Genetic Epidemiology

OFFICIAL JOURNAL
INTERNATIONAL GENETIC
EPIDEMIOLOGY SOCIETY
www.geneticepi.org

WILEY

RESEARCH ARTICLE

A flexible and parallelizable approach to genome-wide polygenic risk scores

Paul J. Newcombe[1] | Christopher P. Nelson[2,3] | Nilesh J. Samani[2,3] | Frank Dudbridge[4]

[1]MRC Biostatistics Unit, School of Clinical Medicine, Cambridge Institute of Public Health, Cambridge Biomedical Campus, Cambridge, UK

[2]Department of Cardiovascular Sciences, Cardiovascular Research Centre, Glenfield Hospital, University of Leicester, Leicester, UK

[3]NIHR Leicester Biomedical Research Centre, Glenfield Hospital, Leicester, UK

[4]Department of Health Sciences, Centre for Medicine, University of Leicester, Leicester, UK

Correspondence
Paul J. Newcombe, MRC Biostatistics Unit, School of Clinical Medicine, Cambridge Institute of Public Health, Forvie Site, Robinson Way, Cambridge Biomedical Campus, Cambridge CB2 0SR, UK.
Email: paul.newcombe@mrc-bsu.cam.ac.uk

Abstract

The heritability of most complex traits is driven by variants throughout the genome. Consequently, polygenic risk scores, which combine information on multiple variants genome-wide, have demonstrated improved accuracy in genetic risk prediction. We present a new two-step approach to constructing genome-wide polygenic risk scores from meta-GWAS summary statistics. Local linkage disequilibrium (LD) is adjusted for in Step 1, followed by, uniquely, long-range LD in Step 2. Our algorithm is highly parallelizable since block-wise analyses in Step 1 can be distributed across a high-performance computing cluster, and flexible, since sparsity and heritability are estimated within each block. Inference is obtained through a formal Bayesian variable selection framework, meaning final risk predictions are averaged over competing models. We compared our method to two alternative approaches: LDPred and lassosum using all seven traits in the Welcome Trust Case Control Consortium as well as meta-GWAS summaries for type 1 diabetes (T1D), coronary artery disease, and schizophrenia. Performance was generally similar across methods, although our framework provided more accurate predictions for T1D, for which there are multiple heterogeneous signals in regions of both short- and long-range LD. With sufficient compute resources, our method also allows the fastest runtimes.

KEYWORDS
Bayesian variable selection, meta-GWAS, polygenic risk scores, risk prediction, summary statistics

1 | INTRODUCTION

The heritability of most complex traits is driven by variation throughout the genome, with a large number of loci contributing small or modest effects (Dudbridge, 2013; 2016). Polygenic risk scores, which combine information on multiple variants genome-wide into weighted sums of trait-associated alleles, have been found to improve prediction of a variety of traits (Dudbridge, 2013; Evans, Visscher, & Wray, 2009; Pharoah, Antoniou, Easton, & Ponder, 2008; Purcell et al., 2009; Stahl et al., 2012; The International Multiple Sclerosis Genetics Consortium (IMSGC), 2010). Even for a modestly heritable trait such as breast cancer, a

comprehensive polygenic score could improve discriminatory power sufficiently for use in a targeted screening program (Pharoah et al., 2002). However, the predictive accuracy of current GWAS "hits" falls well short of what is theoretically possible based on familial and genomic heritability estimates (De Vries et al., 2015; Eriksson et al., 2015; Talmud et al., 2015; Wacholder et al., 2010). Realizing the full potential of polygenic risk prediction will require much larger sample sizes than offered by a typical single cohort GWAS (Chatterjee et al., 2013; Dudbridge, 2013; Wray et al., 2013). In a recent trend large-scale "meta-GWAS", comprising 10s of 1000s of people amassed over multiple studies, have boosted the power of genetic association studies, increasing the number of unambiguously associated regions into the 10s or even 100s for some traits (Lango Allen et al., 2010; Morris et al., 2012; Teslovich et al., 2010). Building predictive models from meta-GWAS results is therefore of great importance, since these consortia often represent the totality of GWAS information for a particular trait.

A simple yet surprisingly effective approach is to select a subset of variants according to a *p*-value threshold and build an additive score, weighting the contribution of individual variants according to their associations with the trait. This is easy to do from summary results alone, and requires minimal computation. The main drawback is that genetic correlations due to linkage disequilibrium (LD) are ignored, which leads to bias in the weights, and, consequently, suboptimal predictive accuracy. Ideally, multivariate regression would be used to estimate weights adjusted for LD. Unfortunately, however, this is complicated for two reasons. The first is that privacy concerns and the logistics of sharing data on such a large scale mean that meta-GWAS typically only conduct one-at-a-time association tests of each variant, based on simple summaries shared between the cohorts. The final univariate results ignore correlations among the variants, and consequently signals tagged by multiple correlated variants would be overrepresented in a polygenic risk score constructed according to a simple threshold on statistical significance. This issue is usually dealt with by pruning variants until they are approximately independent, although information is inevitably discarded (Dudbridge & Newcombe, 2015). The second issue relates to the challenge of dimensionality. Ideally, we would jointly model all important predictors, in order to account for genetic correlations due to LD. However, traditional regression methodology suffers from over-fitting when applied to large numbers of covariates; information is spread too thinly leading to unstable estimates with high standard errors. This latter issue inspired the development of lasso penalized regression by Tibshirani (1996), whereby a large number of predictors are jointly modelled with a penalty term included in the likelihood to encourage sparsity. The penalty term

modifies the likelihood of the regression coefficients, with a large penalty leading to the exclusion of many variables. Typically, the penalty is tuned through cross-validation such that covariates with negligible predictive effects are removed. The over-fitting problem is thus avoided and prediction is improved. Various extensions to the original method have been successfully applied in genomics to explore multi-single nucleotide polymorphism (SNP) models of disease (Vignal, Bansal, & Balding, 2011; Wu, Chen, Hastie, Sobel, & Lange, 2009) or to search for master predictors (Peng, Zhu, & Bergamaschi, 2010). Bayesian versions of the LASSO have also been described (Griffin & Brown, 2010; Park & Casella, 2008) and used for efficient variable selection in genetics (Bottolo et al., 2013; Newcombe, Conti, & Richardson, 2016; Servin & Stephens, 2007; Tachmazidou, Johnson, & De Iorio, 2010; Wallace et al., 2015). Attractive features of Bayesian sparse regression include inference of posterior probabilities for each predictor, posterior inference on competing combinations, and, potentially most importantly, the possibility of incorporating prior information into the analysis. In a related approach, the over-fitting problem has been addressed by recasting the animal model of classical quantitative genetics as a ridge regression model with a Gaussian prior on genetic effects. This does not in itself impose sparsity on the fitted model but has been extended in various ways to allow for a sparse component (Meuwissen, Hayes, & Goddard, 2001; Moser et al., 2015; Zhou, Carbonetto, & Stephens, 2013).

Although most sparse regression methods require individual level data, two frameworks have recently been proposed that allow construction of high-dimensional polygenic risk models from meta-GWAS summaries. "LDPred" is a sparse Bayesian regression framework in which multivariate weights are estimated using a combination of Markov Chain Monte Carlo (MCMC) and empirical Bayes (Vilhjálmsson et al., 2015). The framework is empirical Bayes in that a fixed value is estimated from the data for the residual variance (derived from a heritability estimate). A large proportion of variants are assumed to have no effect, and the proportion of causal variants is selected from a range of values according to performance in the validation data, with those variants assumed to have effects following a Gaussian distribution. The second method, "lassosum", uses a non-Bayesian penalized-regression framework with a Lasso type penalty (Mak et al., 2017). The penalization parameters are optimized according to predictive performance in the validation data, similarly to LDPred, but a "pseudo-validation" approach is also proposed to obtain near-optimal values of the two tuning parameters. Both LDPred and lassosum account for LD using genetic correlation estimates from external reference data. In a comparative

study they performed similarly, both outperformed simple approaches based on pruning and *p*-value thresholding, which is to be expected if there are multiple causal variants in LD (Dudbridge & Newcombe, 2015).

In this work, we propose a new two-step framework for constructing polygenic risk scores from summary data, which builds on our method "JAM", a sparse Bayesian regression model for multivariate fine-mapping from summary data (Newcombe et al., 2016). In the first step, models are fit independently to chromosomal blocks of limited size, and then combined in a second step to account for long-range LD. This obviates a practical need to define LD blocks as required, for example, by LDPred, but, more pertinently, provides the flexibility to capture highly complex genetic models, since sparsity and heritability are calibrated individually within each LD block at step one. A further advantage is that our approach is extremely parallelizable, since the block-specific analyses can be distributed across a high-performance cluster, offering potentially much faster performance when hundreds of computing cores are available. In comparison to penalized regression approaches such as lassosum, our approach provides predictions that are model averaged, that is, they reflect uncertainty in the best selection of variants since they are averaged over competing combinations. Notably, the use of Bayesian model averaging means that all SNPs may enter the polygenic score, even though each evaluated model is sparse.

2 | METHODS

Our aim is to build a high-dimensional sparse regression model using summary data. We start with a brief summary of our previous "JAM" model, which allows multivariate fine-mapping from univariate summary data, and how this framework can be used to construct polygenic risk models. Second, we describe a two-step extension for the adjustment of long-range LD, facilitating genome-wide application.

2.1 | Inference of multivariate polygenic weights from univariate summary data

We start by defining the standard multivariate linear regression of a vector of *n* trait values y on *P* variants in the columns of an $n \times P$ genotype matrix X:

$$y \sim N(X\beta, \sigma^2) \tag{1}$$

β denotes the *P*-length vector of multivariate, that is, correlation-adjusted, genetic effects, and σ^2 denotes the

residual variance. Note that both trait and genotypes are mean-centred, allowing a simplification of the standard regression model to exclude the intercept term. Our previous summary data method JAM (Newcombe et al., 2016) is based on the following model:

$$X'y \sim MVN(X'X\beta, X'X\sigma^2) \tag{2}$$

which is derived from Equation (1) by multiplying through by X'. $X'y$, which has as many elements as genetic variants, can be derived from univariate effect estimates from regressions of each variant and the trait (Newcombe et al., 2016; Verzilli et al., 2008; Yang et al., 2012), as are reported by a typical GWAS or meta-GWAS. Using a plug-in estimate for the genetic correlation matrix $X'X$, as obtained from a reference data set, the model depicted in Equation (2) can therefore be fitted using summary data. Crucially, inference is obtained for the same correlation-adjusted vector of genetic effects, β. To ease the computational burden, we invoke a Cholesky decomposition of $X'X$ to map Equation (2) to a set of independent Gaussian distributions. The Cholesky decomposition of $X'X$ provides an upper triangular and therefore invertible matrix, L, which satisfies:

$$X'X = L'L$$

Multiplying Equation (2) through by L'^{-1}, we obtain:

$$L'^{-1}X'y \sim MVN(L\beta, \sigma^2) \tag{3}$$

that is, a model with the same form as a standard linear regression with independent residual errors.

2.1.1 | Sparsity inducing prior on which variants have predictive effects

To avoid over-fitting in the context of potentially many genetic effects, we use a Bayesian sparse regression framework to draw inference under the model described by Equation (3). This is facilitated by introducing a latent binary vector $\gamma = (\gamma_1, \dots \gamma_P)$ of indices for whether each variant has a nonzero effect, that is, included in the model. Denoting the proportion of variants with nonzero effects as π, sparsity is induced via a β prior distribution:

$$\pi \sim Beta(1, \lambda P) \tag{4}$$

This prior formulation is widely used in Bayesian variable selection due to the intrinsic correction for multiplicity; the marginal prior odds of any single

variable having an effect is $1/\lambda P$, and therefore decreases with the total number of variables P, whereas the global prior odds of any effect is constant at $1/\lambda$ (Wilson, Iversen, Clyde, Schmidler, & Schildkraut, 2010). λ can be chosen to induce more or less sparsity, depending on prior beliefs. In practice, we recommend trying a range of several λ and picking the value which optimizes predictive performance in the validation data. For the analyses presented in the results, we tried $\lambda = 0.001, 0.01, 0.1, 1$. Note that λ is a hyperparameter for the distribution of π, the proportion of variants with nonzero effects, and conditional on λ, we allow the proportion π to be random; thus we allow greater flexibility than approaches that consider a set of fixed values for π (Vilhjálmsson et al., 2015)

2.1.2 | Prior on variant effects

Conditional on a selection of variants indicated in γ, we place a hierarchical normal prior over the corresponding subvector of multivariate effects, which we denote by β_γ:

$$\beta_\gamma \sim N(0, I\sigma_\beta^2)$$

σ_β^2 may be interpreted as the variance among the "true" genetic effects. To estimate this crucial parameter largely from the data, we assign a vague hyper-prior (rather than choosing a fixed value), which we have used previously in an 'omics setting (Newcombe et al., 2014):

$$\sigma_\beta \sim Unif(0.05, 2)$$

Our polygenic prediction model is completed with a prior on the residual variance, σ^2. We use a standard vague prior:

$$\sigma^2 \sim Inv - Gamma(0.01, 0.01)$$

For many traits, there will be prior information available on the heritability, and therefore on the residual variance of a whole genome predictor. However, as we explain below, in the first step of our parallelized approach, the algorithm is only applied to a small number of SNPs simultaneously. Hence the use of the vague priors above.

2.1.3 | Binary traits

So far, our framework has assumed the trait of interest is continuous. This is because we rely on a linear modelling framework to relate univariate summaries to multivariate effects via the linear transformation $X'y$. In the case of binary traits, we derive $X'y$ after first mapping univariate log-odds ratios to approximate linear effects via their z-scores. That is, we infer the univariate effects that would have been estimated if the binary outcome has been modelled by linear regression, allowing construction of $X'y$. This strategy is employed in other linear-based summary data frameworks (Chen et al., 2015) including LDPred (Vilhjálmsson et al., 2015), to which we refer readers for a detailed description of this mapping.

2.1.4 | Inference of a posterior model averaged polygenic risk score via reversible jump Markov Chain Monte Carlo

We cannot derive analytical expressions for the posterior of β and so use reversible jump MCMC (Green, 1995) to sample from the required posterior distribution. The reversible jump sampling scheme starts with an initial model, $\gamma(0)$, which is a selection of variants in Step 1 and corresponding parameter values, $\theta(0)$. To sample the next model and set of parameters, which we denote by $\gamma(1)$ and $\theta(1)$, we propose moving from the current state to another model and/or parameter values, γ^* and θ^*, using a proposal function q $(\gamma^*, \theta^* | \gamma, \theta)$. The proposed model and parameters are accepted with probability equal to the Metropolis-Hastings ratio:

$$MHR = \frac{P(D| \gamma^*, \theta^*)P(\theta^*| \gamma^*)P(\gamma^*)}{P(D| \gamma, \theta)P(\theta | \gamma)P(\gamma)} \times \frac{q(\gamma, \theta | \gamma^*, \theta^*)}{q(\gamma^*, \theta^*| \gamma, \theta)}$$

where D is the observed data, and $P(D|\gamma, \theta)$ is the multivariate likelihood described by Equation (3). $P(\gamma^*)$ is the β-binomial model space prior defined in Equation (4) and $P(\theta^*|\gamma^*)$ is the prior on the parameters conditional on (i.e. included in) the model. The proposed model and parameter values are therefore accepted with a probability proportional to both their likelihood and prior support. If this new set of values is accepted, we set $\gamma(1) = \gamma^*$ and $\theta(1) = \theta^*$, otherwise they are discarded and the current values are retained; $\gamma(1) = \gamma(0)$ and $\theta(1) = \theta(0)$. It can be shown that this produces a sequence of parameter/model samples, which converge to the target posterior distribution (Green, 1995). After obtaining a posterior sample of effects, β, for each variant (note that many of these values may be zero corresponding to exclusion from the model at a particular iteration), we average to obtain the final weighted polygenic risk score across all variants, $\hat{\beta}$. This may also be interpreted as the vector

of posterior mean variant effects, averaged over the posterior distribution of models.

2.1.5 | Two-step approach to the adjustment of long-range linkage disequilibrium

A practical limitation to the use of the method outlined above is that a full rank genotype matrix is required to construct the plug-in estimate for $X'X$. This is due to the necessary inversion during inference (Newcombe et al., 2016). Therefore, the number of variants that can be simultaneously modelled must necessarily be less than the number of individuals in the reference data, and, in practice, depending on the amount of correlation, considerably less. In fine-mapping applications, this is generally not a problem. In the genome-wide context, Mak et al. (2017) regularized $X'X$ with a further penalty parameter, transforming their original Lasso model to an elastic net problem. LDpred uses a Gibbs sampler that essentially models a sliding window of variants, of fixed size. Here we suggest the following two-step approach to build models within blocks, under the Bayesian sparse regression outlined above, and then account for cross-block correlation to arrive at a genome-wide model.

Step 1: Block-specific polygenic scores
In step one, the P variants genome-wide are split into B small blocks of 100 variants each, and JAM is used to derive posterior mean weights for the variants within each block: $\hat{\boldsymbol{\beta}}_b$ for $b = 1, ..B$. For the analysis of each block b, the input data is the set of marginal variant effects as well as the columns of the reference genotype matrix corresponding to the variants in block b. The block size choice of 100 could be varied but we found that 100 worked well in practice (see Section 3).

Step 2: Between-block adjustment
In the absence of correlations across blocks, an unbiased genome-wide polygenic risk score, with weights which we denote by $\hat{\boldsymbol{\beta}}_G$, could then be constructed by simply appending the block-specific scores:

$$\hat{\boldsymbol{\beta}}_G = (\hat{\boldsymbol{\beta}}_1, .., \hat{\boldsymbol{\beta}}_B)$$

However, this risk score will be biased in the presence of correlations *across* blocks, since they were ignored in Step 1. To account for *cross-block* correlations, we introduce a second layer of multivariate weights, $\boldsymbol{\delta} = (\delta_1, .. \delta_B)$, which adjust the "marginal" block-specific

scores for one another. Specifically, we seek to estimate the "block-adjusted" risk score:

$$\hat{\boldsymbol{\beta}}_G = (\delta_1 \hat{\boldsymbol{\beta}}_1, .. \delta_B \hat{\boldsymbol{\beta}}_B)$$

It is instructive to consider the block-specific scores as a set of B covariates, and view $\boldsymbol{\delta}$ as the multivariate vector of effects, we would obtain from a regression of $\boldsymbol{y}$ on an $n \times B$ matrix of the block-specific scores, $\boldsymbol{S}$. For clarity, the element of $\boldsymbol{S}$ corresponding to individual i and block b is:

$$s_{i,b} = \sum_{p=1}^{100} x_{i,b,p} \hat{\beta}_{b,p} \quad (5)$$

where $x_{i,b,p}$ is their genotype at variant p in block b, and $\hat{\beta}_{b,p}$ is the corresponding weight from the block-specific score as estimated in Step 1. It transpires that the estimation of $\boldsymbol{\delta}$ is straightforward, by reapplying the same methodology used to estimate multivariate SNP weights for each block in Step 1, except now we wish to adjust the "marginal" block-scores for the block-block correlation structure $\boldsymbol{S}'\boldsymbol{S}$. The analogy of Equation (2) is:

$$\boldsymbol{S}'\boldsymbol{y} \sim MVN(\boldsymbol{S}'\boldsymbol{S}\boldsymbol{\delta}, \boldsymbol{S}'\boldsymbol{S}\sigma^2) \quad (6)$$

By applying Equation (5) to the reference matrix $\boldsymbol{X}$, we also obtain a plug-in estimate for $\boldsymbol{S}'\boldsymbol{S}$. In the same way $\boldsymbol{X}'\boldsymbol{y}$ for Equation (2) is constructed from the marginal variant effects (and their minor allele frequencies) we can construct $\boldsymbol{S}'\boldsymbol{y}$ from the marginal block-score "effects" and the column means of $\boldsymbol{S}$. By construction, the "marginal" effect of each block-specific score is 1; a unit increase in each risk score is associated with the same unit increase in the trait $\boldsymbol{y}$. Intuitively, all scores have equivalent unit effects on $\boldsymbol{y}$ because the variant-specific effects from which they are composed are on the same scale as per Equation (1), which defines the unit SNP effects within each block. If a block contains variants with small effects, a large number is required for a score of 1. Conversely, for a block containing larger effects, fewer variants are required to achieve the same score. We obtain a maximum likelihood estimate for $\hat{\boldsymbol{\delta}}$ after multiplying both sides of Equation (6) through by the inverse Cholesky decompositon of correlation matrix $\boldsymbol{S}'\boldsymbol{S}$ to obtain a linear regression analogous to Equation (3). Note that model selection is not carried out in Step 2 since sparsity among the genetic effects has already been imposed in Step 1 under the prior in Equation (4), and we do not expect sparsity at the block level.

Obtaining $\hat{\delta}$ completes our genome-wide polygenic score in which correlations are accounted for both within and across regions:

$$\hat{\boldsymbol{\beta}}_G = (\hat{\delta}_1 \hat{\boldsymbol{\beta}}_1, \, .. \hat{\delta}_B \hat{\boldsymbol{\beta}}_B)$$

Note that the first step of our algorithm is why it is highly parallelizable. Although, in principle, joint analysis of multiple blocks would help inform estimation of the residual variance, σ^2, as well as the proportion of "causal" variants, θ, in practice we found no difference compared to estimating each block-specific score, $\hat{\boldsymbol{\beta}}_b$, independently. By running the block-specific analyses independently, we could take advantage of large numbers of CPU cores available to us via a high performance computing cluster (HPC). In each of the following real data applications we ran our algorithm for 200,000 iterations in the Step 1 block-specific analyses.

3 | RESULTS

3.1 | Cross-validation in the Welcome Trust Case Control Consortium

We first compared performance of our proposed method against LassoSum and LDPred using individual level genotypes, measured using the 500K Affymetrix Chip, for seven traits in the Welcome Trust Case Control Consortium (WTCCC, 2007). In total, data were available for 2,835 common controls, 1,827 bipolar disorder, 1,880 coronary artery disease (CAD), 1,684 Crohn's disease, 1,904 hypertension, 1,834 rheumatoid arthritis, 1,933 type 1 diabetes (T1D), and 1,872 type 2 diabetes cases, respectively. For each trait, the WTCCC cases and controls were randomly partitioned such that 2/3rd of the samples were

used to train all models and 1/3rd of the samples were used for testing. The random partitioning was conducted in a stratified manner, such that each of the three folds had the same proportion of cases and controls. After pruning variants with missing rates above 1%, and for maximum LD below r^2 95% using the Plink software package (Purcell et al., 2007), we were left with between 255,781 and 256,925 variants for each trait. LDPred and lassosum were run with default parameters as described in their papers. Results are presented in Table 1. Out-of-sample predictive results were pooled across all three folds before calculating the final performance summaries. 95% confidence intervals (95% CI) for the receiver operating characteristic area under the curves (ROC AUCs) were calculated using 2,000 stratified bootstrap replicates and the pROC R package. We also checked results using a different cross-validation partitioning seed, which were indistinguishable (not shown).

For most traits, the performance of LDPred and LassoSum was similar, although LDPred offered improvements for Crohn's disease, AUC of 0.69 (95% CI: 0.66–0.72) versus 0.65 (95% CI: 0.62–0.68), and T1D, AUC of 0.87 (95% CI: 0.85–0.89) versus 0.83 (0.80–0.85). Our two-step JAM method appeared the most robust, generally resulting in performance on par with the best performing of LDPred and lassosum. Runtimes for our two-step approach ranged between 3 and 4 min when running different prior settings and chromosomes in parallel on different computing nodes, with block parallelization across the 16 cores of each compute node. These runtimes were similar, though slightly faster than lassosum, which typically took an extra minute to run. Conversely, LDPred took several hours. Therefore, with a large numbers of compute cores available, our parallelizable approach was the fastest, while offering typically better predictive performance. However, in terms of total computational cost, LassoSum is the most efficient

TABLE 1 Application of three summary statistics prediction methods in the Welcome Trust Case Control Consortium under three-fold cross-validation

	LassoSum		LDPred		JAM	
Trait	AUC	r^2	AUC	r^2	AUC	r^2
Bipolar disorder	0.67 (0.64, 0.69)	0.09	0.66 (0.63, 0.69)	0.08	0.70 (0.67, 0.73)	0.13
Coronary artery disease	0.59 (0.56, 0.62)	0.02	0.59 (0.56, 0.62)	0.02	0.65 (0.62, 0.67)	0.08
Crohn's disease	0.65 (0.62, 0.68)	0.07	0.69 (0.66, 0.72)	0.10	0.69 (0.66, 0.72)	0.12
Hypertension	0.61 (0.58, 0.64)	0.04	0.59 (0.56, 0.62)	0.03	0.58 (0.55, 0.61)	0.02
Rheumatoid arthritis	0.71 (0.68, 0.73)	0.12	0.72 (0.69, 0.75)	0.14	0.74 (0.71, 0.76)	0.16
Type 1 diabetes	0.83 (0.80, 0.85)	0.30	0.87 (0.85, 0.89)	0.39	0.86 (0.84, 0.88)	0.36
Type 2 diabetes	0.62 (0.60, 0.65)	0.04	0.60 (0.57, 0.63)	0.05	0.64 (0.61, 0.67)	0.08

Note: ROC AUCs and predictive r^2 are presented, with ROC AUC 95% confidence intervals calculated via 2,000 stratified bootstrap samples. For each method, performance is presented for the best performing sparsity.
Abbreviations: AUC, area under the curve; ROC, receiver operating characteristic.

method, achieving these runtimes when only a single compute node is available.

3.2 | Meta-GWAS applications for type 1 diabetes, coronary artery disease, and schizophrenia

Next, we exemplify our method in three case studies using summary statistics from meta-GWAS for T1D (T1DGC; 3,983 cases, 3,999 controls), CAD (cardiogram; 40,170 cases, 97,365 controls), and schizophrenia (PGC; 34,241 cases, 45,604 controls). Summary statistics from the T1DGC study were available for both genotyped SNPs (Illumina 550K) as well as imputed SNPs unique to the older Affymetrix 500K array. The imputation, originally conducted to facilitate meta-analysis with the WTCCC, was based on a substantial subset of controls for which both arrays were available. Details of the original QC, imputation method, and meta-analysis can be found in Barrett et al. (2010). The cardiogram summary statistics came from a meta-analysis after excluding the WTCCC and restricting to the 37 remaining studies with white European ancestry. Studies contributed either genotypes from the Metabochip array or GWAS data imputed using HapMap; further details on QC, imputation, and the meta-analysis may be found in Nikpay et al. (2015). The schizophrenia summary statistics correspond to a meta-analysis of 46 European studies from the Psychiatric Genomics Consortium (PGC), each of which contributed both genotyped SNPs (from various arrays) as well as imputed SNPs using a 1,000 genome reference panel. Further details of the QC, imputation, and meta-analysis may be found in Ripke et al. (2014). Case and control samples from the WTCCC were used as independent testing data for CAD ($n = 4,715$) and T1D ($n = 3,308$). For the latter, the 1958 birth cohort was removed since these samples were used as controls in the T1DGC. For schizophrenia, testing data comprised genotypes measured using the Affymetrix 6.0 array in 5,334 samples from the Molecular Genetics of Schizophrenia (MGS) study (Shi et al., 2009). Since the MGS results were included in the PGC meta-GWAS, we excluded their influence using a technique detailed by Mak et al. (2017), whereby the hypothetical meta-analysis of the PGC excluding MGS is inferred according to the results from each. In each case study, the SNPs used for polygenic model building were the subset that appeared in the intersection of SNPs available in both the summary statistics and testing datasets, and remained after pruning for less than 1% missingness and LD less than r^2 95% in the corresponding testing data. This left 231,510 SNPs for T1D, 211,263 for CAD, and 385,474 for schizophrenia.

All three sparse regression methods—JAM, LDPred, and lassosum—were run with the same parameters as described in the WTCCC cross-validation analyses above. We also compared against a simple p-value thresholding approach, as well as a combined clumping and p-value thresholding approach, which selectively removes less significant SNPs to reduce LD (Wray et al., 2014). For the p-value thresholding, we used the set of p-values {5e-8, 1e-5, 1e-4, 1e-3, 0.0015, 0.0025,...0.995}, and when combined with clumping, we used r^2 thresholds of 0.2, 0.5, and 0.8.

For T1D, all sparse regression methods surpassed simpler p-value thresholding, with JAM offering the best predictive performance; r^2 of 0.38 compared to 0.35 for LDPred and 0.30 for lassosum, AUC of 0.86 (95% CI: 0.85–0.88) compared to 0.85 (95% CI: 0.83–0.86) for LDPred, and 0.82 (95% CI: 0.81–0.84) for Lassosum. We suspect the improved performance from JAM is due to the large number of correlated signals in regions of both short- and long-range LD within the major histocompatibility complex (MHC) on chromosome 6, for which a more sophisticated model search and averaging algorithm should offer greater accuracy. To confirm, we reran JAM, LDPred, and lassosum after excluding chromosome 6 from the training data. As expected, performance was considerably diminished for all three methods but was indistinguishable between JAM, LDPred, and lassosum (r^2 of 0.08 and AUC of 0.66 for all), indicating that JAM's improved performance for T1D is indeed driven by a more flexible model for the MHC. For schizophrenia and CAD, where the polygenic signal is weaker and more dispersed through the genome, all methods performed similarly (Figure 1).

A unique feature of our method is the ability to adapt sparsity within local SNP blocks. To demonstrate how this looked in practice, we plotted the posterior mean selected number of variants for each block from each case study (Figure 2). As expected for T1D, a number of blocks within the MHC on chromosome 6 had considerably more SNPs selected, indicating that JAM adapted to impose less sparsity, in a region containing strong signals. For CAD and schizophrenia, the spread of block-specific selections was more similar across chromosomes, but there was still variation, demonstrating block-to-block adaption. Note that the larger number of total SNPs selected for CAD and schizophrenia was due to much larger training samples, which allowed the estimation of many more smaller effects—see Table 2. The pattern of runtimes across all three case studies was similar to the WTCCC cross-validation study above, with JAM offering the best runtimes when different priors and chromosomes were run in parallel, lassosum not far behind, and LDPred significantly slower (Table 2).
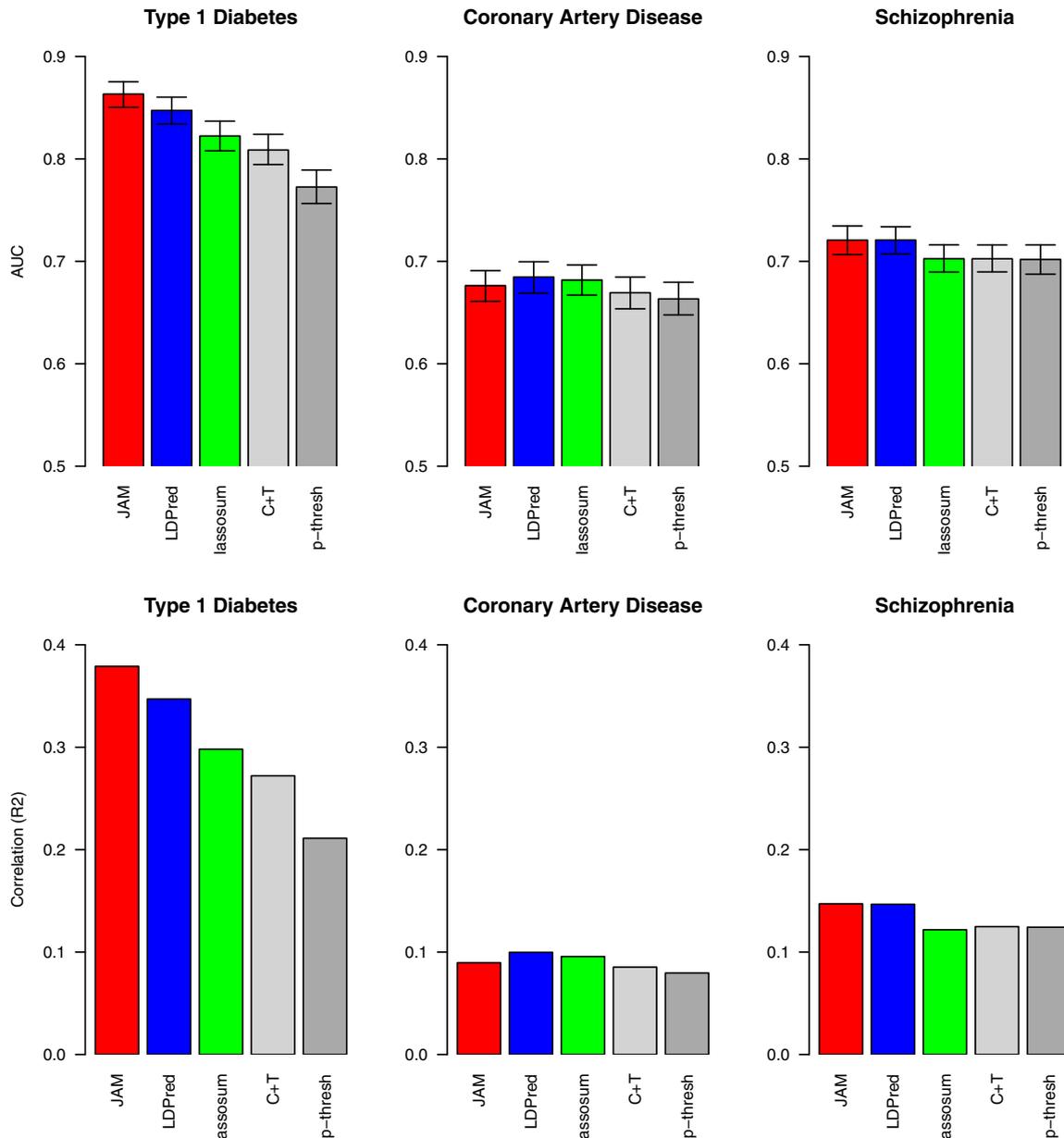
FIGURE 1 Receiver operating characteristic area under the curves ROC AUCs and predictive r^2 for various predictive methods in three case studies training polygenic predictive models using meta-GWAS summaries. For type 1 diabetes, the T1DGC ($n = 8,005$) was used for training and the Welcome Trust Case Control Consortium for validation. For coronary artery disease (CAD), cardiogram ($n = 137,535$) was used for training and the WTCCC for testing. For schizophrenia, the Psychiatric Genomics Consortium ($n = 74,511$) was used for training and the MGS study for testing. For each analysis and method, results are presented for the best performing sparsity. ROC AUC 95% confidence intervals were calculated using 2,000 stratified bootstrap replicates

4 │ DISCUSSION

We present a novel, flexible, and parallelizable approach to the construction of genome-wide polygenic risk scores from meta-GWAS summary statistics. In cross-validation analyses of all seven traits in the WTCCC, and three case studies using meta-GWAS summary results, our method offered similar predictive performance to two alternative approaches, lassosum and LDPred, while achieving the fastest runtimes.

The improved runtimes were due to the highly parallelizable nature of our algorithm, whereby genetic correlations are adjusted for in two steps, allowing the analysis of many small groups of SNPs independently at Step 1, in which we only account for short-range LD. Because we partition the genome into blocks rather arbitrarily at Step 1, we allow for LD straddling block boundaries, as well as longer range LD, in Step 2. The adjustment for both short- and long-range LD is novel to our method, and is more satisfactory than
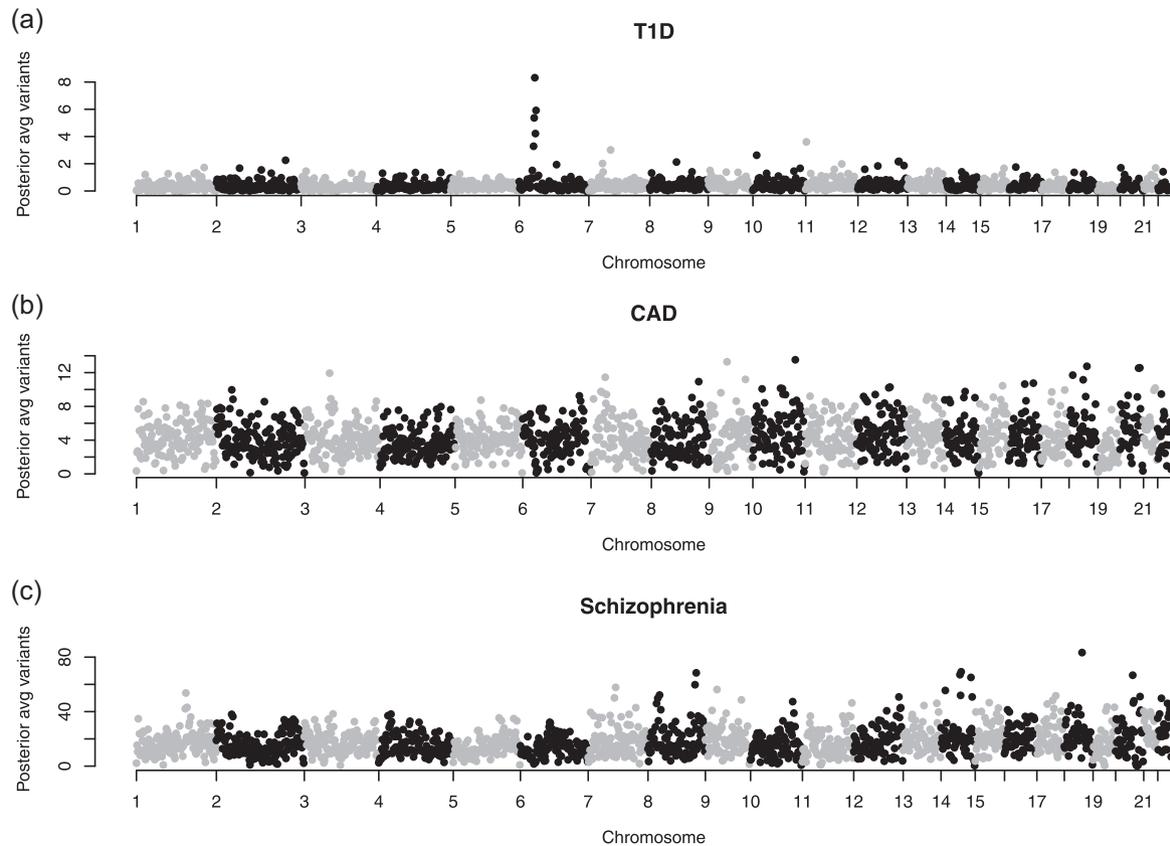
FIGURE 2 Block-specific posterior mean numbers of selected single nucleotide polymorphisms by the JAM method in each of the three meta-GWAS case studies. The vertical spread indicates the variation in block-specific adapted sparsities from step one of our proposed framework. Note that the global average varies across the case studies owing to different optimal λs, which control the global sparsity. For CAD and schizophrenia, summary statistics were available from considerably larger training datasets, allowing the estimation of many more small effects (see Table 2). CAD, coronary artery disease

simple block-wise model fitting since LD decays stochastically and any attempt to partition the genome will result in some correlations across blocks. The most important instance of long-range LD in the genome is in the human leukocyte antigen (HLA) complex, and it is those diseases with HLA associations where our method, and others that account for LD, did best. There is a particularly strong and complex HLA signal for T1D, consisting of multiple heterogeneous signals in regions of both short- and long-range LD, and it was here, in both the WTCCC and T1DGC

case studies, that performance gains were greatest compared to LDPred and lassosum, which only model short-range LD. A further novel feature of our method, compared to LDPred and lassosum, is that we allow for different genetic models, that is, signal sparsities, across different blocks. This is achieved by treating sparsity and local heritability as random quantities rather than fixed hyperparameters, which should offer increased statistical robustness over repeated replication studies, as well as inference on where highly polygenic predictive signals are concentrated (see Figure 2).

TABLE 2 Computational aspects of JAM and runtimes (in minutes) of the different methods applied to the meta-GWAS case studies

Case study	Total SNPs	JAM λ	JAM SNPs	JAM Runtime	lassosum Runtime	LDPred Runtime
T1D	231,510	0.01	712	4.6	10.3	61.5
CAD	211,263	0.001	7233	5.1	7.5	27.5
Schizophrenia	385,474	1E-04	30544	16.6	17.4	157.2

Note: The total number of SNPs analyzed (i.e., after QC), for all methods, is shown in the first column. The next two columns correspond to the optimal value of for use with JAM—smaller values encourage more sparsity—and the posterior average number of SNPs selected into the corresponding optimal JAM model
Abbreviations: CAD, coronary artery disease; SNP, single-nucleotide polymorphism; T1D, type 1 diabetes.

Although we demonstrate the two-step approach in a Bayesian model averaging framework, the idea is generic to regression and could be leveraged for gains in other frameworks too. There are, however, some conceptual advantages to using a formal Bayesian model averaging approach. First, we do not have to fix an assumed proportion of causal variants, but instead treat this as an unknown parameter with a prior that is integrated over separately for each block. Consequently, we observed that predictive performance under different priors was more robust than, analogously, LDPred and lassosum across the range of their respective sparsity tuning parameters. Although the standard practice is to choose the sparsity tuning parameter for each method according to best performance in the test data, as we do here, this will, in principle, lead to a degree of optimism, that is, overfitting. Ideally, an independent data set should be used to select tuning parameters such that the predictive algorithm is entirely finalized before application in the test data. This was not practical in the data sets we considered here, but the observation that our fully Bayesian approach is more robust to the sparsity choice provides confidence that results from the case studies are less likely to suffer from optimism. A further appeal of our approach is that it provides a unified analytical model for fine mapping and polygenic score estimation. In the fine mapping context, the genetic effects can be integrated out as nuisance parameters (Newcombe et al., 2016), with the objects of inferences being the probabilities of including each SNP in the model. Here, by contrast, but under the same framework, the SNP effects are obtained by averaging over the inclusion probabilities.

Further flexibility is possible through the prior on effect sizes for selected SNPs. We have used a conjugate Gaussian prior, which after averaging over sparse models leads to a marginal posterior distribution similar to that obtained from a slab-and-spike prior. Thus our approach is conceptually similar to LDpred at the block level, but we allow for different posterior distributions across blocks, related by a common vague prior distribution. Now our approach can readily be extended by letting each SNP belong to one of several classes, with class membership probabilities drawn from a multinomial distribution and distinct Gaussian distributions for the effects in each class. After model averaging, this would be analogous to some mixture models recently proposed for individual level data (Moser et al., 2015; Zhou & Stephens, 2014), with the advantage that we can use summary statistics and fit distinct models to blocks of

SNPs in parallel. Thus, while the present work provides a proof of principle with comparable accuracy to competing methods for summary statistics, it is readily extensible in ways that mirror the best current models for individual level data. Furthermore, a mixture modelling formulation may provide a natural way to accommodate external SNP-specific functional annotation information from public resources, such as annotation databases, expression, and methylation quantitative trait locus analyses. An extension is conceivable whereby SNP-specific functional annotations would influence class assignments within the mixture of Gaussian distributions. For example, rather than using a naïve multinomial distribution, an informative β prior could be constructed across the class assignment ratios. Indeed, it has previously been shown that reflection of prior annotation when constructing polygenic risk scores is advantageous (Shi et al., 2016). We plan to pursue all these ideas in future work.

Assuming availability of a large number of computing nodes, our method offered the fastest performance with similar predictive performance to lassosum and LDPred. The level of computing resources required for these runtimes (~100 compute nodes) is available to many researchers today, however, our parallelization approach can take advantage of even more resources as they inevitably become available over the coming years, since many of the individual block analyses are still being run sequentially in Step 1. With sufficient computing resources, in principle, every single 100 SNP block could be run be run in parallel, which would reduce runtimes to well under a minute in our three case studies. Relatedly, while we found the use of 100 SNP blocks in Step 1 worked well in practice, this choice is likely to be dependent on marker density, which will determine the average genomic length of these blocks. Using microarray data, as in our case studies, the 100 SNP blocks span genomic lengths the order of 100 kb for microarray data, but, when, for example, using imputed sequence data larger blocks may lead to better performance. We intend to explore this in more detail in future work.

We have incorporated our algorithm "JAMPred" into our existing fully documented R package for Bayesian model selection "R2BGLiMS", which also contains the original "JAM" software, and is freely available to download via github https://github.com/pjnewcombe/R2BGLiMS. We have included scripts demonstrating the JAMPred syntax, and how to distribute a genome-wide analysis on an HPC, in the Supporting Information.

ACKNOWLEDGMENTS

DATA AVAILABILITY

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Paul J. Newcombe http://orcid.org/0000-0002-5611-6702
Frank Dudbridge http://orcid.org/0000-0002-8817-8908

REFERENCES

Barrett, J. C., Clayton, D. G., Concannon, P., Akolkar, B., Cooper, J. D., Erlich, H. A., ... Rich, S. S. (2010). Genome-wide association study and meta-analysis finds over 40 loci affect risk of type 1 diabetes. *Nature Genetics*, 41(6), 703–707.

Bottolo, L., Chadeau-Hyam, M., Hastie, D. I., Zeller, T., Liquet, B., Newcombe, P., ... Richardson, S. (2013). GUESS-ing polygenic associations with multiple phenotypes using a GPU-based evolutionary stochastic search algorithm. *PLoS Genetics*, 9(8), e1003657. Edited by G. Gibson.

Chatterjee, N., Wheeler, B., Sampson, J., Hartge, P., Chanock, S. J., & Park, J. H. (2013). Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. *Nature Genetics*, 45(4), 400–405. Nature Publishing Group.

Chen, W., Larrabee, B. R., Ovsyannikova, I. G., Kennedy, R. B., Haralambieva, I. H., Poland, G. A., & Schaid, D. J. (2015). Fine mapping causal variants with an approximate Bayesian method using marginal test statistics. *Genetics*, 200(3), 719–736.

Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. *PLoS Genetics*, 9(3), e1003348.

Dudbridge, F. (2016). Polygenic epidemiology. *Genetic Epidemiology*, 40(4), 268–272.

Dudbridge, F., & Newcombe, P. J. (2015). Accuracy of gene scores when pruning markers by linkage disequilibrium. *Human Heredity*, 80(4), 178–186.

Eriksson, J., Evans, D. S., Nielson, C. M., Shen, J., Srikanth, P., Hochberg, M., ... Ohlsson, C. (2015). Limited clinical utility of a genetic risk score for the prediction of fracture risk in elderly subjects. *Journal of Bone and Mineral Research*, 30(1), 184–194.

Evans, D. M., Visscher, P. M., & Wray, N. R. (2009). Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk. *Human Molecular Genetics*, 18(18), 3525–3531.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4), 711–732.

Griffin, J. E., & Brown, P. J. (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis*, 5(1), 171–188.

Lango Allen, H., Estrada, K., Lettre, G., Berndt, S. I., Weedon, M. N., Rivadeneira, F., ... Hirschhorn, J. N. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*, 467(7317), 832–838.

Mak, T. S. H., Porsch, R. M., Choi, S. W., Zhou, X., & Sham, P. C. (2017). Polygenic scores via penalized regression on summary statistics. *Genetic Epidemiology*, 1–12.

Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157(4), 1819–1829.

Morris, A. P., Voight, B. F., Teslovich, T. M., Ferreira, T., Segrè, AV, Steinthorsdottir, V., ...McCarthy, M. I. (2012). Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nature Genetics*, 44(9), 981–990. Nature Publishing Group.

Moser, G., Lee, S. H., Hayes, B. J., Goddard, M. E., Wray, N. R., & Visscher, P. M. (2015). Simultaneous discovery, estimation and prediction analysis of complex traits using a Bayesian mixture model. *PLoS Genetics*, 11(4), 1–22.

Newcombe, P., Raza Ali, H., Blows, F., Provenzano, E., Pharoah, P., Caldas, C., & Richardson, S. (2014). Weibull regression with Bayesian variable selection to identify prognostic tumour markers of breast cancer survival. *Statistical Methods in Medical Research*, 26, 414–436. https://doi.org/10.1177/0962280214548748p

Newcombe, P. J., Conti, D. V., & Richardson, S. (2016). JAM: A scalable Bayesian framework for joint analysis of marginal SNP effects. *Genetic Epidemiology*, 40(3), 188–201.

Nikpay, M., Goel, A., Won, H. H., Hall, L. M., Willenborg, C., Kanoni, S., ... Farrall, M. (2015). A comprehensive 1000 genomes-based genome-wide association meta-analysis of coronary artery disease. *Nature Genetics*, 47(10), 1121–1130.

Park, T., & Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association*, 103(482), 681–686.

Peng, J., Zhu, J., & Bergamaschi, A. (2010). Regularized multivariate regression for identifying master predictors with application to integrative genomics study of breast cancer. *The Annals of Applied Statistics*, 4(1), 53–77.

Pharoah, P. D. P., Antoniou, A., Bobrow, M., Zimmern, R. L., Easton, D. F., & Ponder, B. A. J. (2002). Polygenic susceptibility

to breast cancer and implications for prevention. *Nature Genetics*, 31(1), 33–36.

Pharoah, P. D. P., Antoniou, A. C., Easton, D. F., & Ponder, B. A. J. (2008). Polygenes, risk prediction, and targeted prevention of breast cancer. *The New England Journal of Medicine*, 358(26), 2796–2803.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., ... Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81(3), 559–575.

Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., Sullivan, P. F., ... Moran, J. L. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, 10, 8192–8192.

Ripke, S., Neale, B. M., Corvin, A., Walters, J. T., Farh, K. H., Holmans, P. A., ... O'Donovan, M. C. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, 511(7510), 421–427.

Servin, B., & Stephens, M. (2007). Imputation-based analysis of association studies: Candidate regions and quantitative traits. *PLoS Genetics*, 3(7), e114. Public Library of Science.

Shi, J., Park, J. H., Duan, J., Berndt, S. T., Moy, W., Yu, K., ... Chatterjee, N. (2016). Winner's curse correction and variable thresholding improve performance of polygenic risk modeling based on genome-wide association study summary-level data. *PLoS Genetics*, 12(12), 1–24.

Shi, J., Levinson, D. F., Duan, J., Sanders, A. R., Zheng, Y., Pe'er, I., ... Gejman, P. V. (2009). Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature*, 460(7256), 753–757. Nature Publishing Group.

Stahl, E. A., Wegmann, D., Trynka, G., Gutierrez-Achury, J., Do, R., Voight, B. F., ... Plenge, R. M. (2012). Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nature Genetics*, 44(5), 483–489.

Tachmazidou, I., Johnson, M. R., & De Iorio, M. (2010). Bayesian variable selection for survival regression in genetics. *Genetic Epidemiology*, 34(7), 689–701.

Talmud, P. J., Cooper, J. A., Morris, R. W., Dudbridge, F., Shah, T., Engmann, J., ... Humphries, S. E. (2015). Sixty-five common genetic variants and prediction of type 2 diabetes. *Diabetes*, 64(5), 1830–1840.

Teslovich, T. M., Musunuru, K., Smith, A. V., Edmondson, A. C., Stylianou, I. M., Koseki, M., ... Kathiresan, S. (2010). Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*, 466(7307), 707–713. Nature Publishing Group.

The International Multiple Sclerosis Genetics Consortium (IMSGC) (2010). Evidence for polygenic susceptibility to multiple sclerosis-The shape of things to come. *American Journal of Human Genetics*, 86(4), 621–625.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B*, 58(1), 267–288.

Verzilli, C., Shah, T., Casas, J. P., Chapman, J., Sandhu, M., Debenham, S. L., ... Hingorani, A. D. (2008). Bayesian meta-analysis of genetic association studies with different sets of markers. *American Journal of Human Genetics*, 82(4), 859–872.

Vignal, C. M., Bansal, A. T., & Balding, D. J. (2011). Using penalised logistic regression to fine map HLA variants for rheumatoid arthritis. *Annals of Human Genetics*, 75(6), 655–664.

Vilhjálmsson, B. J., Yang, J., Finucane, H. K., Gusev, A., Lindström, S., Ripke, S., ... Zheng, W. (2015). Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *American Journal of Human Genetics*, 97(4), 576–592.

De Vries, P. S., Kavousi, M., Ligthart, S., Uitterlinden, A. G., Hofman, A., Franco, O. H., & Dehghan, A. (2015). Incremental predictive value of 152 single nucleotide polymorphisms in the 10-year risk prediction of incident coronary heart disease: The Rotterdam Study. *International Journal of Epidemiology*, 44(2), 682–688.

Wacholder, S., Hartge, P., Prentice, R., Garcia-Closas, M., Feigelson, H. S., Diver, W. R., ... Hunter, D. J. (2010). Performance of common genetic variants in breast-cancer risk models. *The New England Journal of Medicine*, 362(11), 986–993.

Wallace, C., Cutler, A. J., Pontikos, N., Pekalski, M. L., Burren, O. S., Cooper, J. D., ... Wicker, L. S. (2015). Dissection of a complex disease susceptibility region using a Bayesian stochastic search approach to fine mapping. *PLoS Genetics*, 11(6), e1005272.

Wilson, M. A., Iversen, E. S., Clyde, M. A., Schmidler, S. C., & Schildkraut, J. M. (2010). Bayesian model search and multilevel inference for SNP association studies. *Annals of Applied Statistics*, 4(3), 1342–1364.

Wray, N. R., Lee, S. H., Mehta, D., Vinkhuyzen, A. A. E., Dudbridge, F., & Middeldorp, C. M. (2014). Research review: Polygenic methods and their application to psychiatric traits. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 55(10), 1068–1087.

Wray, N. R., Yang, J., Hayes, B. J., Price, A. L., Goddard, M. E., & Visscher, P. M. (2013). Pitfalls of predicting complex traits from SNPs. *Nature Reviews Genetics*, 14(7), 507–515.

WTCCC (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447(7145), 661–678.

Wu, T. T., Chen, Y. F., Hastie, T., Sobel, E., & Lange, K. (2009). Genome-wide association analysis by lasso penalized logistic regression. *Bioinformatics*, 25(6), 714–721.

Yang, J., Ferreira, T., Morris, A. P., Medland, S. E., Madden, P. A. F., Heath, A. C., ... Visscher, P. M. (2012). Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nature Genetics*, 44(4), 369–375. Nature Publishing Group.

Zhou, X., Carbonetto, P., & Stephens, M. (2013). Polygenic modeling with Bayesian sparse linear mixed models. *PLoS Genetics*, 9(2), e1003264.

Zhou, X., & Stephens, M. (2014). Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nature Methods*, 11(4), 407–409.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.