

Bayesian meta-analysis across genome-wide association studies of diverse phenotypes

Short title: “MetABF: Bayesian meta-analysis for GWAS”

Holly Trochet^{*1,2}, Matti Pirinen^{*3,4,5}, Gavin Band¹, Luke Jostins^{1,7,8,9},
Wellcome Trust Case-Control Consortium 2⁶, Gilean McVean^{1,7}, and Chris
C. A. Spencer¹

¹Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK

²Institut de Cardiologie de Montréal (Centre de Recherche), Université de Montréal, Montréal, Québec, Canada

³Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland

⁴Helsinki Institute for Information Technology HIIT and Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland

⁵Department of Public Health, University of Helsinki, Helsinki, Finland

⁶www.wtccc.org.uk

⁷Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Roosevelt Drive, Oxford, UK

⁸Kennedy Institute of Rheumatology, University of Oxford, Roosevelt Drive, Oxford, UK

⁹Christ Church, University of Oxford, St Aldates, Oxford, UK

^{*}Joint first author.

Correspondence can be sent to holly.trochet@umontreal.ca.

March 11, 2019

Abstract

Genome-wide association studies (GWAS) are a powerful tool for understanding the genetic basis of diseases and traits, but most studies have been conducted in isolation, with a focus on either a single or a set of closely related phenotypes. We describe MetABF, a simple Bayesian framework for performing integrative meta-analysis across multiple GWAS using summary statistics. The approach is applicable across a wide range of study designs and can increase the power by 50% compared to standard frequentist tests when only a subset of studies have a true effect. We demonstrate its utility in a meta-analysis of 20 diverse GWAS which were part of the Wellcome Trust Case-Control Consortium 2. The novelty of the approach is its ability to explore, and assess the evidence for, a range of possible true patterns of association across studies in a computationally efficient framework.

Keywords: gwas, meta-analysis, pleiotropy, summary statistics

1 Introduction

In the past decade, a large number of genome-wide association studies (GWAS) have been performed, and they have identified thousands of associations between genotypes and various biological traits (Hindorff et al., n.d.). These associations expand our understanding of both unique and shared molecular mechanisms across different phenotypes (Price, Spencer, & Donnelly, 2015). With so much data available, there is now an increased interest in combining datasets in order to learn about regions of the genome that affect multiple traits and, consequently, the shared biological pathways that underlie them.

While the search for genetic associations shared between traits can be conducted using individual-level genotype data from the participating GWAS (Ellinghaus et al., 2016), in practice, most of the readily available data come in the form of summary statistics. While the exact summary statistics reported vary from study to study, they usually include information about the genetic variant (usually a single nucleotide polymorphism, SNP) and its estimated effect on the phenotype under study. Information about the variant usually includes its genomic location in a given build of the human genome, as well as its alleles, and sometimes the frequencies of the alleles. Information about the association often includes the effect size estimate, its standard error or a 95% confidence interval, and the p -value of the association between the variant and the phenotype. The summary statistic information typically comes in files that are orders of magnitude smaller than the original genotype data files, and reduce the risk of revealing personal information about study participants.

As a result of their availability, a number of methods have been introduced to combine GWAS summary statistics to find genetic loci that influence multiple traits across the genome (Bhattacharjee et al., 2012; Cotsapas et al., 2011; Flutre, Wen, Pritchard, & Stephens, 2013; Majumdar, Haldar, Bhattacharya, & Witte, 2018; Turley et al., 2018; Wen & Stephens, 2014a). These meta-analyses can also find novel associations that were overlooked in individual GWAS.

In this paper we introduce MetABF, an approach to meta-analyse GWAS summary

statistics at a single SNP which is simple, efficient, and which can be easily programmed using any standard statistical package—we offer an implementation in R. We have found similar methods to be useful in previous studies (Band et al., 2013; Bellenguez et al., 2012; Gilchrist et al., 2018; Rautanen et al., 2016) and present a unified framework for others to use to explore their data. Unlike most other methods available, ours allows for a direct probabilistic assessment of different models of association across studies, given appropriate prior assumptions. We demonstrate the broad applicability of the approach using simulations of association studies, and by applying it to a diverse set of 20 GWAS that were performed by the Wellcome Trust Case-Control Consortium 2 (WTCCC2, www.wtccc.org). The R implementation is available at <https://github.com/trochet/metabf>.

2 The Method

Traditionally, meta-analyses of GWAS have focused on combining results of multiple studies on the same or similar traits. Figure 1 illustrates the inference problem. Given a set of effect size estimates ($\hat{\beta}$), and the uncertainty in these estimates ($SE_{\hat{\beta}}$) what can we infer about the true effect sizes, β_i , underlying them?

Figure 1

Interpretation of the data depends on assumptions about the heterogeneity in effect between studies. For example, an assessment of the data might be that there is little evidence from studies 1, 2 and 3 because their confidence intervals overlap zero. Even if we were to assume that all studies estimate the same effect, the inconsistency in observed effect sizes, in particular for β_2 and β_4 , means that we are unlikely to find strong evidence for a non-zero effect. However, if we knew that the exact trait under examination in studies 1 and 2 was slightly different to those in 3 and 4 then the strong evidence that $\beta_4 < 0$ might suggest that $\beta_3 < 0$ as well, without necessarily suggesting the same for β_1 and β_2 . Our framework aims to make it easy to quantify the statistical evidence for these kinds of heterogeneous models

and to provide the corresponding effect size estimates.

2.1 Bayesian approach

Our method uses a Bayesian approach that provides a natural way of capturing the reasoning described above for Figure 1 in a statistical framework. To measure the evidence for association, we calculate Bayes factors, which consider the ratio of the posterior probabilities of two models M_1 (an alternative model) and M_0 (the null model) given some data. By Bayes' Theorem,

$$\underbrace{\frac{P(M_1|\text{data})}{P(M_0|\text{data})}}_{\text{Posterior odds}} = \underbrace{\frac{P(\text{data}|M_1)}{P(\text{data}|M_0)}}_{\text{Bayes factor}} \times \underbrace{\frac{P(M_1)}{P(M_0)}}_{\text{Prior odds}}. \quad (1)$$

A Bayes factor greater than 1 suggests that the evidence from the data favors M_1 over M_0 , while a Bayes factor smaller than 1 suggests the opposite. Note that Bayes factors lie in the range $(0, \infty)$ but when they are presented on a logarithmic scale, negative values are also possible and correspond to the Bayes factors favoring the null model. The interpretation of Bayes factors in the context of GWAS has been discussed earlier (Consortium, 2007; Stephens & Balding, 2009). Because Bayes factors calculate the probability of the data under the null as well as under the alternative model, they naturally account for power to detect effects. In GWAS, this means that Bayes factors can be easier to calibrate across varying study sizes and minor allele frequencies than p -values which only consider the tail probabilities under the null model (Wakefield, 2009). Bayes factors require a prior distribution on the model parameters which can be used to describe different alternative models.

In practice, Bayes factors can be computationally expensive to calculate as they commonly involve an integration over the model parameters with respect to the prior distribution. Often, there is no closed form solution of the necessary integrals, necessitating numerical procedures. In the context of GWAS, Jon Wakefield developed an approximate Bayes factor (ABF) that can be calculated directly from GWAS summary statistics (Wakefield, 2007,

2009). Asymptotically, it gives similar results to a Bayes factor calculated using a Laplace approximation, and the study sizes required for a good approximation are of the order of hundreds of participants (Wakefield, 2009). Here we describe our approach, using approximate Bayes factors (ABFs) in genome-wide association studies (Wakefield, 2007, 2009) which is similar to those applied elsewhere (for example, (Asimit et al., 2015; Pickrell et al., 2016; Wen & Stephens, 2014b)). We compare our MetABF method to two other methods CPBayes (Majumdar et al., 2018) and MTAG (Turley et al., 2018).

2.2 Statistical model

For each variant, a GWAS produces a point estimate, $\hat{\beta}$, of the effect of a given allele on a trait, typically adjusted for relevant covariates like age, sex, or ethnicity. In quantitative traits, $\hat{\beta}$ represents the direct effect on the phenotype measurement. In binary traits, $\hat{\beta}$ estimates the natural logarithm of the odds ratio (OR) corresponding to each additional copy of the effect allele, that is, $\exp(\hat{\beta})$ is the odds ratio estimate. Each $\hat{\beta}$ comes with a standard error, $SE_{\hat{\beta}}$, which is estimated along with $\hat{\beta}$ and, in most settings, is largely determined by the sample size and variant frequency. Wakefield’s ABF assumes that the given $SE_{\hat{\beta}}$ and $\hat{\beta}$ capture the information in the study data about the true effect size, β such that

$$\hat{\beta} \sim \mathcal{N}(\beta, SE_{\hat{\beta}}^2). \quad (2)$$

For the large sample sizes found in most GWAS, equation 2 is expected to be a reasonable approximation at all but rare variants (less than 1% frequency). Further discussion of this can be found in section 3.1 of the supplementary material.

Like all Bayesian methods, Wakefield’s ABF includes a prior distribution, which encodes our beliefs about the true effect size β . The prior is determined by a single scaling parameter σ that is set by the analyst,

$$\beta \sim \mathcal{N}(0, \sigma^2). \quad (3)$$

Large σ corresponds to a belief that true effect sizes can be large, and analogously for small σ . Values commonly used in the literature for disease studies are 0.2 and 0.4 (Marchini & Band, n.d.; Stephens & Balding, 2009). The value $\sigma = 0$ encodes a belief that the variant has no effect on the trait under investigation, and will be used as the null model, M_0 in our analysis.

For given $\hat{\beta}$, $\text{SE}_{\hat{\beta}}$, and prior σ , let $f(x; m, s^2)$ be the probability density function of the Gaussian distribution with mean m and variance s^2 , evaluated at x . Then Wakefield’s ABF is simply

$$\text{ABF} = \frac{f(\hat{\beta}; 0, \text{SE}_{\hat{\beta}}^2 + \sigma^2)}{f(\hat{\beta}; 0, \text{SE}_{\hat{\beta}}^2)}. \quad (4)$$

Because this is a ratio of normal densities, it has a closed form expression that can be evaluated very quickly.

We note that the prior distribution on the true effect size, β , shown in Equation 3, has most of its probability mass close to zero, which is where the null model has all of its mass. Thus, the Bayes factor in equation 4 has a minimum, non-zero value at $\hat{\beta} = 0$, while the maximum value is unbounded. It will rarely provide strong evidence in favor of the null (M_0) as the alternative model includes very small effects close to zero. Other non-local priors (where the prior density of the alternative model drops to zero near zero), have been described (Johnson & Rossell, 2010). The lack of strong evidence in favor of the null reflects the fact that, in the GWAS context, small effect sizes are difficult to rule out. Furthermore, a GWAS typically aims to identify SNPs that show strong evidence (say, $\text{ABF} > 1000$) for an association with a trait, rather than those that show strong evidence for no association.

2.3 Approximate Bayes factors for meta-analysis

We describe our multivariate extension of the ABF to meta-analysis. Instead of calculating an ABF for an association between a given variant and a given trait measured in one GWAS,

we calculate an ABF for an association between a variant and an arbitrary number of traits, n , measured in independent or (partially) overlapping studies. Recently, multivariate analysis of GWAS summary statistics has been implemented by the MTAG approach (Turley et al., 2018) and a framework involving Bayes factors to compare models of association has been developed to help determine relevant tissues in eQTL data (Flutre et al., 2013). When extending Wakefield’s ABF to the multivariate case, $\hat{\beta}$, which was a single effect size estimate from a single GWAS, becomes $\hat{\boldsymbol{\beta}}$, an n -vector of effect size estimates from each of the n studies included in the meta-analysis.

$$\hat{\boldsymbol{\beta}} = \left(\hat{\beta}_1, \dots, \hat{\beta}_n \right)^\top.$$

Similarly, $\text{SE}_{\hat{\beta}}$ is replaced by the study covariance matrix, $\mathbf{V}_{\hat{\boldsymbol{\beta}}}$, of dimension $n \times n$,

$$\mathbf{V}_{\hat{\boldsymbol{\beta}}} = \begin{bmatrix} \text{SE}_1^2 & r_{1,2}\text{SE}_1\text{SE}_2 & \cdots & r_{1,n}\text{SE}_1\text{SE}_n \\ r_{1,2}\text{SE}_1\text{SE}_2 & \text{SE}_2^2 & \cdots & r_{2,n}\text{SE}_2\text{SE}_n \\ \vdots & \vdots & \ddots & \vdots \\ r_{1,n}\text{SE}_1\text{SE}_n & r_{2,n}\text{SE}_2\text{SE}_n & \cdots & \text{SE}_n^2 \end{bmatrix}. \quad (5)$$

If the studies are all independent—that is, they do not share samples—then the $r_{i,j}$ terms in $\mathbf{V}_{\hat{\boldsymbol{\beta}}}$ are all 0, and $\mathbf{V}_{\hat{\boldsymbol{\beta}}}$ is simply a diagonal matrix of study-wise variances. If studies are not independent, then some or all of the off-diagonal elements of $\mathbf{V}_{\hat{\boldsymbol{\beta}}}$ will be non-zero. If the amount of overlap between studies is known, then the values of r can be calculated from formulas provided by Zaykin and Kozbur (Zaykin & Kozbur, 2010) and Bhattacharjee *et al.* (Bhattacharjee et al., 2012).

However, having information on exactly which samples were included in which consortia and analyses is increasingly difficult, especially when dealing with summary statistics, necessitating approaches that estimate the covariance between studies directly from the data. The authors of the MTAG approach (Turley et al., 2018) used the intercept of the pairwise LD score regression (B. K. Bulik-Sullivan et al., 2015) to create an analogous matrix. This

can be applied to our method as well. As we show later, it is also possible to get similar estimates by simply calculating empirical correlations in a chosen subset of $\hat{\beta}$ s between the two studies.

Next we consider a multivariate extension of σ , which we call Σ . As in the univariate case, each study $i \in \{1, \dots, n\}$ has a prior parameter σ_i determining the expected effect sizes. However, combining multiple studies introduces the possibility of correlated true effects between pairs of them, which we capture in the terms $\rho_{i,j} : i, j \in \{1, \dots, n\}$.

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho_{1,2}\sigma_1\sigma_2 & \cdots & \rho_{1,n}\sigma_1\sigma_n \\ \rho_{1,2}\sigma_1\sigma_2 & \sigma_2^2 & \cdots & \rho_{2,n}\sigma_2\sigma_n \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1,n}\sigma_1\sigma_n & \rho_{2,n}\sigma_2\sigma_n & \cdots & \sigma_n^2 \end{bmatrix}. \quad (6)$$

This matrix has a similar role as the Ω matrix in the MTAG approach (Turley et al., 2018). Thus, by writing $f(\mathbf{x}; \mathbf{m}, \mathbf{V})$ for the density of a multivariate normal distribution with mean vector \mathbf{m} and covariance matrix \mathbf{V} , evaluated at \mathbf{x} , Equation 4 becomes

$$\text{ABF} = \frac{f(\hat{\beta}; \mathbf{0}, \mathbf{V}_{\hat{\beta}} + \Sigma)}{f(\hat{\beta}; \mathbf{0}, \mathbf{V}_{\hat{\beta}})}. \quad (7)$$

2.4 Choosing prior parameter values

Unlike $\mathbf{V}_{\hat{\beta}}$, which is in theory can be defined by the data, Σ reflects the prior beliefs of the researcher about the similarity of the effects between studies. One of the most common ways to meta-analyze a set of GWAS summary statistics is the inverse-variance weighted fixed-effects model. This model assumes that the underlying true effect of an associated variant is the same in all cohorts, and the differences in the estimated effect sizes arise due to statistical noise. Larger studies produce more accurate estimates of the underlying true effect producing estimators with smaller variances. Thus, weighting by the inverse of the

variance gives a larger study a greater contribution to the meta-analyzed estimate than a smaller one. In our framework, this fixed-effects model is encoded by setting all the $\rho_{i,j}$ terms of Σ equal to 1.

Another common meta-analysis method is the random-effects model. This assumes heterogeneity in true effect sizes across studies. There are many ways of implementing this assumption, many of which have inherently reduced power to detect effects compared to the fixed effects approach (Han & Eskin, 2011). Our approach allows for a number of different priors that account for heterogeneity in effects found in different studies. They are summarized, along with the fixed-effects prior, in Table 1.

Table 1

The prior models described above allow us to quantify the evidence of association through an ABF in which the alternative model can take a range of forms. The simplest models assume that all studies draw their effects from the same marginal distribution and are equally correlated with one another, such that the prior can be specified with just a single σ and a single ρ parameter. When $\rho < 1$, the prior allows some studies to have large effects while others can still have effects very close to zero since the prior on β is always centered on zero. In this way, we expect an ABF which assumes all studies have an effect with $0 < \rho < 1$ to show evidence for the alternative model, that is, $\text{ABF} > 1$, even when only a subset of the studies have truly nonzero effects.

2.5 Subset exhaustive model averaging

In the analysis of a diverse set of traits, it is natural to consider the possibility that only a subset of studies have non-zero effects ($\sigma_i^2 > 0$) while the rest are null ($\sigma_i^2 = 0$). For n studies and fixed values of ρ and σ , there are 2^n possible subset models, including the null model of no association. A natural Bayesian approach for detecting associations under these assumptions is to apply model averaging over the subset models to obtain a single

summary of the evidence for models of association against the null model of no association. By storing the ABF for each possible model, the analysis can also generate the full posterior distribution. These analyses require specifying a prior probability on each model. When all studies are considered equally likely to have an effect, then possible prior distributions include:

Uniform prior: This assumes that every model of association is equally likely, and thus has the prior probability p of

$$p = \frac{1}{2^n}. \quad (8)$$

Combinatorial prior: When calculating the total prior weight on the models with m associated traits out of all possible n traits under the uniform prior, we see that the models with $\frac{n}{2}$ traits ($\frac{n}{2} \pm 0.5$ if n is odd) have the largest prior weight since the number of these models is the largest. The combinatorial prior forces the prior weights on the number of associations to be the same. That is, for a model with a total of m associations, the prior probability p on the model is

$$p = \frac{m!(n-m)!}{(n+1)!}. \quad (9)$$

Binomial prior: Both the uniform prior and the combinatorial prior assume that a model with no associations is equally likely as a model where all n traits are associated. In reality, we might assume that each of the studies has some probability $q_i : i \in \{1, \dots, n\}$ of having a true association with a variant. Alternatively, these q_i can be thought of as the proportion of SNPs with which the i th study is expected to be associated. This is similar to π , the prior on the expected proportion of true effects that Stephens and Balding (Stephens & Balding, 2009) discuss for the calculation of the posterior probability of association. For a given model with subset $K \subseteq \{1, \dots, n\}$ of studies showing associations with the variant and subset $L = \{1, \dots, n\} \setminus K$ showing no association, the prior probability of the model is

$$p = \prod_{i \in K} q_i \prod_{j \in L} (1 - q_j). \quad (10)$$

The uniform prior can be recreated in this prior by setting all $q_i = 0.5$.

In general, we do not recommend using the combinatorial prior, as it places relatively high prior weight on the null model or the model with all traits associated. This becomes more pronounced as the number of studies in the meta-analysis, n , increases. This is because for all values of $n > 1$, there is exactly one null model and one model where all n traits are associated, while the number of models with one to $n - 1$ associations increases with n , so the prior weight on any one of these models decreases with n , which also decreases their relative weights compared to the null and n associated models. Furthermore, in real world applications, it is unlikely that one would have an *a priori* belief in the number of true underlying associations while being completely agnostic to which specific traits are associated.

The binomial prior provides a principled way of upweighting particular subset models. Setting a low prior probability on each trait (that is, $q_i < 0.5, \forall i \in \{1, \dots, n\}$) increases the weight on models with fewer associations, which might be appropriate if sharing of associations is deemed unlikely *a priori*. Conversely, setting $q_i > 0.5$ encodes a belief that associations are more likely, and for any pair of studies with prior probabilities greater than 0.5, sharing of associations is assumed to be more likely than not. In this way, assumptions about the expected level of pleiotropy can be incorporated via the binomial prior.

By making use of the generality of the prior covariance matrix Σ , a range of other models can be defined based on particular assumptions on the relationship between traits. Where it might be appropriate to assume that the marginal effect at a variant is representative of the genome average, the estimates of the genetic correlation between two studies might be appropriate as prior information, and it has been shown that this can increase power in some scenarios (Turley et al., 2018).

In practice, for analyses of large numbers of studies, we advocate selecting a relatively small set of prior matrices which best represent possible models as a way to overcome the exponentially growing model space and the fact that prior probability on any one subset model becomes smaller with as the number of studies increases. The rationale is similar to the approach taken in the analysis of eQTL data (Flutre et al., 2013). We can average across all of the ABFs calculated under the various priors to get a single quantity that accounts for our uncertainty about the exact prior model.

It is also of interest to assess models in which effects are assumed to be negatively correlated between some pairs of studies. A challenge is that the number of possible configurations of positive and negative association grows even more quickly than the subset models. Taking absolute values of the effect sizes changes the sampling distribution to the folded-Gaussian distributions which are difficult to work with. If negative correlations are of interest, or should be entertained as possible, then assuming independent effects ($\rho = 0$) is a practical option, although with the cost of losing the dependency information between the effect sizes.

2.5.1 Shotgun stochastic search

If the number of studies in the meta-analysis is so large that an exhaustive search over all subset models is not feasible for a given Σ , a shotgun stochastic search (SSS) (Hans, Dobra, & West, 2007) can be performed to determine the most likely models of association. This iterative method calculates the ABFs for the set of models in the “neighborhood” of the current index model. This neighborhood contains models that (i) add an association to the index model, (ii) remove an association from the index model, or (iii) replace exactly one nonzero study from the index model with one that is zero in the index model. The algorithm then chooses the next index model based on a probability distribution generated from normalizing the ABFs of the models in the neighborhood of the index model. This type of search has been used to fine-map genetic loci (Benner et al., 2016) and is well-suited to

our ABF analysis, as it quickly finds the models with high ABFs. In every scenario we have investigated, in both real and simulated data, the distribution of ABFs has been unimodal, meaning that the search is unlikely to miss the global maximum of the distribution due to being caught in a local one.

2.6 Posterior distribution of effect sizes

Within the Bayesian framework described above, it is natural to compute the posterior on the true effect sizes under different assumptions about the presence and size of an effect within a study, and the correlation in effects across studies. Often when describing the interpretation of a forest plot, like that illustrated in Figure 1, the analyst informally imposes these beliefs in drawing conclusions. It is therefore of interest to assess formally the impact of these beliefs on our inference about the true effects.

Using standard results for the posterior distribution of the mean given a multivariate normal likelihood and a normal prior:

$$\boldsymbol{\beta} \sim \mathcal{MVN}\left((\boldsymbol{\Sigma}^{-1} + \mathbf{V}_{\hat{\boldsymbol{\beta}}}^{-1})^{-1}(\mathbf{V}_{\hat{\boldsymbol{\beta}}}^{-1}\hat{\boldsymbol{\beta}}), (\boldsymbol{\Sigma}^{-1} + \mathbf{V}_{\hat{\boldsymbol{\beta}}}^{-1})^{-1}\right) \quad (11)$$

A posterior forest plot can then be drawn with point estimates based on the posterior mean, and with the marginal variance on the posterior effect size taken from the diagonal elements of the posterior variance-covariance matrix. The impact of the prior matrices described above on the posterior forest plot typically includes: shrinkage of effects towards zero, when they are believed to be non-zero; increased certainty in the effect size estimates of two studies with similar effects when they are assumed to be correlated, with the certainty at its maximum when the studies are independent; and increased uncertainty in true effects which are assumed to be correlated when the observed effects differ significantly. We note that it would be possible model average the posterior distribution over models by numeric evaluation. In this case, the marginal posterior distribution of effect size within a study is

no longer necessarily unimodal.

3 Results

3.1 Simulations and statistical properties

To understand the statistical properties of our method compared to standard approaches to meta-analysis, we performed simulations of 5 independent genome-wide association studies at a single variant. We found that, assuming sample sizes are moderate (> 1000 individuals) and allele frequencies not too rare (> 0.01), it was possible to simulate effect sizes and standard errors directly from the model described above, where true effects can either be fixed at a given value, or sampled from a prior distribution (see Supplementary Material 6.3). The ability to efficiently simulate directly from the approximate model is useful for assessing frequentist properties of the Bayesian approach.

As an illustrative example Figure 2 shows the estimated power of a standard inverse-variance weighted fixed effects meta-analysis approach and Fisher’s method of combining p -values, as well the power of a MetABF which assumes that all studies have an effect. Results are shown for MetABFs that assume effects are uncorrelated ($\rho = 0$), or highly correlated ($\rho = 0.96$). The MetABF with highly correlated effects performs well when there is an effect in all 5 studies, although not quite as well as the fixed effects approach when the fixed effects assumption is correct. However it has substantially more power when only a subset of studies—for example 2 out of 5—have an effect in which case the highly correlated MetABF has 50% more power than the fixed effects approach.

Figure 2

We used simulations under the null to estimate p -values for the ABF using priors which assumed effects were present in all 5 studies (as in Figure 2), and the MetABF model that explicitly averages over all possible subsets of true effects. We note that when the ABF is

based on a single variance-covariance matrix for the prior, i.e. it is not model averaged, then it is possible to derive the distribution of ABFs under the null, given a frequentist test statistic (see Supplementary Material 6.4). We performed simulations under a range of different subset effects with different priors on effect sizes and correlations and compare statistical power to the two frequentist methods described above as well as a random effects approach with a closed-form calculation (DerSimonian & Laird, 1986) (see Figure 8 in the Supplementary Material).

Briefly, we found that the MetABF approach was similar in power to the best frequentist tests, was more powerful when a subset of the traits in the meta-analysis were truly associated. Of the three alternative approaches we tested, Fisher’s method came closest in terms of power with the Bayes factor approaches when there are three or more studies with associations to find. We also found that while the MetABF sometimes loses power when the prior correlation on the true effects is different from the true underlying correlation, the loss is not large — the results of the MetABF analyses using different values of ρ (but the same prior σ) tended to differ by less than 0.005. This suggests that the MetABF approach is fairly robust to the choice of prior ρ .

3.2 Comparison to other methods

We compared our method to both CPBayes (Majumdar et al., 2018) and MTAG (Turley et al., 2018) using simulated data. Section 7 in the supplement provides a more detailed discussion of our findings. Briefly, at any given false positive rate (proportion of spurious associations in the posterior model), CPBayes had the highest true positive (proportion of true associations in the posterior model) rate of the three methods, while MetABF and MTAG were very similar in performance. However, neither MTAG nor CPBayes allow the user to explore the posterior probability space the way MetABF does. Additionally, the gain in accuracy in CPBayes comes at a considerable time cost. It took roughly 8 minutes on average to analyze a dataset of 60,000 SNPs with MetABF, while CPBayes took an average

of 2.3 days when studies were assumed to have no cryptic relatedness between them, and an average of 6.6 days when they were not. Both CPBayes and MetABF are implemented as R packages. MTAG, which is implemented in Python, was by far the fastest method, taking an average of 15 seconds on the full data; however, this does not include the time taken to create the LD score reference panels.

3.3 Application to WTCCC2 data

To investigate the applicability of the approach across a large number of heterogeneous GWAS, we applied it to summary statistic data from the following Wellcome Trust Case Control Consortium 2 (WTCCC2) studies. These included

Autoimmune diseases: 1) ankylosing spondylitis (AS); 2) multiple sclerosis, UK cohort (MS_UK); 3) multiple sclerosis, non-UK European cohort (MS_nonUK); 4) psoriasis (PS); 5) ulcerative colitis (UC).

Infectious diseases: 6) bacteremia, all types (BS_overall); 7) pneumococcal bacteremia (BS_pneumococcus); 8) visceral leishmaniasis, Indian cohort (VL_India); 9) visceral leishmaniasis, Brazilian cohort (VL_Brazil).

Stroke cohorts: 10) ischemic stroke, large vessel subtype (IS_TOAST_1); 11) ischemic stroke, small vessel subtype (IS_TOAST_2); 12) ischemic stroke, cardioembolic subtype (IS_TOAST_3).

Reading and mathematics cohort: 13) reading scores (RM_reading); 14) mathematics scores (RM_maths).

Psychiatric traits: 15) schizophrenia (SP); 16) psychosis endophenotypes (PE).

Studies of unique traits: 17) Parkinson’s disease (PD); 18) Barrett’s esophagus (BO); 19) metformin response (PR); 20) glaucoma (GL).

Studies were mainly conducted in European cohorts, but two studies (BS and VL) used non-Europeans samples. Three different genotyping arrays were used, and four studies were imputed to increase the number of SNPs in the dataset (see Supplementary Tables 7 and 8

for details). As a result, summary statistic data were typically available for only a subset of studies at any given SNP. Our analysis included all SNPs for which there were summary statistics in at least two of the studies. Before performing our analysis we harmonized the data using a pipeline to align the SNPs in each study to the forward strand, thus ensuring that the “A” and “B” alleles of each SNP were the same across all studies, and corresponded to the reference/alternative alleles in the 1000 Genomes database. Effect sizes were estimated for the alternative allele. Details about the processing and availability of the data can be found in Section 6.7.1 of the Supplementary Material.

To search for signals of association we calculated a model averaged ABF at each SNP across the genome. To capture the range of possible patterns of association across studies, we selected 12 models with different assumptions about the correlation structure of effect sizes among the studies. These are described in section 3.4 below. We found that subset models that assumed an effect only within one study (and explicitly no effect in other studies) often had low probability, presumably because conditional on one study showing a true effect, the chance that no other study shows any effect becomes increasingly small (particularly for related traits) as the number of studies increases. We calculated the MetABF for each of these models with different values of σ (0.1, 0.2 and 0.4). We then took the mean across all resulting MetABFs, assuming them to be equally likely *a priori*, to obtain a single summary of the association evidence across all studies.

3.4 Prior matrices

Briefly, the 12 priors used in our analyses assume non-zero effects across all studies. They differ in the presumed correlation structure, ranging from completely uncorrelated across all studies (model 1), through to prior matrices where correlated effects only occur within a subset of studies (for example IS, MS, BS and RM, models 5-7) or within study classes (for example autoimmune disease; AS, PS, MS and UC: models 8-12). Figure 3 illustrates each of the prior models used. The traits are listed in the order described at the start of Section

3.3, grouped by disease/trait class.

Figure 3

3.5 Study matrix

Among the 20 WTCCC2 cohorts, nine of them used samples from a shared pool of controls. Additionally, one of the bacteremia (BS) cohorts was a subset of the other, while the two reading and mathematics cohorts (RM) comprised of two different phenotypes measured on the same individuals. This introduced correlation in the effect size estimates due to shared statistical noise, and therefore non-zero covariance terms between some of the studies in the $\mathbf{V} = \mathbf{V}_{\hat{\beta}}$ matrix (see Equation 5), which we needed to estimate (Table 2).

When estimating the correlation directly from the genome-wide $\hat{\beta}$ -statistics, we used several approaches: taking the intercept from LD score regression (LDSC) (B. Bulik-Sullivan et al., 2015; B. K. Bulik-Sullivan et al., 2015; Turley et al., 2018), which provides an estimate of trait covariance due to undetected shared controls; using all the available data and empirical correlation of $\hat{\beta}$ -estimates over three subsets of the SNPs (all SNPs; thinning out the SNPs by the recombination distance; and removing SNPs that had significant associations with $p < 5 \times 10^{-8}$ as well as those in the MHC region). Broadly, the estimated correlations were similar across approaches, however the removal of significant associations and the MHC region tended to increase the estimated correlation, as did the LDSC analysis. These observations suggest that strong association signals create additional variance in the effect size estimates that can reduce the correlation if not accounted for.

For the analysis that follows, we assumed $\mathbf{V}_{i,j,p} = r_{i,j}SE_{i,p}SE_{j,p}$, where i and j indexed the study pairs and p indexed the SNP. Because LD score regression was not able to run on all of our data, due to small ($<100,000$) overlaps between the markers in software’s reference panels and the markers in some of our cohorts, the pairwise correlations $r_{i,j}$ were based on the genome-wide correlations in $\hat{\beta}$ estimated after the removal of the MHC region and other

significant associations, and $SE_{j,p}$ were taken directly from the original association analyses.

Table 2

Due to simplicity of the multivariate normal approximation, it is quick to simulate effect size estimates and standard errors under the null for all 20 traits in our analyses. Using these simulated data, we can calculate ABFs under a range of models, across all traits, and simultaneously account for the non-independence of the studies. It is then possible to check, via a quantile-quantile comparison, whether the observed ABFs are distributed as expected under the null at the majority of SNPs, which would provide confidence that positive ABFs represent evidence for genuine departures from the null model, as opposed to being a systematically mis-calibrated test statistic. Supplementary Figure 13 shows the result of the simulations across the genome for each of the 12 prior models. The resulting quantiles are closely matched at the majority of SNPs suggesting that the null model fits the data at the majority of SNPs. As expected there is a deviation in the tail, where the observed ABFs are bigger than expected, suggesting the alternative model is a better explanation for the data. We caution that in the analysis of genome-wide association data, these simulations do not account for the correlation between SNPs due to linkage disequilibrium.

3.6 Genome-wide analysis

Figure 4

The results of the genome-wide analysis in Figure 4 show the mean MetABF at each SNP across all 36 combinations of 12 prior matrices and three values of σ ($\sigma = \{0.1, 0.2, 0.4\}$), using the model that assumes an association in every trait. Additionally, we curated two lists: one of SNPs reported in each WTCCC2 publication as being implicated by previous studies, and the other of novel loci identified as genome-wide significant (often after replication) by each publication. These are given in the Supplementary Material (Tables 9 and 10). Of

the 35 previously identified associations, 18 had a model averaged MetABF $> 10^4$ in our analysis.

For regions showing a strong evidence of association (model-averaged MetABF $> 10^4$) we further explored the patterns of association, examining all possible subset models across all 12 prior correlation matrices and three values of prior effect size parameter σ . For markers with information for at least 19 studies, we used a shotgun stochastic search (Hans et al., 2007) to avoid making all 2^{19} or 2^{20} calculations. Because of the large number of possible models, and because we assumed each to be equally likely *a priori* (uniform prior), this approach imposes a strong prior belief in the models with an intermediate number of non-zero effects. We show the marginal probability of the study being included (averaged across all possible subset models) and the most likely model in Supplementary Figure 14. Outside of the MHC region (the strong peak of association on chromosome 6), several loci stand out as potentially having effects across a range of traits. To illustrate the impact of the MetABF analysis on the interpretation of the pattern of association we highlight three SNPs. The forest plots of the original association summary statistics and the posterior distribution of effects under the most likely model are shown in Figure 5. We discuss these illustrative examples below.

Figure 5

- *SLC44A2* locus - rs8106664. The analysis highlights a SNP near the gene *SLC44A2* which shows a strong association with MS, as well as associations with both psoriasis and ischemic stroke. The SNP was not identified in the 2011 multiple sclerosis analysis (Sawcer et al., 2011), but is in strong LD ($r^2 = 0.9347$ in 1000 Genomes EUR populations) with a missense variant that was subsequently confirmed (Beecham et al., 2013). Its prominence in this analysis is driven by the signals in other auto-immune diseases which boost the signal in MS and increases the confidence in the effect size. The most likely model also includes an effect in Parkinson’s disease, which under the

prior correlation matrix (prior model 12, which assumes high correlations among traits from the same class, but no correlations otherwise) is independent of the other studies, allowing it to have an effect in the opposite direction. *SLC44A2* is upstream of *LDLR*, a well known lipid locus, and downstream of *TYK2*, which has a complex association with multiple auto-immune diseases. One possibility is that this SNP is linked to multiple signals of association and therefore is not necessarily the causal one for every, or even any, of the traits showing association (see Supplementary Figure 15). In any case, this analysis highlights a genetic variant that is a marker for susceptibility to multiple traits.

- Chromosome 17 inversion - rs1981997. There is a large polymorphic inversion on chromosome 17 (Stefansson et al., 2005; Tobin et al., 2008) which contains several important genes, including the gene *MAPT* which shows a strong signal of association with Parkinson’s disease (PD) (Skipper et al., 2004; Steinberg et al., 2012; Tobin et al., 2008). A number of SNPs which tag the inversion show strong evidence of association in our analysis. The size of the effect in PD leads to the top model having an effect size prior with a variance of 0.2^2 , which increases the plausibility of large effects in other studies. Interestingly, the top models include associations in glaucoma in the opposite direction to the PD risk, and Barrett’s oesophagus (BO) showing effects in the same direction with PD.
- *ERAP1* mutation - rs30187. The strongest signal of association outside of the MHC was at rs30187, which was previously reported as being significantly associated with both PS and AS, and these effects also drive the evidence for association in our analysis. The variant is a missense change (Lys528Arg) with the effect allele associated with impaired peptide trimming, which is more strongly protective in the presence of a pre-disposing HLA allele (Evans et al., 2011). The top model identifies this effect as being specific to AS and PS. This model also assumes effect sizes are strongly correlated between all diseases that have an effect, with a prior variance of 0.2^2 . These assumptions couple

the effect size estimates, marginally reducing the posterior estimates of the study-wise standard errors for AS and PS, and increase the effect in PS to become more consistent with AS.

The above examples highlight how the subset models can explicitly quantify the extent to which that the data are consistent with no effect at all in a subset of studies, and assess assumptions about the size and correlation in true effects across the remaining studies. The analysis explores the posterior probability on $2^{20} \times 12 \times 3 = 37,748,736$ models, and therefore has substantial flexibility to reveal patterns from the data. An advantage of this analysis is that any two models can be compared directly against each other to explore how well the data support one model over the other.

4 Discussion

Here we have introduced MetABF, a method for searching for cross-trait associations using GWAS summary statistics. The Bayesian approach allows the expected relationships between studies or traits to be encoded in the analysis. When effects are assumed to be correlated between studies, a strong effect in one study in the meta-analysis automatically adjusts the threshold of evidence required to discover extra associations among related studies. Because the calculations are fast, we can compare different models of association under different priors directly to determine which fits the data best. This allows us to make probabilistic statements about which studies showed true effects at a given marker.

This type of approach is attractive in its ability to combine a large number of heterogeneous GWAS, so to demonstrate how it can be applied, we jointly analyzed 20 different traits across the WTCCC2 data. Our results highlighted loci that showed multiple associations across traits studied which did not achieve genome-wide significance in the individual GWAS. Two different genotyping protocols were used across all the WTCCC2 studies, and some studies contained imputed markers, meaning that there were data for all traits at fewer

than 5% of the markers in our meta-analysis. In general, our method would tend to favor a SNP for which there is information on more studies over a SNP for which there is less data, but which shows a slightly stronger effect in one of the studies.

Although we have largely focused our motivation on meta-analyzing different traits, the approach is also applicable to the meta-analysis of the same trait in different populations, as was performed in a study of severe malaria in different African ethnic groups (Band et al., 2013). In this setting, the prior correlations could be derived from a measure of similarity between populations, rather than from a measure of the similarity of traits. Some consideration is needed for the assumed source of heterogeneity in effect size between studies carried out in different ancestry groups. Heterogeneity in effect size might derive from causal variants having different correlations with genotyped SNPs, due to different patterns of linkage disequilibrium. Or, if it assumed that the causal variant is typed directly, then heterogeneity in effect might derive from an interaction between the causal variant and other variants in the genome, or with the environment.

Cross-trait analyses, including ours, have limitations. Specifically, markers with apparent cross-study associations may be tagging multiple distinct causal loci. This is a problem inherent in GWAS in humans, where the typed marker is rarely the causal variant, but in linkage disequilibrium with it (Visscher, Brown, McCarthy, & Yang, 2012). Implicitly this up-weights SNPs that tag multiple causal SNPs and complicates direct interpretation of the patterns of associations across studies. Even when the same typed marker is statistically significant for multiple GWAS, the underlying causal markers may be different, and this may be exacerbated by differences in the patterns of linkage disequilibrium between populations under examination. Colocalization methods (Giambartolomei et al., 2014; Hormozdiari et al., 2016; Wen, Pique-Regi, & Luca, 2017) may be employed to determine if multiple causal variants are likely for regions that show associations with multiple traits.

An additional challenge is that for every study that is added, the number of possible models of association doubles. In a large meta-analysis, it is computationally challenging to

fully explore the possible subset space, and to conceptualize their prior probability because each model becomes more unlikely on its own. In the discovery phase of the application to real data described here, we used a set of prior correlation matrices to cover the model space in which we thought true associations might exist, each with equal prior weight. These were then fully explored using a uniform prior on each possible model of association. When the prior probability of there being a true association in any given study is small, the binomial prior may be of interest as it will naturally apply shrinkage to the posterior estimate of the studies that have true effects, thereby reducing the prior probability on a pleiotropic models. However, the model averaging approach is valid for any prior weighting on models, and these should depend on beliefs about the prior probability of each specific combination of effects where possible.

These challenges increase further if combinations of both positive and negative correlations are explicitly considered. It is possible that when the effect size estimates are precise (low SE), many of the subset models will have almost zero probability. In this scenario, the subset configuration space might be more efficiently explored by approaches such as Markov chain Monte Carlo (Smith & Roberts, 1993), or shotgun stochastic search (Hans et al., 2007). Here we have set parameters for the appropriate prior distributions. There is however considerable scope for fitting or estimating the parameters on the size and correlation of effects across models, and explicitly inferring the prior probability of SNPs that derive from each of these models across the genome.

The genome-wide association study has become a standard approach to gaining insights into the etiology of traits and disease—so common that it has been performed in an automated manner on thousands of traits in the UK Biobank (Canela-Xandri, Rawlik, & Tenesa, 2018; Howrigan, Abbott, Churchhouse, & Palmer, n.d.). As more biobanks are gathered, and as researchers start to peruse them for genetic associations, MetABF is a useful tool to determine cross-trait associations from GWAS. Its reliance on summary statistics means that access to the raw genetic data is not necessary. For a reasonable number of traits it

can be applied genome-wide, overcoming some of the limitations of a traditional phenome-wide association study (PheWAS) (Cortes et al., 2017; Denny et al., 2010; Pendergrass et al., 2011). We have found the approach to aid in the assessment of which models of true associations are most consistent with the observed summary statistics at a variant.

5 Acknowledgements

We thank the participants in the studies that generated the data for analysis within this manuscript. GM was supported by a Wellcome Trust grant (100956/Z/13/Z). CS was supported by a Wellcome Trust Career Development Fellowship grant (097364/Z/11/Z). MP was supported by grants from the Academy of Finland (288509, 294050, 312076). LJD is supported by the Wellcome Trust and the Royal Society (208750/Z/17/Z), and the Kennedy Trust for Rheumatology Research. CS is a shareholder and employee of Genomics PLC. GM is a director of and shareholder in Genomics PLC and a partner in Peptide Groove LLP. MP has provided consultancy services for Genomics plc. LJD has provided consultancy services for Genomics plc.

References

- Asimit, J. L., Panoutsopoulou, K., Wheeler, E., Berndt, S. I., the GIANT consortium, the arcOGEN consortium, Cordell, H. J., ... Barroso, I. (2015, December). A Bayesian Approach to the Overlap Analysis of Epidemiologically Linked Traits. *Genetic Epidemiology*, 39(8), 624–634. Retrieved 2016-07-15, from <http://doi.wiley.com/10.1002/gepi.21919> doi: 10.1002/gepi.21919
- Band, G., Le, Q. S., Jostins, L., Pirinen, M., Kivinen, K., Jallow, M., ... Spencer, C. C. (2013, May). Imputation-based meta-analysis of severe malaria in three African populations. *PLoS Genet.*, 9(5), e1003509.
- Beecham, A. H., Patsopoulos, N. A., Xifara, D. K., Davis, M. F., Kempainen, A., Cotsapas, C., ... McCauley, J. L. (2013, September). Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nature Genetics*, 45(11), 1353–1360. Retrieved 2017-05-14, from <http://www.nature.com/doi/10.1038/ng.2770> doi: 10.1038/ng.2770
- Bellenguez, C., Bevan, S., Gschwendtner, A., Spencer, C. C. A., Burgess, A. I., Pirinen, M., ... Markus, H. S. (2012, March). Genome-wide association study identifies a variant in HDAC9 associated with large vessel ischemic stroke. *Nat Genet*, 44(3), 328–333. Retrieved from <http://dx.doi.org/10.1038/ng.1081> doi: 10.1038/ng.1081
- Benner, C., Spencer, C. C., Havulinna, A. S., Salomaa, V., Ripatti, S., & Pirinen, M. (2016, May). FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics*, 32(10), 1493–1501. Retrieved 2017-03-28, from <https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btw018> doi: 10.1093/bioinformatics/btw018
- Bhattacharjee, S., Rajaraman, P., Jacobs, K. B., Wheeler, W. A., Melin, B. S., Hartge, P., ... Chatterjee, N. (2012, May). A Subset-Based Approach Improves Power and Interpretation for the Combined Analysis of Genetic Association Studies of Heterogeneous Traits. *The American Journal of Human Genetics*, 90(5), 821–835. Retrieved 2016-

- 07-15, from <http://linkinghub.elsevier.com/retrieve/pii/S0002929712001590>
doi: 10.1016/j.ajhg.2012.03.015
- Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P.-R., ... Neale, B. M. (2015, September). An atlas of genetic correlations across human diseases and traits. *Nature Genetics*, 47(11), 1236–1241. Retrieved 2016-07-14, from <http://www.nature.com/doifinder/10.1038/ng.3406> doi: 10.1038/ng.3406
- Bulik-Sullivan, B. K., Loh, P.-R., Finucane, H. K., Ripke, S., Yang, J., Patterson, N., ... Neale, B. M. (2015, February). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, 47(3), 291–295. Retrieved 2016-07-13, from <http://www.nature.com/doifinder/10.1038/ng.3211> doi: 10.1038/ng.3211
- Canela-Xandri, O., Rawlik, K., & Tenesa, A. (2018, November). An atlas of genetic associations in UK Biobank. *Nature Genetics*, 50(11), 1593–1599. Retrieved from <https://doi.org/10.1038/s41588-018-0248-z> doi: 10.1038/s41588-018-0248-z
- Consortium, T. W. T. C. C. (2007, June). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447(7145), 661–678. Retrieved from <http://dx.doi.org/10.1038/nature05911> doi: 10.1038/nature05911
- Cortes, A., Dendrou, C. A., Motyer, A., Jostins, L., Vukcevic, D., Dilthey, A., ... McVean, G. (2017, Sep). Bayesian analysis of genetic association across tree-structured routine healthcare data in the UK Biobank. *Nat. Genet.*, 49(9), 1311–1318.
- Cotsapas, C., Voight, B. F., Rossin, E., Lage, K., Neale, B. M., Wallace, C., ... on behalf of the FOCiS Network of Consortia (2011, August). Pervasive sharing of genetic effects in autoimmune disease. *PLoS Genetics*, 7(8), e1002254. Retrieved 2014-07-02, from <http://dx.plos.org/10.1371/journal.pgen.1002254> doi: 10.1371/journal.pgen.1002254
- Denny, J. C., Ritchie, M. D., Basford, M. A., Pulley, J. M., Bastarache, L., Brown-Gentry, K., ... Crawford, D. C. (2010, May). PheWAS: demonstrating the feasibility of a

- phenome-wide scan to discover gene-disease associations. *Bioinformatics*, 26(9), 1205–1210. Retrieved 2016-07-13, from <http://bioinformatics.oxfordjournals.org/cgi/doi/10.1093/bioinformatics/btq126> doi: 10.1093/bioinformatics/btq126
- DerSimonian, R., & Laird, N. (1986, September). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7(3), 177–188. Retrieved 2016-07-12, from <http://linkinghub.elsevier.com/retrieve/pii/0197245686900462> doi: 10.1016/0197-2456(86)90046-2
- Ellinghaus, D., Jostins, L., Spain, S. L., Cortes, A., Bethune, J., Han, B., ... Franke, A. (2016, March). Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nature Genetics*, 48(5), 510–518. Retrieved 2017-05-19, from <http://www.nature.com/doifinder/10.1038/ng.3528> doi: 10.1038/ng.3528
- Evans, D. M., Spencer, C. C. A., Pointon, J. J., Su, Z., Harvey, D., Kochan, G., ... Donnelly, P. (2011, July). Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. *Nature Genetics*, 43(8), 761–767. Retrieved 2016-07-14, from <http://www.nature.com/doifinder/10.1038/ng.873> doi: 10.1038/ng.873
- Flutre, T., Wen, X., Pritchard, J., & Stephens, M. (2013, May). A Statistical Framework for Joint eQTL Analysis in Multiple Tissues. *PLoS Genetics*, 9(5), e1003486. Retrieved 2018-05-03, from <http://dx.plos.org/10.1371/journal.pgen.1003486> doi: 10.1371/journal.pgen.1003486
- Giambartolomei, C., Vukcevic, D., Schadt, E. E., Franke, L., Hingorani, A. D., Wallace, C., & Plagnol, V. (2014, May). Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genetics*, 10(5), e1004383. Retrieved 2017-03-28, from <http://dx.plos.org/10.1371/journal.pgen.1004383> doi: 10.1371/journal.pgen.1004383
- Gilchrist, J. J., Rautanen, A., Fairfax, B. P., Mills, T. C., Naranbhai, V., Trochet, H., ...

- Hill, A. V. S. (2018, March). Risk of nontyphoidal *Salmonella* bacteraemia in African children is modified by STAT4. *Nature Communications*, 9(1), 1014. Retrieved from <https://doi.org/10.1038/s41467-017-02398-z> doi: 10.1038/s41467-017-02398-z
- Han, B., & Eskin, E. (2011, May). Random-Effects Model Aimed at Discovering Associations in Meta-Analysis of Genome-wide Association Studies. *The American Journal of Human Genetics*, 88(5), 586–598. Retrieved 2016-07-29, from <http://linkinghub.elsevier.com/retrieve/pii/S0002929711001558> doi: 10.1016/j.ajhg.2011.04.014
- Hans, C., Dobra, A., & West, M. (2007, June). Shotgun Stochastic Search for “Large p ” Regression. *Journal of the American Statistical Association*, 102(478), 507–516. Retrieved 2016-11-07, from <http://www.tandfonline.com/doi/abs/10.1198/016214507000000121> doi: 10.1198/016214507000000121
- Hindorff, L., MacArthur, J., Morales, J., Junkins, H., Hall, P., Klemm, A., & Manolio, T. (n.d.). *A Catalog of Published Genome-Wide Association Studies*. <http://www.genome.gov/gwastudies>. (Accessed: 2 July 2014)
- Hormozdiari, F., van de Bunt, M., Segrè, A. V., Li, X., Joo, J. W. J., Bilow, M., ... Eskin, E. (2016, December). Colocalization of GWAS and eQTL Signals Detects Target Genes. *American Journal of Human Genetics*, 99(6), 1245–1260. doi: 10.1016/j.ajhg.2016.10.003
- Howrigan, D., Abbott, L., Churchhouse, C., & Palmer, D. (n.d.). *Details and considerations of the UK Biobank GWAS — Neale lab*. Retrieved 2018-08-27, from <http://www.nealelab.is/blog/2017/9/11/details-and-considerations-of-the-uk-biobank-gwas>
- Johnson, V. E., & Rossell, D. (2010, March). On the use of non-local prior densities in Bayesian hypothesis tests. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(2), 143–170. Retrieved 2017-06-24, from <http://doi.wiley.com/10.1111/j.1467-9868.2009.00730.x> doi: 10.1111/j.1467-9868.2009.00730.x

- Majumdar, A., Haldar, T., Bhattacharya, S., & Witte, J. S. (2018, February). An efficient Bayesian meta-analysis approach for studying cross-phenotype genetic associations. *PLOS Genetics*, *14*(2), e1007139. Retrieved 2018-05-03, from <http://dx.plos.org/10.1371/journal.pgen.1007139> doi: 10.1371/journal.pgen.1007139
- Marchini, J., & Band, G. (n.d.). *SNPTEST*. Retrieved from https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html
- Pendergrass, S., Brown-Gentry, K., Dudek, S., Torstenson, E., Ambite, J., Avery, C., ... Ritchie, M. (2011, July). The use of phenome-wide association studies (PheWAS) for exploration of novel genotype-phenotype relationships and pleiotropy discovery. *Genetic Epidemiology*, *35*(5), 410–422. Retrieved 2017-02-01, from <http://doi.wiley.com/10.1002/gepi.20589> doi: 10.1002/gepi.20589
- Pickrell, J. K., Berisa, T., Liu, J. Z., Segurel, L., Tung, J. Y., & Hinds, D. A. (2016, July). Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet*, *48*(7), 709–717. Retrieved from <http://dx.doi.org/10.1038/ng.3570>
- Price, A. L., Spencer, C. C., & Donnelly, P. (2015, Dec). Progress and promise in understanding the genetic basis of common diseases. *Proc. Biol. Sci.*, *282*(1821), 20151684.
- Rautanen, A., Pirinen, M., Mills, T. C., Rockett, K. A., Strange, A., Ndungu, A. W., ... Spencer, C. C. (2016, Jun). Polymorphism in a lincRNA Associates with a Doubled Risk of Pneumococcal Bacteremia in Kenyan Children. *Am. J. Hum. Genet.*, *98*(6), 1092–1100.
- Sawcer, S., Hellenthal, G., Pirinen, M., Spencer, C. C. A., Patsopoulos, N. A., Moutsianas, L., ... Compston, A. (2011, August). Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature*, *476*(7359), 214–219. Retrieved 2016-07-14, from <http://www.nature.com/doifinder/10.1038/nature10251> doi: 10.1038/nature10251
- Skipper, L., Wilkes, K., Toft, M., Baker, M., Lincoln, S., Hulihan, M., ... Farrer, M. (2004, October). Linkage Disequilibrium and Association of MAPT H1 in Parkinson Disease.

- The American Journal of Human Genetics*, 75(4), 669–677. Retrieved 2017-07-19, from <http://linkinghub.elsevier.com/retrieve/pii/S0002929707627189> doi: 10.1086/424492
- Smith, A. F. M., & Roberts, G. O. (1993). Bayesian Computation Via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society. Series B (Methodological)*, 55(1), 3–23. Retrieved from <http://www.jstor.org/stable/2346063>
- Stefansson, H., Helgason, A., Thorleifsson, G., Steinthorsdottir, V., Masson, G., Barnard, J., ... Stefansson, K. (2005, February). A common inversion under selection in Europeans. *Nature Genetics*, 37(2), 129–137. doi: 10.1038/ng1508
- Steinberg, K. M., Antonacci, F., Sudmant, P. H., Kidd, J. M., Campbell, C. D., Vives, L., ... Eichler, E. E. (2012, July). Structural diversity and African origin of the 17q21.31 inversion polymorphism. *Nature Genetics*, 44(8), 872–880. Retrieved 2017-07-19, from <http://www.nature.com/doifinder/10.1038/ng.2335> doi: 10.1038/ng.2335
- Stephens, M., & Balding, D. J. (2009, October). Bayesian statistical methods for genetic association studies. *Nature Reviews Genetics*, 10(10), 681–690. Retrieved 2016-07-27, from <http://www.nature.com/doifinder/10.1038/nrg2615> doi: 10.1038/nrg2615
- Tobin, J. E., Latourelle, J. C., Lew, M. F., Klein, C., Suchowersky, O., Shill, H. A., ... Myers, R. H. (2008, July). Haplotypes and gene expression implicate the MAPT region for Parkinson disease: The GenePD Study. *Neurology*, 71(1), 28–34. Retrieved 2017-07-19, from <http://www.neurology.org/cgi/doi/10.1212/01.wnl.0000304051.01650.23> doi: 10.1212/01.wnl.0000304051.01650.23
- Turley, P., Walters, R. K., Maghzian, O., Okbay, A., Lee, J. J., Fontana, M. A., ... Benjamin, D. J. (2018, February). Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nature Genetics*, 50(2), 229–237. Retrieved from <https://doi.org/10.1038/s41588-017-0009-4> doi: 10.1038/s41588-017-0009-4
- Visser, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012, January). Five Years of

- GWAS Discovery. *The American Journal of Human Genetics*, 90(1). Retrieved 2016-07-26, from <http://linkinghub.elsevier.com/retrieve/pii/S0002929711005337> doi: 10.1016/j.ajhg.2011.11.029
- Wakefield, J. (2007). A Bayesian Measure of the Probability of False Discovery in Genetic Epidemiology Studies. *The American Journal of Human Genetics*, 81(2), 208-227. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0002929707611884> doi: <http://dx.doi.org/10.1086/519024>
- Wakefield, J. (2009, January). Bayes factors for genome-wide association studies: comparison with P -values. *Genetic Epidemiology*, 33(1), 79-86. Retrieved 2014-07-02, from <http://doi.wiley.com/10.1002/gepi.20359> doi: 10.1002/gepi.20359
- Wen, X., Pique-Regi, R., & Luca, F. (2017, March). Integrating molecular QTL data into genome-wide genetic association analysis: Probabilistic assessment of enrichment and colocalization. *PLOS Genetics*, 13(3), e1006646. Retrieved 2017-09-11, from <http://dx.plos.org/10.1371/journal.pgen.1006646> doi: 10.1371/journal.pgen.1006646
- Wen, X., & Stephens, M. (2014a, March). Bayesian methods for genetic association analysis with heterogeneous subgroups: From meta-analyses to gene-environment interactions. *The Annals of Applied Statistics*, 8(1), 176-203. Retrieved 2016-07-15, from <http://projecteuclid.org/euclid.aoas/1396966283> doi: 10.1214/13-AOAS695
- Wen, X., & Stephens, M. (2014b, March). Bayesian methods for genetic association analysis with heterogeneous subgroups: From meta-analyses to gene-environment interactions. *The Annals of Applied Statistics*, 8(1), 176-203. Retrieved 2016-07-15, from <http://projecteuclid.org/euclid.aoas/1396966283> doi: 10.1214/13-AOAS695
- Zaykin, D. V., & Kozbur, D. O. (2010, November). P-value based analysis for shared controls design in genome-wide association studies. *Genetic Epidemiology*, 34(7), 725-738. Retrieved 2016-07-21, from <http://doi.wiley.com/10.1002/gepi.20536> doi: 10.1002/gepi.20536

Figure Legends

Figure 1: Statement of the problem. Given a set of observed effect sizes estimated from data in $n = 4$ studies, we want to make inferences about the true effect sizes. Our joint estimate of the true effects depends on our assumptions of the similarity between the $\hat{\beta}_i$ s being estimated, for example the similarity between traits being studied. A key idea in our approach is to be able to easily assess the evidence for models with various assumptions of similarity between the studies.

Figure 2: Plot of power for two frequentist test and the ABF with two different prior parameter settings for the assumed correlation between studies (ρ). Simulation were performed with a true effect (Odds ratio = 1.1) present in a subset (y-axis) of 5 studies with 1500 cases and 1500 controls. Power was calculated using a significance threshold $\alpha = 0.01$. ABFs were calculated with $\sigma = 0.2$.

Figure 3: Visualizations of the prior correlation matrices of WTCCC2 analyses. Dark red squares correspond to $\rho = 1$, red squares correspond to $\rho = 0.96$, yellow-orange squares correspond to $\rho = 0.5$, light yellow squares correspond to $\rho = 0.1$, and white squares correspond to $\rho = 0$. The traits are listed in the order at the start of Section 3.3 and grouped by disease/trait class.

Figure 4: MetABF analysis of genome-wide association analysis across 20 WTCCC2 studies. Adjacent chromosomes are colored with different shades of blue with SNPs of interest colored as denoted in the legend. The y-axis has a threshold at 20. We highlight the markers with the highest mean ABF in our analysis for each non-MHC region with markers whose mean MetABFs were greater than 10^4 . Yellow dots show markers that had been established as loci associated with one or more of the WTCCC2 traits. Red dots highlight the SNPs with the highest mean MetABF (for a given region) that show an association with at least one trait in the meta-analysis, where the mean MetABF $\geq 10^4$, and which are not genome-wide significant in any of the original WTCCC2 analyses.

Figure 5: Bayesian forest plots of three SNPs highlighted in the genome-wide scan under the prior assumption for the most likely posterior model. The dotted lines and open circles show the original effect size and 95% confidence intervals in the WTCCC2 studies. The filled circles and the solid lines show the mean and 95% credible regions from the posterior distribution under the top model (crosses indicate the SNP was missing). Posterior effects at zero reflect that the top model assumed no effect in the study. Grey densities at the bottom show the prior distribution on the effect size under the top model.

Tables

	Variance of effect size	Correlation in effect sizes	Notes
Null model	$\sigma^2 = 0$	-	Assumes variant has no effect on the phenotype.
Fixed effects	$\sigma^2 > 0$	$\rho = 1$	Equivalent to first applying an inverse-variance weighted meta-analysis to estimates $\hat{\beta}_i$ and $SE_{\hat{\beta}_i}$, and then using these estimates as a single study in equation (4).
Independent effects	$\sigma^2 > 0$	$\rho = 0$	Equivalent to multiplying the single study ABFs across studies, assuming the studies are independent. Similar to Fisher’s method of combining p -values.
Correlated effects	$\sigma^2 > 0$	$0 < \rho < 1$	Similar to a related-effects approach where the effects are assumed to be correlated after being drawn from a common distribution (see text).
Subset model	$\sigma_i^2 > 0$ for i in a subset $I \subseteq \{1, \dots, n\}$	$0 \leq \rho_{ij} \leq 1$ for $i, j \in I$	Models heterogeneous patterns of effects, where only a subset of the underlying parameters are nonzero. May be most appropriate to model as fixed, independent, or correlated within the set of nonzero effects, depending on the setting.

Table 1: Different models of association across studies and their relation to parameters of the prior model. Here σ and ρ are as defined in text and assumed to be the same across studies.

Table 2: Comparison of the 39 non-zero correlations calculated under the null between each pair of studies using different methods: using the set of SNPs shared between the two phenotypes with genome-wide significant SNPs and those in the MHC region removed (“Signals removed”), the set of SNPs shared between the two phenotypes, thinned so that all SNPs are at least 0.25 centiMorgans apart (“Thinned”), the set of all SNPs shared between the two phenotypes (“All”), and the intercept of the genetic covariance calculated by LD Score regression B. Bulik-Sullivan et al. (2015); B. K. Bulik-Sullivan et al. (2015) (“LDSC”).

		Signals			
Trait 1	Trait 2	removed	Thinned	All	LDSC
AS	BO	0.192	0.162	0.177	0.196
AS	IS TOAST 1	0.097	0.0946	0.0932	0.0939
AS	IS TOAST 2	0.0995	0.0892	0.0868	0.0998
AS	IS TOAST 3	0.0885	0.0817	0.0804	0.096
AS	MS UK	0.184	0.143	0.147	0.199

(Continued on next page)

(continued from previous page)

Trait 1	Trait 2	Signals		All	LDSC
		removed	Thinned		
AS	PD	0.198	0.185	0.169	0.206
AS	PS	0.198	0.147	0.152	0.206
AS	UC	0.213	0.203	0.209	0.225
BO	IS TOAST 1	0.122	0.134	0.122	0.127
BO	IS TOAST 2	0.133	0.162	0.132	0.135
BO	IS TOAST 3	0.113	0.121	0.113	0.126
BO	MS UK	0.229	0.199	0.226	0.241
BO	PD	0.237	0.227	0.237	0.236
BO	PS	0.284	0.277	0.273	0.294
BO	UC	0.244	0.251	0.244	0.268
BS overall	BS pneumo.	0.619	0.623	0.619	NA
IS TOAST 1	IS TOAST 2	0.121	0.133	0.120	0.127
IS TOAST 1	IS TOAST 3	0.169	0.175	0.168	0.164
IS TOAST 1	MS UK	0.115	0.098	0.111	0.118
IS TOAST 1	PD	0.111	0.118	0.111	0.110
IS TOAST 1	PS	0.129	0.119	0.124	0.127
IS TOAST 1	UC	0.114	0.112	0.115	NA
IS TOAST 2	IS TOAST 3	0.117	0.137	0.118	0.114
IS TOAST 2	MS UK	0.124	0.127	0.121	0.127
IS TOAST 2	PD	0.128	0.117	0.127	0.128
IS TOAST 2	PS	0.134	0.136	0.127	0.139
IS TOAST 2	UC	0.124	0.161	0.123	0.128
IS TOAST 3	MS UK	0.112	0.0866	0.11	0.116

(Continued on next page)

(continued from previous page)

Trait 1	Trait 2	Signals		All	LDSC
		removed	Thinned		
IS TOAST 3	PD	0.112	0.102	0.112	0.104
IS TOAST 3	PS	0.113	0.108	0.107	0.122
IS TOAST 3	UC	0.109	0.131	0.109	NA
MS UK	MS nonUK	0.0529	0.132	0.0927	0.0191
MS UK	PD	0.243	0.245	0.242	0.252
MS UK	PS	0.243	0.218	0.222	0.254
MS UK	UC	0.233	0.224	0.231	0.244
PD	PS	0.249	0.269	0.237	0.253
PD	UC	0.24	0.251	0.239	0.246
PS	UC	0.257	0.247	0.240	0.271
RM reading	RM maths	0.535	0.535	0.535	0.543

Figures

Figure 1:

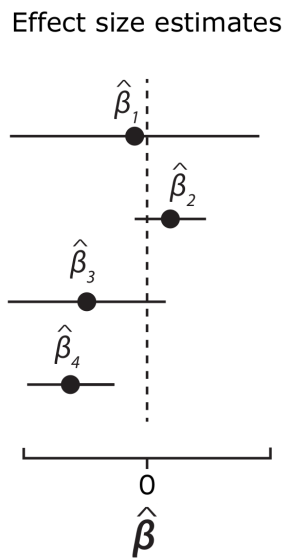


Figure 2:

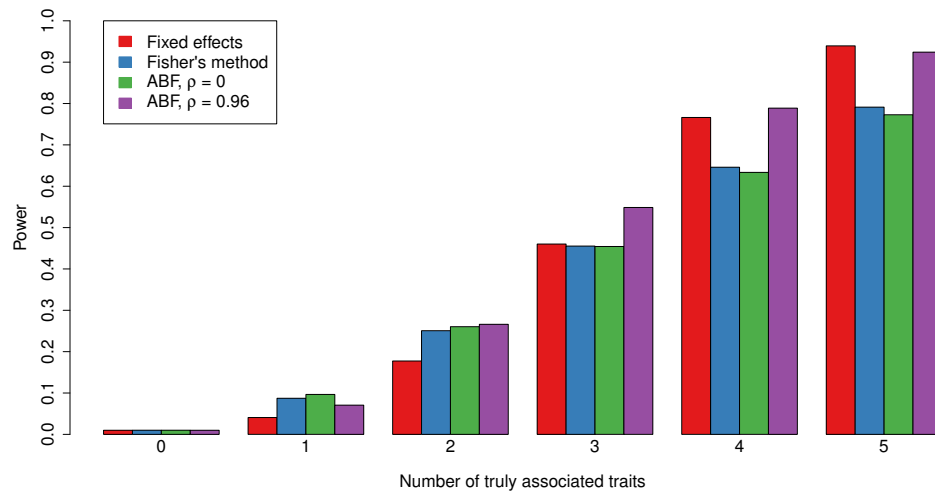


Figure 3:

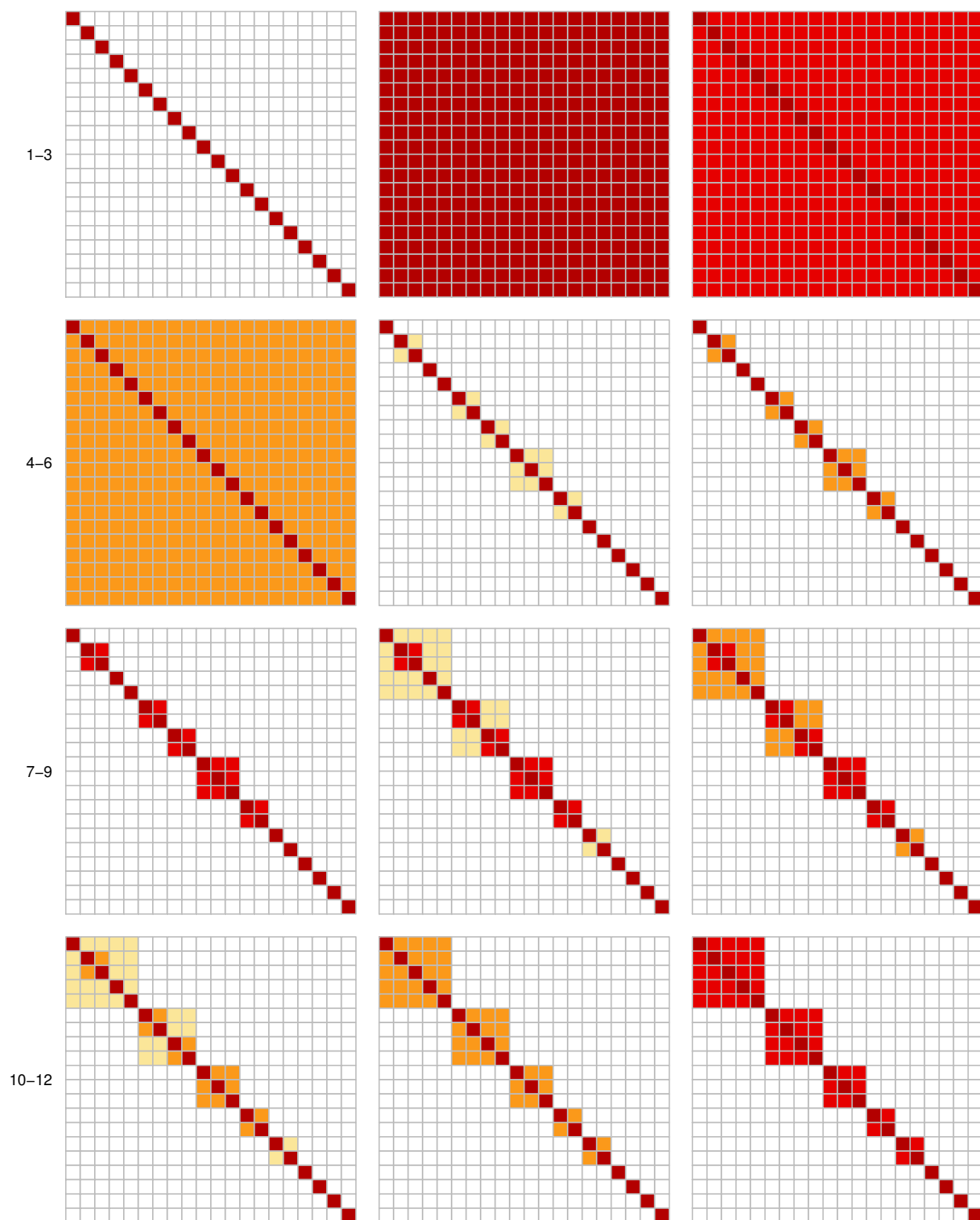


Figure 4:

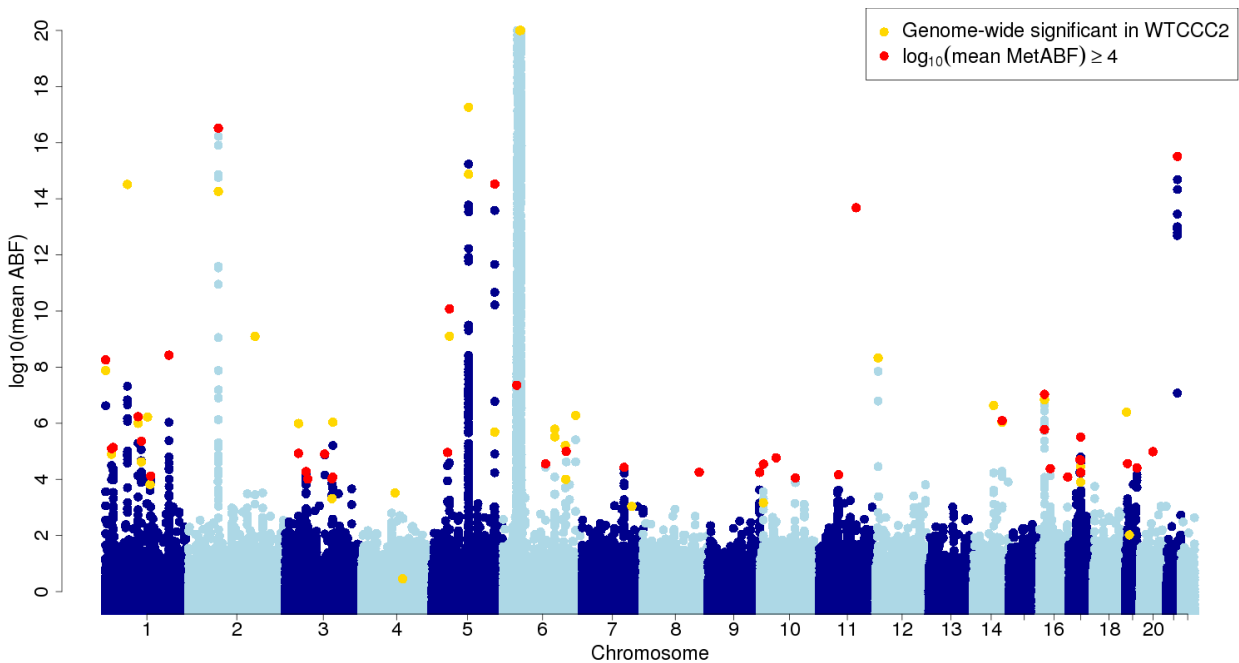


Figure 5:

