationship < 0.05 estimated from all SNPs after QC using GCTA-GRM[77]. After all the QC steps, a total of 12,942 unrelated individuals and 564,959 SNPs remained. A quantitative trait was simulated by choosing 1,000 SNPs at random as causal variants with their effects sampled from a standard normal distribution. To simulate a spectrum of relationships between MAF and effect size, the effect size was multiplied by $[2p_j(1-p_j)]^S$ where $S = -2, -1, 0, 1,$ or 2, representing negative to positive relationship between effect size and MAF including the case of independence when $S = 0$. An individual phenotype specific to a given value of S was generated by adding a random normal residual with the variance identical to the genetic variance, giving each simulated trait a heritability of 0.5. The simulation process was repeated for 100 times. BayesS and the nested BayesS were used to analyse the simulated data with and without the causal variants in the model. To evaluate the robustness of our method to the starting values of parameters, we used an arbitrary value of 0 for S, 0.2 for heritability, and 0.05 for π, respectively to start the MCMC. In the nested BayesS, the length of window was set to be 200-Kb with 2 SNPs *a priori* fitted in the model. It is noteworthy that the distribution of genetic variance explained by each causal variant was not identical for different

scenarios of S in the true model. Under HWE, the genetic variance at locus j is $[2p_j(1-p_j)]^{S+1}\beta_j^2$ with $\beta_j \sim N(0,1)$ in the simulation. Compared to a trait with $S < 0$, a trait with $S > 0$ has a larger proportion of loci each explaining a small proportion of variance, given an identical distribution of MAF at the causal variants between traits (**Supplementary Fig. 14**). Thus, in the scenario of $S > 0$, it is more difficult to capture the causal variants by SNP markers if causal variants are not observed.

Analyis of the UK Biobank data. We have access to 46 complex traits in the UK Biobank[26], where the phenotype data were collected from over 500,000 individuals aged between 40 and 69 across the United Kingdom. The interim release contains genotypes for 152,736 samples at 806,466 SNPs on a customized Affymetrix Axiom array after QC procedures[78]. We selected a subset of 140,408 individuals that had a self-reported gender identical to the genetically inferred gender and a European ethnicity derived from a principal component analysis together with self-reported ethnicity. Furthermore, we removed individuals with genomic relatedness > 0.05 estimated from all SNPs using GCTA-GRM[77] and SNPs with genotype missing rate > 5%, Hardy-Weinberg equilibrium test $P < 10^{-6}$, or MAF < 1%. The final data set consisted of 126,752 individuals of European ancestry with 483,634 common SNPs (MAF > 1%). After removal of 5 duplicated traits and 5 traits with sample size (N) < 20,000, we had 36 traits remained for analysis, including 32 quantitative traits (anthropometric, cardiovascular and reproductive), 2 categorical traits – male pattern baldness (MPB) and years of schooling (educational attainment) and 2 diseases – type 2 diabetes (T2D) and major depressive disease (MDD). The sample sizes of the traits are shown in **Supplementary Table 2**, where most traits had N > 100,000. The prevalence of T2D and MDD in the sample was 5.35% and 6.70%, respectively. The phenotypes of quantitative traits were standardized within each sex group after regressing out the age effect. For educational attainment, the years of schooling are pre-adjusted by sex, a third order polynomial of year-of-birth and year-of-birth by sex interactions. We used BayesS for the analysis, where the first 20 principal components (PC) of GRM were fitted as fixed effects to account for the effects due to population stratification. For the disease traits, sex and age were fitted as covariates in addition to PCs, and for MPB, only age was fitted as the additional covariate.

Consortium for the Architecture of Gene Expression (CAGE) data set. We analyzed the mRNA levels for 36,778 transcript expression traits (probes) from the Consortium for the Architecture of Gene Expression (CAGE)[27] data set using the nested BayesS method. The CAGE data comprised of measurements from 36,778 gene expression probes in peripheral blood, with a subset of 1,748 unrelated (genomic relatedness > 0.05) European individuals from the total

2,765 individuals used for this analysis. Full details of the cohorts contributing to CAGE, and their sample preparation, normalization and genotype imputation are outlined in Lloyd-Jones *et al.*[27]. Briefly, the quantification of gene expression for each cohort was performed using the Illumina Whole-Genome Expression BeadChips. The gene expression levels in each cohort were initially normalized, followed by a quantile adjustment to standardize the distribution of expression levels across samples. We corrected for age, gender, cell counts and batch effects as well as hidden heterogeneous sources of variability. The rank-based inverse-normal transformation was used to normalize the measurements for each probed to be normally-distributed with mean 0 and variance 1. Probes measuring expression levels of genes located on chromosomes X and Y were removed from the analysis. The initial CAGE dataset consisted of seven unique cohorts that were genotyped on different SNP arrays. Therefore, genotype data were imputed to the 1000 Genomes Phase 1 Version 3 reference panel[79], resulting in 7,763,174 SNPs passing quality control of which 1,066,738 SNPs overlapped with HapMap3 and were used for analysis.

Forward simulation for different types of natural selection. We ran forward simulations using SLiM[80] to confirm that the relationship between effect size and MAF is subject to different types of natural selection. We simulated a 10-Mb region where new mutations occurred with probability of 0.95 to be neutral and of 0.05 to be a causal variant with an effect sampled from a standard normal distribution. The mutation rate was set to be 1.65×10^{-8}[81]. The phenotype of an individual was simulated based on the genotypic values at all segregating causal variants in the current generation with a heritability of 0.1. We simulated the evolution of a population of 1,000 individuals over 10,000 generations (this is equivalent to 100,000 generations in a population of 10,000 individuals[82]). The first 5,000 generations were used a burn-in period, where the phenotype did not affect fitness and all variants (including the causal variants) were under neutral variation. From generation 5,001, we related the standardized phenotype with mean zero and variance one to fitness through a hypothetical function that represents different types of selection (**Fig. 7**, top row). For directional selection, the phenotype was positively or negatively correlated to fitness through a simple linear function. For stabilizing selection, we used a normal curve to model that fitness achieved optimum at intermediate phenotype value. For disruptive selection, a reversed normal curve was used to model that the phenotypes at the two tails produced highest fitness. In the last generation of selection, we investigated the joint distribution of effects and frequencies of the derived alleles, the joint distribution of effects and frequencies of the coded alleles (arbitrarily chosen as in reality where derived alleles are unknown), and the relationship between the variance of the coded-allele effects and MAF. We collected results from 200 simulation replicates.

Acknowledgments

This research was supported by the Australian Research Council (DP160101343), the Australian National Health and Medical Research Council (1107258, 1078901, 1078037, and 1113400), and the Sylvia & Charles Viertel Charitable Foundation (Senior Medical Research Fellowship). R.d.V. acknowledges funding from an ERC consolidator grant (647648 EdGe, awarded to Philipp Koellinger). This study makes use of data from dbGaP (accession number: phs000090.v3.p1) and UK Biobank Resource (Application number: 12514). A full list of acknowledgements to these data sets can be found in the **Supplementary Note**. At the end, we wish to acknowledge The University of Queensland's Research Computing Centre (RCC) for its support in this research.

Author contributions

J.Y., P.M.V. and R.d.V. conceived the study. J.Y., J.Z. and P.M.V. designed the experiment. J.Z. derived the analytical methods, conducted all analyses and developed the software with assistance and guidance from J.Y., Y.W., M.R.R., L.L.J., L.Y.D., C.Y. A.X. and J.S.. L.L.J., A.F.M., J.E.P., G.W.M., A.M., T.E., G.G. and P.M.V. provided the CAGE data. J.Z. and J.Y. wrote the manuscript with the participation of all authors. All authors reviewed and approved the final manuscript.

References

1. Hansen, T.F., Alvarez-Castro, J.M., Carter, A.J., Hermisson, J. & Wagner, G.P. Evolution of genetic architecture under directional selection. *Evolution* **60**, 1523-36 (2006).
2. Pritchard, J.K. & Di Rienzo, A. Adaptation - not by sweeps alone. *Nat Rev Genet* **11**, 665-7 (2010).
3. Hancock, A.M. *et al.* Colloquium paper: human adaptations to diet, subsistence, and ecoregion are due to subtle shifts in allele frequency. *Proc Natl Acad Sci U S A* **107 Suppl 2**, 8924-30 (2010).
4. Pritchard, J.K., Pickrell, J.K. & Coop, G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr Biol* **20**, R208-15 (2010).
5. Smith, J.M. & Haigh, J. The hitch-hiking effect of a favourable gene. *Genet Res* **23**, 23-35 (1974).
6. Wright, S. The Distribution of Gene Frequencies in Populations of Polyploids. *Proc Natl Acad Sci U S A* **24**, 372-7 (1938).
7. Eyre-Walker, A. Evolution in health and medicine Sackler colloquium: Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. *Proc Natl Acad Sci U S A* **107 Suppl 1**, 1752-6 (2010).
8. Fuchsberger, C. *et al.* The genetic architecture of type 2 diabetes. *Nature* **536**, 41-7 (2016).
9. Park, J.H. *et al.* Distribution of allele frequencies and effect sizes and their interrelationships for common genetic susceptibility variants. *Proc Natl Acad Sci U S A* **108**, 18026-31 (2011).

10. Berndt, S.I. *et al.* Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nature Genetics* **45**, 501-U69 (2013).
11. Slatkin, M. Genotype-specific recurrence risks as indicators of the genetic architecture of complex diseases. *American Journal of Human Genetics* **83**, 120-126 (2008).
12. Liu, C.T., Raghavan, S. & Maruthur, N. Trans-ethnic meta-analysis and functional annotation illuminates the genetic architecture of fasting glucose and insulin. *The American Journal of ...* (2016).
13. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-7 (2014).
14. Manolio, T.A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747-53 (2009).
15. Lee, S.H., Wray, N.R., Goddard, M.E. & Visscher, P.M. Estimating missing heritability for disease from genome-wide association studies. *Am J Hum Genet* **88**, 294-305 (2011).
16. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* **42**, 565-9 (2010).
17. Yang, J. *et al.* Genome partitioning of genetic variation for complex traits using common SNPs. *Nature Genetics* **43**, 519-U44 (2011).
18. Gratten, J., Wray, N.R., Keller, M.C. & Visscher, P.M. Large-scale genomics unveils the genetic architecture of psychiatric disorders. *Nat Neurosci* **17**, 782-90 (2014).
19. Habier, D., Fernando, R.L., Kizilkaya, K. & Garrick, D.J. Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics* **12**, 186 (2011).
20. Moser, G. *et al.* Simultaneous discovery, estimation and prediction analysis of complex traits using a bayesian mixture model. *PLoS Genet* **11**, e1004969 (2015).
21. Meuwissen, T.H.E., Hayes, B.J. & Goddard, M.E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819-1829 (2001).
22. de Los Campos, G., Hickey, J.M., Pong-Wong, R., Daetwyler, H.D. & Calus, M.P. Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics* **193**, 327-45 (2013).
23. Zhou, X., Carbonetto, P. & Stephens, M. Polygenic modeling with bayesian sparse linear mixed models. *PLoS Genet* **9**, e1003264 (2013).
24. Lloyd-Jones, L.R. *et al.* Inference on the Genetic Basis of Eye and Skin Colour in an Admixed Population via Bayesian Linear Mixed Models. *Genetics* (2017).
25. Mancuso, N. *et al.* The contribution of rare variation to prostate cancer heritability. *Nat Genet* **48**, 30-5 (2016).
26. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* **12**, e1001779 (2015).
27. Lloyd-Jones, L.R., Holloway, A., McRae, A. & Yang, J. The genetic architecture of gene expression in peripheral blood. *The American Journal of ...* (2017).
28. Neal, R.M. MCMC Using Hamiltonian Dynamics. *Handbook of Markov Chain Monte Carlo*, 113-162 (2011).
29. Fernando, R.L., Dekkers, J.C. & Garrick, D.J. A class of Bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole-genome analyses. *Genet Sel Evol* **46**, 50 (2014).
30. Gelman, A., Carlin, J.B., Stern, H.S. & Rubin, D.B. Book Reviews. *Journal of the American Statistical Association* **109**, 1325-1337 (2014).
31. Fernando, R.L. *et al.* Controlling the proportion of false positives in multiple dependent tests. *Genetics* **166**, 611-9 (2004).
32. Storey, J.D. The optimal discovery procedure: a new approach to simultaneous significance testing. *Journal of the Royal Statistical Society Series B-Statistical Methodology* **69**, 347-368 (2007).
33. Psaty, B.M. *et al.* Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ Cardiovasc Genet* **2**, 73-80 (2009).

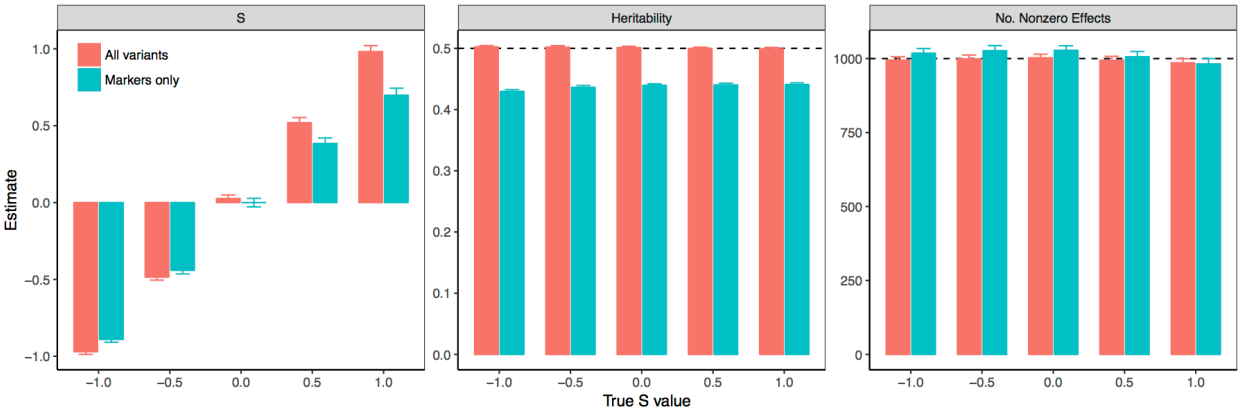
34. Wood, A.R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet* **46**, 1173-86 (2014).
35. Locke, A.E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197-206 (2015).
36. Yang, J. *et al.* Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat Genet* **47**, 1114-20 (2015).
37. Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499-503 (2015).
38. Marouli, E. *et al.* Rare and low-frequency coding variants alter human adult height. *Nature* **542**, 186-190 (2017).
39. Robinson, M.R. *et al.* Population genetic differentiation of height and body mass index across Europe. *Nat Genet* **47**, 1357-62 (2015).
40. Field, Y. *et al.* Detection of human adaptation during the past 2000 years. *Science* **354**, 760-764 (2016).
41. Matthew R. Robinson, G.E., Gerhard Moser, Luke R. Lloyd-Jones, Marcus A. Triplett, Zhihong Zhu, Ilja M. Nolte, Jana V. van Vliet-Ostaptchouk, Harold Snieder, The LifeLines Cohort Study, Tonu Esko, Lili Milani, Reedik Mägi, Andres Metspalu, Patrik K. E. Magnusson, Nancy L. Pedersen, Erik Ingelsson, Magnus Johannesson, Jian Yang, David Cesarini and Peter M. Visscher. Genotype-covariate interaction effects and the heritability of adult body mass index. *Nature Genetics* (In press).
42. Koning, D.L., Merchant, A.T. & Pogue, J. Waist circumference and waist-to-hip ratio as predictors of cardiovascular events: meta-regression analysis of prospective studies. *European heart journal* **28.7**, 850-856 (2007).
43. Wass, P., Waldenstrom, U., Rossner, S. & Hellberg, D. An android body fat distribution in females impairs the pregnancy rate of in-vitro fertilization-embryo transfer. *Hum Reprod* **12**, 2057-60 (1997).
44. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236-41 (2015).
45. Hill, W.G., Goddard, M.E. & Visscher, P.M. Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet* **4**, e1000008 (2008).
46. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *The Lancet* **381**, 1371-1379 (2013).
47. Das, S.K. & Elbein, S.C. The Genetic Basis of Type 2 Diabetes. *Cellscience* **2**, 100-131 (2006).
48. Konarzewski, M. & Ksiazek, A. Determinants of intra-specific variation in basal metabolic rate. *J Comp Physiol B* **183**, 27-41 (2013).
49. Nyholt, D.R., Gillespie, N.A., Heath, A.C. & Martin, N.G. Genetic basis of male pattern baldness. *J Invest Dermatol* **121**, 1561-4 (2003).
50. Liu, F. *et al.* Prediction of male-pattern baldness from genotypes. *Eur J Hum Genet* **24**, 895-902 (2016).
51. Hyde, C.L. *et al.* Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat Genet* **48**, 1031-6 (2016).
52. de Moor, M.H. *et al.* Meta-analysis of Genome-wide Association Studies for Neuroticism, and the Polygenic Association With Major Depressive Disorder. *JAMA Psychiatry* **72**, 642-50 (2015).
53. Davies, G. *et al.* Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N=112 151). *Mol Psychiatry* **21**, 758-67 (2016).
54. International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52-8 (2010).
55. Barbosa-Morais, N.L. *et al.* A re-annotation pipeline for Illumina BeadArrays: improving the interpretation of gene expression data. *Nucleic Acids Res* **38**, e17 (2010).
56. Bustamante, C.D. *et al.* Natural selection on protein-coding genes in the human genome. *Nature* **437**, 1153-7 (2005).

57. Casto, A.M. & Feldman, M.W. Genome-wide association study SNPs in the human genome diversity project populations: does selection affect unlinked SNPs with shared trait associations? *PLoS Genet* **7**, e1001266 (2011).
58. Berg, J.J. & Coop, G. A population genetic signal of polygenic adaptation. *PLoS Genet* **10**, e1004412 (2014).
59. Hernandez, R.D. *et al.* Classic Selective Sweeps Were Rare in Recent Human Evolution. *Science* **331**, 920-924 (2011).
60. Turchin, M.C. *et al.* Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nature Genetics* **44**, 1015-+ (2012).
61. Kong, A. *et al.* Selection against variants in the genome associated with educational attainment. *Proc Natl Acad Sci U S A* **114**, E727-E732 (2017).
62. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279-83 (2016).
63. McVicker, G., Gordon, D., Davis, C. & Green, P. Widespread genomic signatures of natural selection in hominid evolution. *PLoS Genet* **5**, e1000471 (2009).
64. Sohail, M. *et al.* Negative selection in humans and fruit flies involves synergistic epistasis. *Science* **356**, 539-542 (2017).
65. Romero, I.G., Ruvinsky, I. & Gilad, Y. Comparative studies of gene expression and the evolution of gene regulation. *Nat Rev Genet* **13**, 505-16 (2012).
66. Gilad, Y., Oshlack, A. & Rifkin, S.A. Natural selection on gene expression. *Trends Genet* **22**, 456-61 (2006).
67. Petrovski, S., Wang, Q., Heinzen, E.L., Allen, A.S. & Goldstein, D.B. Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet* **9**, e1003709 (2013).
68. Itan, Y. *et al.* The human gene damage index as a gene-level approach to prioritizing exome variants. *Proc Natl Acad Sci U S A* **112**, 13615-20 (2015).
69. Grossman, S.R. *et al.* A composite of multiple signals distinguishes causal variants in regions of positive selection. *Science* **327**, 883-6 (2010).
70. Chen, G.B., Lee, S.H., Zhu, Z.X., Benyamin, B. & Robinson, M.R. EigenGWAS: finding loci under selection through genome-wide association studies of eigenvectors in structured populations. *Heredity (Edinb)* **117**, 51-61 (2016).
71. Lee, S.H. *et al.* Estimation of SNP heritability from dense genotype data. *Am J Hum Genet* **93**, 1151-5 (2013).
72. Speed, D., Hemani, G., Johnson, M.R. & Balding, D.J. Improved heritability estimation from genome-wide SNPs. *Am J Hum Genet* **91**, 1011-21 (2012).
73. Speed, D. *et al.* Reevaluation of SNP heritability in complex human traits. *Nature Genetics* (2017).
74. Guan, Y. & Stephens, M. Bayesian variable selection regression for genome-wide association studies and other large-scale problems. *The Annals of Applied Statistics*, 1780-1815 (2011).
75. Zeng, J. Whole genome analyses accounting for structures in genotype data. *PhD diss., IOWA STATE UNIVERSITY* (2015).
76. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
77. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).
78. Ukbiobank, U.K. Genotyping and quality control of UK Biobank, a large-scale, extensively phenotyped prospective resource. *Available at biobank. ctsu. ox. ac. uk/crystal/docs/genotyping_qc. pdf. Accessed April 1* (2015).
79. 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65 (2012).
80. Messer, P.W. SLiM: simulating evolution with selection and linkage. *Genetics* **194**, 1037-9 (2013).

766 81. Palamara, P.F. *et al.* Leveraging Distant Relatedness to Quantify Human Mutation and
767 Gene-Conversion Rates. *Am J Hum Genet* **97**, 775-89 (2015).
768 82. Enard, D., Messer, P.W. & Petrov, D.A. Genome-wide signals of positive selection in
769 human evolution. *Genome Res* **24**, 885-95 (2014).
770

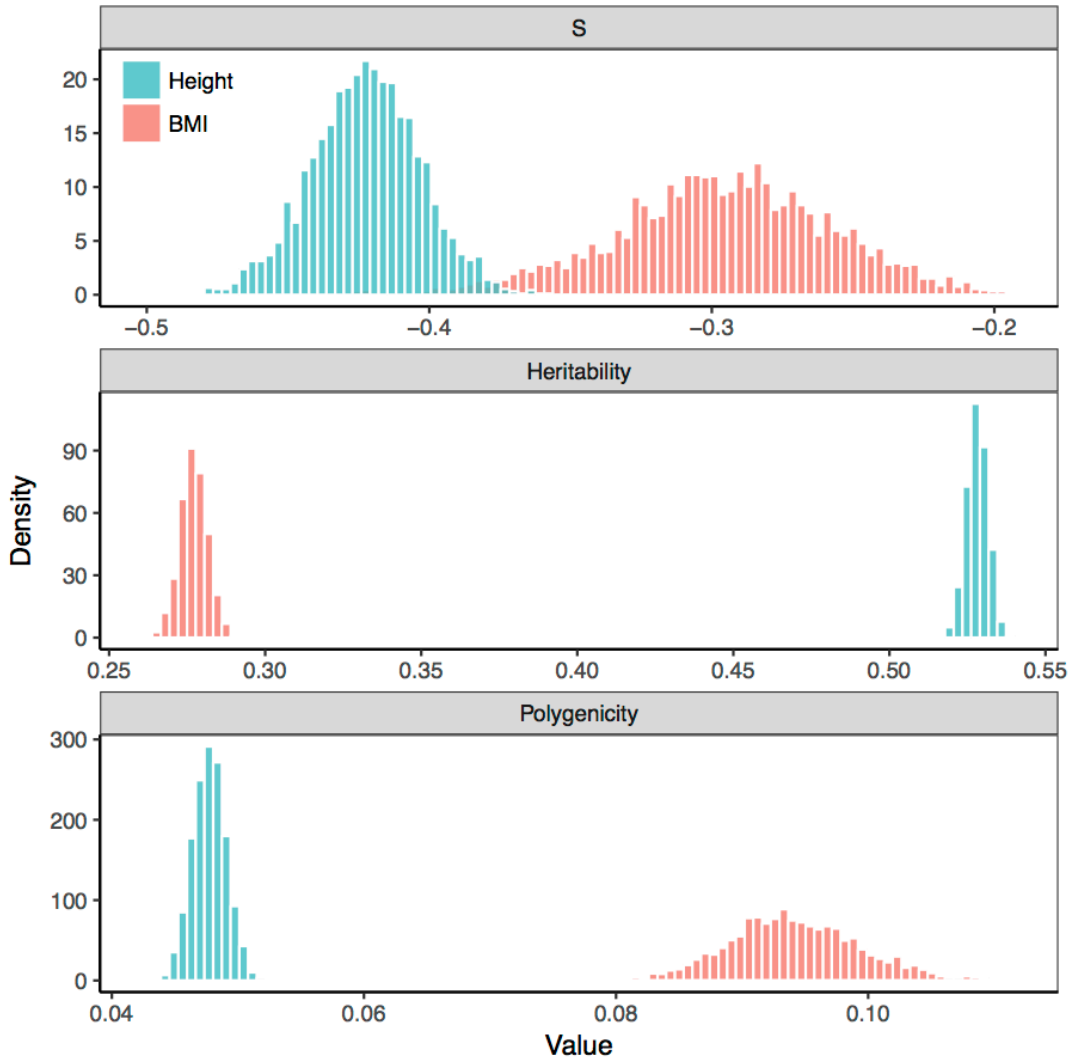
771 **Figures**

772 Figure 1: Estimation of the genetic architecture parameters, e.g. *S*, heritability and polygenicity, for a
773 simulated trait using the ARIC data. Results are the mean estimates with s.e.m. (cap) over 100
774 simulation replicates for a spectrum of *S* parameter values. Colour indicates the results of BayesS
775 with both causal variants and SNP markers (red) or with SNP markers only (blue). The heritability at
776 the 1,000 randomly selected causal variants was 0.5. The number of nonzero effects is the number
777 of SNPs with nonzero effects averaged over MCMC iterations.
778



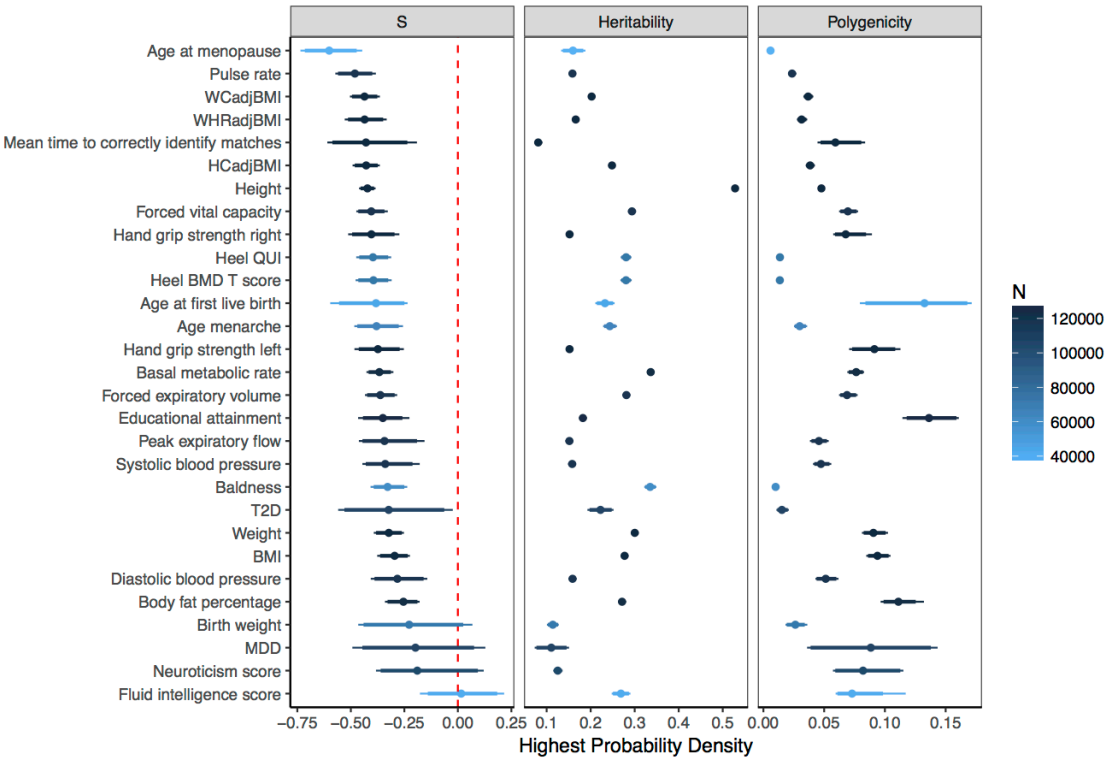
779
780

781 Figure 2: Posterior distributions of the genetic architecture parameters for height versus BMI using
782 data from UKB. S measures the relationship between SNP effect size and MAF. Polygenicity is
783 defined as the proportion of SNPs with nonzero effects.
784



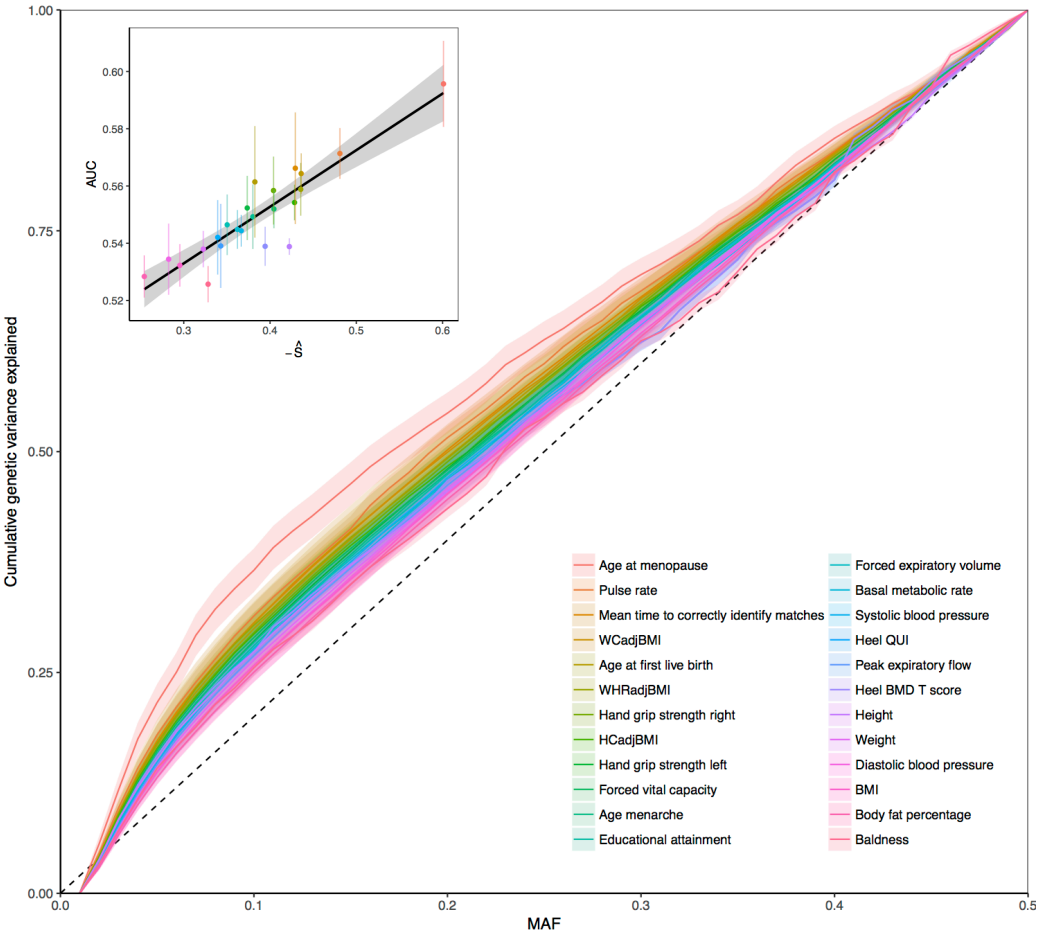
785

786 Figure 3: Posterior modes with credible intervals for the genetic architecture parameters using
787 BayesS. Results are for the 28 UKB complex traits that had passed convergence tests on the MCMC
788 chain. The bold line represents 95% credible interval (highest posterior density, HPD) and the thin
789 line represents 90% credible interval. Sample size *N* for each trait is shown by the colour gradient.
790 Polygenicity is defined as the proportion of genome-wide SNPs with nonzero effects on the trait.
791



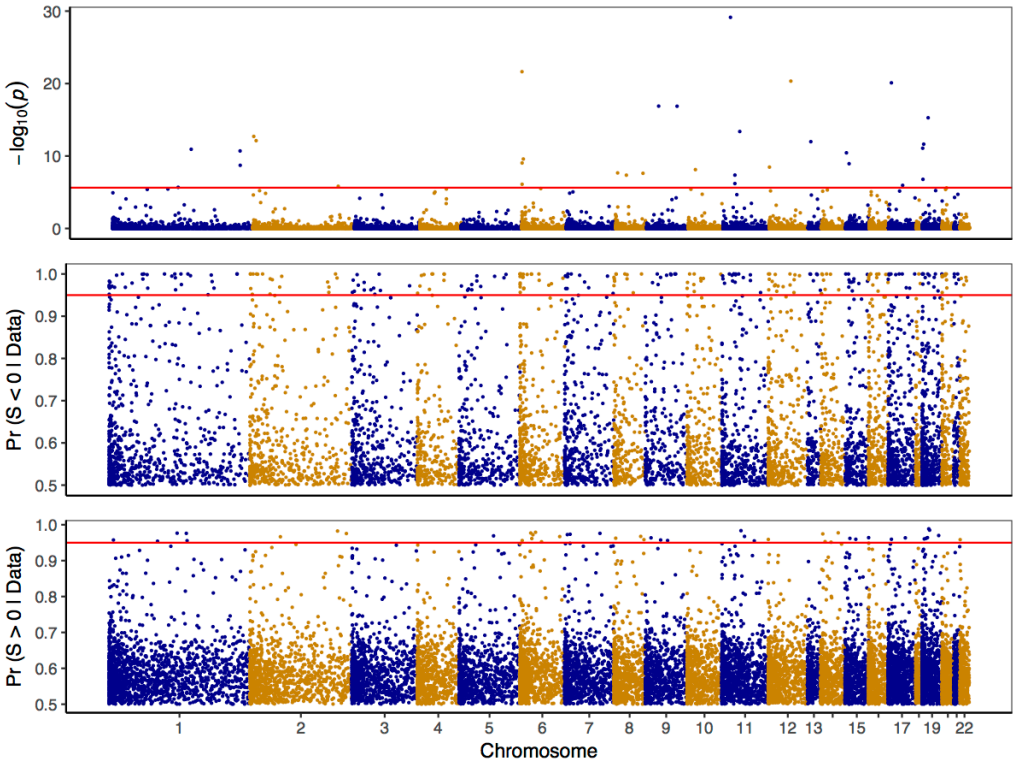
792

793 Figure 4: Cumulative genetic variance explained by SNPs with MAF smaller than a threshold on the *x*-
794 axis. The lines are the posterior means for the 23 UKB complex traits from UKB for which the
795 estimates of S were significantly different from zero. Shadow shows the posterior standard error.
796 The inner graph shows the relationship between the area under the curve (AUC) of the cumulative
797 genetic variance and negative $\hat{S}$ (bar shows s.e.) across traits.
798

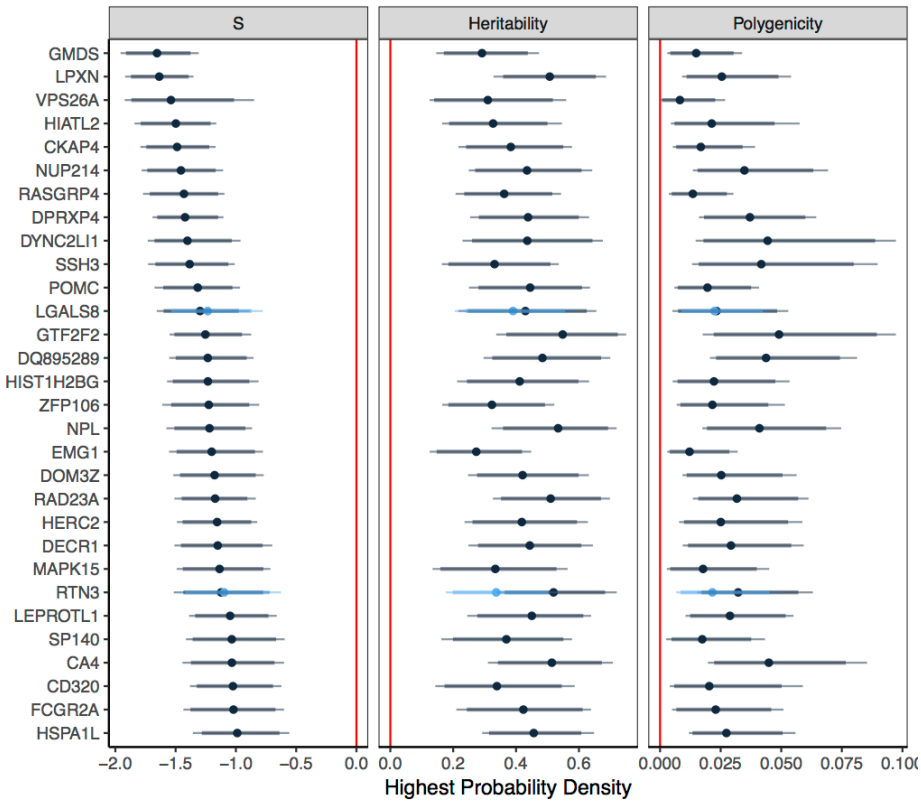


799

Figure 5: Genome-wide evidence of selection from the p-values to test against $S = 0$ and the posterior probability of $S < 0$ or $S > 0$ for 21,303 probes in the CAGE data after QC. The red line shows the significant threshold of 0.05 after Bonferroni correction (p-value = 2.3×10^{-6}) or 0.95 for the posterior probabilities.



806 Figure 6: Estimation of the genetic architecture parameters for 30 genes (corresponding to 32
807 probes) with significant $\hat{S}$ (p-value < 0.05/21,303) in the analysis of CAGE data. Results are the
808 posterior modes with credible intervals obtained from the nested BayesS model. The bold line
809 represents 95% credible interval (highest posterior density, HPD) and the thin line represents 90%
810 credible interval. Polygenicity is defined as the proportion of 200-Kb windows with nonzero effects
811 in the genome. The light colour shows the results of the second probe tagging one gene.
812



813
814

Figure 7: Forward simulations with different types of selection. A quantitative trait was generated based on a simulated chromosome segment of 10Mb (5% causal and 95% neutral mutations in each generation). The trait heritability was 0.1. The top row shows the functions used to relate the phenotype (normally distributed) to fitness in different modes of selection: neutral variation, directional selection with the phenotype positively (+) or negatively (-) correlated to fitness, stabilizing selection and disruptive selection. The 2rd row shows the joint distributions of the coded allele effects and frequencies, where the coded allele at each causal variant was chosen at random from the derived and ancestral alleles, and the red line shows the means of coded allele effects in allele frequency intervals of 0.05. The bottom row shows the relationships between the variance of coded allele effects and MAF. Results were collected from 200 replicates of simulation.