

RESEARCH ARTICLE

SMetABF: A rapid algorithm for Bayesian GWAS meta-analysis with a large number of studies included

Jianle Sun¹, Ruiqi Lyu¹, Luoja Deng¹, Qianwen Li¹, Yang Zhao^{2*}, Yue Zhang^{1*}

1 Department of Bioinformatics and Biostatistics, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China, **2** Department of Biostatistics, Nanjing Medical University School of Public Health, Nanjing, Jiangsu, China

* yzhao@njmu.edu.cn (YAZ); yue.zhang@sjtu.edu.cn (YUZ)



OPEN ACCESS

Citation: Sun J, Lyu R, Deng L, Li Q, Zhao Y, Zhang Y (2022) SMetABF: A rapid algorithm for Bayesian GWAS meta-analysis with a large number of studies included. *PLoS Comput Biol* 18(3): e1009948. <https://doi.org/10.1371/journal.pcbi.1009948>

Editor: Carl Herrmann, Heidelberg University, GERMANY

Received: September 8, 2021

Accepted: February 21, 2022

Published: March 14, 2022

Copyright: © 2022 Sun et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All data and code used for running experiments, model fitting, and plotting is available on a GitHub repository at https://github.com/sjl-sjtu/GWAS_meta. The detailed results of application is available at https://figshare.com/articles/dataset/Table_S_zip/19179179.

Funding: YUZ was supported by the National Natural Science Foundation of China (11901387), the National Competition of Health and Longevity of China (JC2021CL029), Three-year Action

Abstract

Bayesian methods are widely used in the GWAS meta-analysis. But the considerable consumption in both computing time and memory space poses great challenges for large-scale meta-analyses. In this research, we propose an algorithm named SMetABF to rapidly obtain the optimal ABF in the GWAS meta-analysis, where shotgun stochastic search (SSS) is introduced to improve the Bayesian GWAS meta-analysis framework, MetABF. Simulation studies confirm that SMetABF performs well in both speed and accuracy, compared to exhaustive methods and MCMC. SMetABF is applied to real GWAS datasets to find several essential loci related to Parkinson's disease (PD) and the results support the underlying relationship between PD and other autoimmune disorders. Developed as an R package and a web tool, SMetABF will become a useful tool to integrate different studies and identify more variants associated with complex traits.

Author summary

MetABF is a Bayesian GWAS meta-analysis framework but the efficiency is restricted by the number of studies included. In this article, we propose SMetABF by introducing SSS, an improved edition of traditional MCMC, to speed the MetABF algorithm. We develop an R package and a web tool based on R Shiny to make SMetABF practical for biomedical research. Comparing with the exhaustive approach and MCMC, we validate the effectiveness of SSS in terms of speed and accuracy through simulations. We applied SMetABF to identify several important variants associated with Parkinson's disease and other autoimmune diseases, and explore the relationship between them. We hope this method can benefit future GWAS meta-analyses, help to identify more risk variants associated with complex traits, and improve the prediction of diseases.

Program of Shanghai Municipality for Strengthening the Construction of Public Health System Big Data and Artificial Intelligence Application (GWV-10.1-XK05), the Non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences (2021-JKCS-028), and Shanghai Jiao Tong University "Jiaotong Star" Plan Medical Engineering Cross Research Project (No:YG2021QN07). YAZ was supported by the National Natural Science Foundation of China (82173620). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

1 Introduction

Genome-wide association study (GWAS), a powerful tool to find out the associations between genetic variations and phenotypes, has received more and more attention in the field of statistical genetics and epidemiology [1]. Numerous variants, typically many common single nucleotide polymorphisms (SNPs), are identified linked with complex traits. However, since single variant's genetic effect on polygenic traits is relatively small, large sample sizes are often required to increase the statistical power [2]. Besides, due to the population stratification and other unobserved confounders, the estimated effect sizes in different studies are divided or even contradictory [3]. Therefore, it has become an increasingly essential challenge to make sufficient use of summary statistics derived from a wide range of studies and to attain pooled statistics through meta-analysis [4], especially when the requirement of data security and privacy makes individual-level data increasingly difficult to obtain [5, 6].

Either the fixed-effect model (FEM) [7] or the random-effect model (REM) [8] is conventionally used to derive a pooled effect size, depending on the assumption on heterogeneity [9]. However, the p -value is dependent on the sample size and minor allele frequency (MAF) of the variant. Therefore, it is improper to use a single threshold [10]. Besides, the relationships between true effect sizes in different studies are hard to be considered in both FEM and REM [11]. On the contrary, it is easy to involve them into the model as a prior in the Bayesian framework. The Bayesian method is also prevalent for researchers for it is more intuitively explainable [12]. Recently, a promising method based on the Bayesian framework named MetABF has already been proposed [13]. With GWAS summary statistics, it could conveniently estimate the pooled associations between multiple traits and genetic variations in different associated models across studies. However, with the rapidly increasing data of GWAS, the method also confronts the challenge of exponential explosion in both time and space consumption. Since it requires traversing 2^n subsets represented by n -dimensional vectors to compute the optimal ABF, the considerable time and memory consumption required makes the computation almost impossible as the number of studies n increases.

In this article, we propose SMetABF, a method based on the Markov chain Monte Carlo (MCMC) method and its extension named shotgun stochastic search (SSS) [14] to speed the process of subset selection. SSS is proved to be superior in speed, accuracy, and stability through simulation. Based on SSS, we introduce SMetABF to obtain the maximum ABF in a large-scale meta-analysis quickly. SMetABF is implemented as an R package and the code is available at https://github.com/sjl-sjtu/GWAS_meta.

2 Method

2.1 Asymptotic Bayes factor

Different from the traditional statistical framework based on the p -value for statistical inference, the Bayes factor (BF) is used in Bayesian statistical framework. BF is defined as the relative size of the likelihood to observe data under the null hypothesis (H_0) or the alternative one (H_1),

$$BF = \frac{P(D|H_1)}{P(D|H_0)} = \frac{\int_{\beta} \int_{\gamma} P(Y|\beta, \gamma) \pi(\beta, \gamma) d\beta d\gamma}{\int_{\gamma} P(Y|\beta = 0, \gamma) \pi(\beta = 0, \gamma) d\gamma},$$

where D stands for the data observed, β is the effect parameter we are interested in, γ is the parameter vector of confounders, and $\pi(\cdot)$ stands for the prior of β and γ . In general, $BF > 1$ means more inclined to accept H_1 , and on the contrary, $0 < BF < 1$ means more inclined to accept H_0 . Since BF is difficult to calculate directly in many studies, an asymptotic Bayes factor

(ABF) is proposed as an alternative [10]. If $P(Y|\beta, \gamma)$ is replaced by the asymptotic distribution $P(\hat{\beta}, \hat{\gamma}|\beta, \gamma)$, and the marginal prior for β instead of the joint prior $\pi(\beta, \gamma)$ is considered, the probability of obtaining the parameter β under a certain hypothesis could replace the probability of observing data Y , written as

$$ABF = \frac{P(\beta|H_1)}{P(\beta|H_0)}.$$

For a study aimed to measure the association between several risk factors and specific outcomes, like GWAS, let $\hat{\beta}$ be the estimated size of the association. It is assumed to obey the normal distribution

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

where β is the true effect size of the variant, and $SE_{\hat{\beta}}$ represents the estimated standard error. The true effect size β is also assumed to follow a normal distribution

$$\beta \sim N(0, \sigma^2),$$

where σ^2 represents the prior variance of the true effect size. When $\sigma = 0$, the distribution of β degenerates to a point, which means $\beta = 0$. In other words, the genetic variant has no effect on the outcome. Under $H_0: \sigma = 0$, ABF can be calculated as

$$ABF = \frac{f(\hat{\beta}; 0, SE_{\hat{\beta}}^2 + \sigma^2)}{f(\hat{\beta}; 0, SE_{\hat{\beta}}^2)}.$$

where $f(x; m, s^2)$ is the probability density of normal distribution $N(m, s^2)$ at x .

Each study included in the meta-analysis provides an estimated effect size, $\hat{\beta}$. When there are n studies in the meta-analysis, let $\hat{\beta}$ be the estimated effect vector

$$\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_n)^T,$$

and $\hat{\beta}$ follows a multivariate normal distribution $N(\beta, V_{\hat{\beta}})$, where β stands for the true effect vector, and $V_{\hat{\beta}}$ represents the covariance matrix of the estimated standard errors, expressed as

$$V_{\hat{\beta}} = \begin{pmatrix} SE_1^2 & r_{1,2}SE_1SE_2 & \cdots & r_{1,n}SE_1SE_n \\ r_{1,2}SE_1SE_2 & SE_2^2 & \cdots & r_{2,n}SE_2SE_n \\ \vdots & \vdots & \ddots & \vdots \\ r_{1,n}SE_1SE_n & r_{2,n}SE_2SE_n & \cdots & SE_n^2 \end{pmatrix},$$

in which SE_i is the standard error of $\hat{\beta}_i$ in the i -th study, and $r_{i,j}$ is the correlation between the i -th and j -th studies. For each study, the prior effect size is σ_i , and the prior correlation coefficient between two studies is $\rho_{i,j}$, then the prior matrix Σ is

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho_{1,2}\sigma_1\sigma_2 & \cdots & \rho_{1,n}\sigma_1\sigma_n \\ \rho_{1,2}\sigma_1\sigma_2 & \sigma_2^2 & \cdots & \rho_{1,n}\sigma_2\sigma_n \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1,n}\sigma_1\sigma_n & \rho_{1,n}\sigma_2\sigma_n & \cdots & \sigma_n^2 \end{pmatrix}.$$

Table 1. Different models of prior across studies [13].

parameter	name	set
σ	null model	$\sigma^2 = 0$
	complete model	$\sigma^2 > 0$
	subset model	$\sigma^2 > 0$ for subset $I \subset \{1, \dots, n\}$
ρ	fixed effect	$\rho = 1$
	independent effect	$\rho = 0$
	correlated effect	$0 < \rho < 1$

<https://doi.org/10.1371/journal.pcbi.1009948.t001>

With the estimated effect vector and covariance matrix of the estimated standard errors, ABF for meta-analysis could be calculated as

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

Similarly, H_0 is defined that Σ equals to the zero matrix.

2.2 Prior

The assumption on the heterogeneity among studies is critical in prior selection. Table 1 provides different models for prior under the assumption that σ and ρ remain the same for all studies included in the meta-analysis.

Since both the null model and the complete model can be regarded as special cases of the subset model, the subset model is adopted in the meta-analysis. But it also brings a tricky issue to determine the optimal subsets. It is preferred to get a higher ABF score in the meta-analysis, for it means that H_0 is harder to be accepted and the probability of the type II error decreases. In other words, it increases the statistical power, which equals one minus the probability of the type II error.

2.3 Model selection

2.3.1 Subset-exhaustive. The former study [13] performs model selection by traversing subsets to select the highest ABF, named the subset-exhaustive method (EXH). For a meta-analysis including n studies, there are in total 2^n different subsets. It requires taking all these subsets as a prior one by one to calculate ABF, and then find the optimal one. The time consumed will explode exponentially as the number of studies included increases. At the same time, the memory required to store all subsets (2^n n -dimensional 0-1 vectors) has also expanded dramatically. Therefore, it becomes quite essential to introduce a method to get higher ABF quickly.

2.3.2 MCMC. A commonly used method to quickly find the best subset is the Markov chain Monte Carlo method (MCMC). Here MC³ algorithm [15, 16] is used to define the transition function and the Metropolis-Hastings algorithm is used for sampling. The whole process is carried out as Algorithm 1.

Algorithm 1 MCMC Pseudocode

Require: Ω : Universe of subsets.
Ensure: x : Sample with stable distribution.
 $x_0 \leftarrow$ random sample from Ω
for $t = 0 \dots T$ **do**

```

Update  $nb d(x_t)$ 
 $\Pi(x_t) \leftarrow \{P_{i \in nb d(x_t)} = \frac{1}{|nb d(x_t)|}, P_{i \notin nb d(x_t)} = 0\}$ 
Generate  $y$  based on  $\Pi(x_t)$ 
 $h = \min\{1, \frac{ABF(y)}{ABF(x_{t-1})}\}$ 
Generate  $u$  from uniform distribution  $U(0, 1)$ 
if  $u < h$  then
     $x_{t+1} = y$ 
else
     $x_{t+1} = x_t$ 
end if
end for

```

▷ Neighborhood
 ▷ Transition Probability
 ▷ Alternative Model
 ▷ Discriminant Function

1. Randomly select a subset as the initial prior model x_0 , and calculate the ABF.
2. For the current model x_t , define the neighborhood as a set constituted of subsets formed by adding or deleting an element from the current subset, as well as the current model itself. The proposal distribution is defined by equalizing the sampling probability of all models in the neighborhood. In other words, the sampling probability of all models in the neighborhood remains equal, and the transition probability of all models outside the neighborhood is 0. Since each model has the same size of the neighborhood, the proposal distribution is symmetric.
3. Generate the alternative prior model y according to the transition probability, and then calculate the ABF. The discriminant function is defined as $h = \min\{1, \frac{ABF(y)}{ABF(x_{t-1})}\}$, where $ABF(x)$ represents the ABF value with subset x as the prior.
4. Generate a random number u that follows the uniform distribution $U(0, 1)$. If $u < h$, accept y as a new step of x_{t+1} , and otherwise, $x_{t+1} = x_t$.
5. Repeat steps 2-4 until the maximum number of iterations or stable distribution is reached.

The first half of the entire iterative sequence is used for the warm-up and the second for the final sampling.

2.3.3 Shotgun stochastic search. Here we introduce an extension of MCMC for variable selection named shotgun stochastic search (SSS) [14]. It can be used to fast detect the optimal ABF following the procedures as below (see Algorithm 2):

1. Let Γ donate a set containing up to B optimal models. Randomly select the initial model x_0 , set $\Gamma = \{x_0\}$, and calculate the score of the model $S(x) = ABF(x)$.
2. For the current model x_t , define models that add or delete or replace an element from the current subset to constitute the sets Γ^+ , Γ^- , and Γ° , respectively, and then define the neighborhood

$$nb d(x_t) = \Gamma^+ \cup \Gamma^- \cup \Gamma^\circ.$$

Then update $\Gamma = \Gamma \cup nb d(x_t)$. If $|\Gamma| > B$, remove $(|\Gamma| - B)$ models with the lowest scores.

3. Sample x^+ , x^- , and x° from Γ^+ , Γ^- , and Γ° , with the score $S(x)$ as sampling weight, respectively.
4. Then take a sample from x^+ , x^- , and x° , with the score $S(x)$ as sampling weight, and let the sample be the new model x_{t+1} .
5. Repeat steps 2-4 until the maximum number of iterations is reached.

Algorithm 2 SSS Pseudocode

Require: Ω : Universe of subsets.
Ensure: x : Sample with stable distribution.
 $x_0 \leftarrow$ random sample from Ω
 $\Gamma = \{x_0\}$
 $S(x) \leftarrow ABF(x)$
for $t = 0 \dots T$ **do**
 Constitute $\Gamma^+, \Gamma^-, \Gamma^\circ$
 $nbid(x_t) = \Gamma^+ \cup \Gamma^- \cup \Gamma^\circ$
 Update $\Gamma = \Gamma \cup nbid(x_t)$
 if $|\Gamma| > B$ **then**
 Remove $(|\Gamma| - B)$ models with lowest S .
 end if
 Sample x^+ from Γ^+ , weight = $S(x)$
 Sample x^- from Γ^- , weight = $S(x)$
 Sample x° from Γ° , weight = $S(x)$
 Sample x_{t+1} from x^+, x^-, x° , weight = $S(x)$
 if x_{t+1} satisfies stable distribution **then**
 break
 end if
end for

The former study [13] has provided R code for EXH. Here R functions for meta-analysis by MCMC and SSS are constructed.

3 Simulation

3.1 The construction of simulated datasets

Several parameters are given to build the simulated datasets, including the incidence of the disease in the population (p), the frequency of the major allele of the studied variant (f , which equals to 1-MAF under the assumption that there are only two alleles in the SNP), the effect size (odds ratio, OR), and the sample size of both case and control groups (which is assumed to be the same, n). For example, suppose A is the risk allele while G is the non-risk allele. If the dominant model is applied, both AA and AG can be considered as equivalent risk genotypes while GG is non-risk. Suppose baseline effect is α , the increased effect on prevalence by risk genotype is θ , then

$$P(D|GG) = \alpha,$$

$$P(D|AA + AG) = \alpha(1 + \theta),$$

where D represents the outcome (disease). Then we can get

$$OR = \frac{P(D|GG)}{P(\bar{D}|GG)} \cdot \frac{P(\bar{D}|AA + AG)}{P(D|AA + AG)} = \frac{1 - \alpha}{\alpha} \cdot \frac{\alpha(1 + \theta)}{1 - \alpha(1 + \theta)},$$

and

$$p = f^2\alpha + (1 - f^2)\alpha(1 + \theta).$$

Then α and θ can be calculated. According to the Bayes Theorem, the probability of risk and non-risk genotypes in the case and control groups can be calculated.

$$P(GG|D) = \frac{P(D|GG)P(GG)}{P(D)} = \frac{f^2}{p},$$

$$P(GG|\bar{D}) = \frac{(1 - \alpha)f^2}{1 - p}$$

And then, the simulated genotypes in both the case and control groups could be randomly generated under binomial distribution. The estimated \widehat{OR} can be calculated. The effect size is defined as $\beta = \ln OR$, and similarly, $\hat{\beta} = \ln \widehat{OR}$. The standard error can be estimated from the contingency table, as $SE_{\hat{\beta}} = \sqrt{\frac{1}{n_{11}} \cdot \frac{1}{n_{12}} \cdot \frac{1}{n_{21}} \cdot \frac{1}{n_{22}}}$.

Suppose there are N studies included in the meta-analysis. For each study, the true effect OR_i obeys the normal distribution $N(OR, SE^2)$. The sample size n_i in each study is generated as a random integer in a given range, and p and f remain the same for all studies. Through the process above, $\hat{\beta}_i$ and $SE_{\hat{\beta}_i}$ of each study can be estimated.

3.2 Results

The ABFs calculated under different true ORs are shown in Fig 1. The overall trends of the ABF obtained by EXH, MCMC, and SSS remain consistent, corresponding to the p -value obtained by the traditional method. When the true OR approaches 1, the p -value increases, while the ABF value decreases to 0. However, when the sample size of the study included is small (Fig 1A), the change of p -value will be unstable if true OR is near to 1, which will affect the analysis. Besides, the ABF calculated by SSS almost coincides with the optimal ABF curve obtained by EXH, which shows the validity of SMetABF.

Figs 2 and 3 show the performance of each algorithm in accuracy and speed under different priors and iterations, respectively. Since in SSS, ABF is calculated under all models in the neighborhood in one iteration, much more models will be calculated by the SSS with the same number of iterations. Therefore, the number of iterations of SSS is set to be 100, 200, 500, 1000 and 2000; while that of MCMC is set to be 1000, 5000, 10000 and 20000. To compare the averaged ABF and time consumed, the algorithm is repeated 100 times under each condition. The

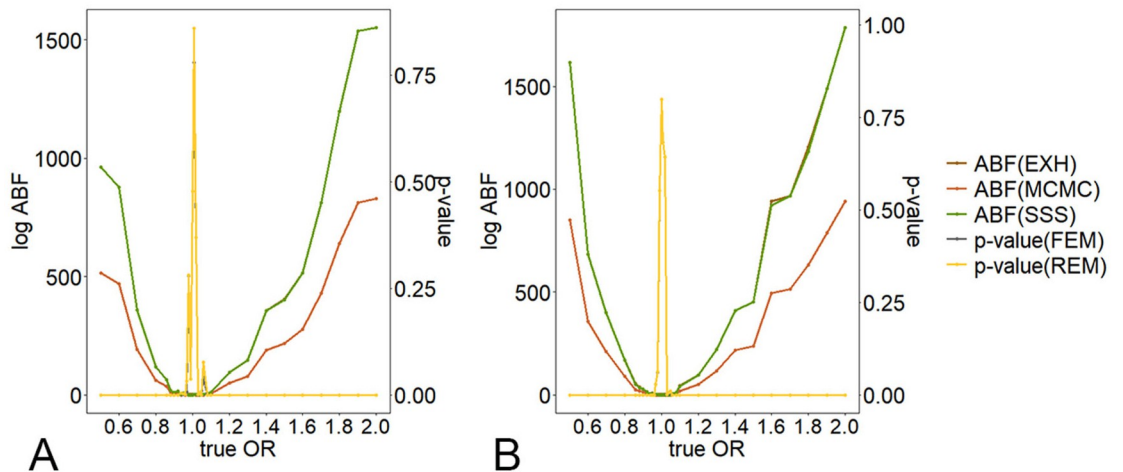


Fig 1. The comparison under various true ORs. Curves representing ABF (EXH) and ABF (SSS) are nearly coincide. Curves representing p -value (FEM) and p -value (REM) are nearly coincide as well. The parameters are set as follows: $p = 0.05, f = 0.8$ (which equally means $MAF = 0.2$), the number of studies included (N) is set to be 20 (Fig 1A) and 25 (Fig 1B) respectively. For the i -th study, $OR_i \sim N(OR, 0.01)$, the sample size n_i is sampled from 100 to 2000 and 100 to 5000, respectively.

<https://doi.org/10.1371/journal.pcbi.1009948.g001>

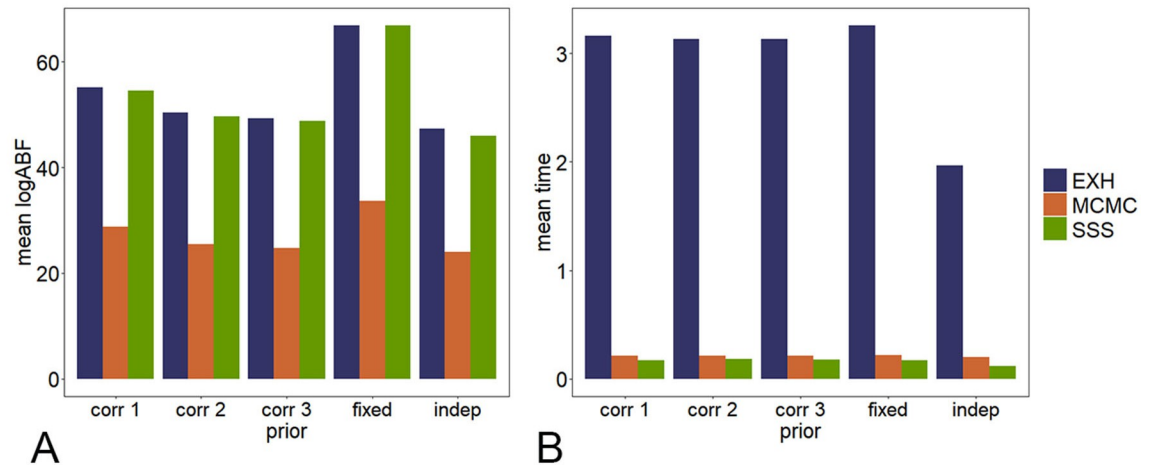


Fig 2. The comparison in accuracy and speed of the three algorithms under different priors. Priors setting: corr 1 (correlated model, $\sigma = 0.5, \rho = 0.7$); corr 2 (correlated model, $\sigma = 0.5, \rho = 0.3$); corr 3 (correlated model, $\sigma = 0.8, \rho = 0.7$); fixed (fixed model, $\sigma = 0.5$); indep (independent model, $\sigma = 0.5$).

<https://doi.org/10.1371/journal.pcbi.1009948.g002>

SSS algorithm can reach the maximum ABF in a short time with a small number of iterations. On the contrary, the MCMC algorithm can hardly find the maximum ABF in even longer time.

When repeating 100 times of MCMC (10000 and 20000 iterations) and SSS (500 and 1000 iterations), as shown in Fig 4, the ABF values obtained by MCMC are relatively small, while the results of SSS are relatively stable, very close to the maximum ABF.

4 Application

4.1 Meta-analysis on the variants related to PD and other autoimmune disorders

Here an application is performed to measure the risk variants associated with Parkinson’s disease (PD), a common chronic neurodegenerative disease among the elderly population. Its

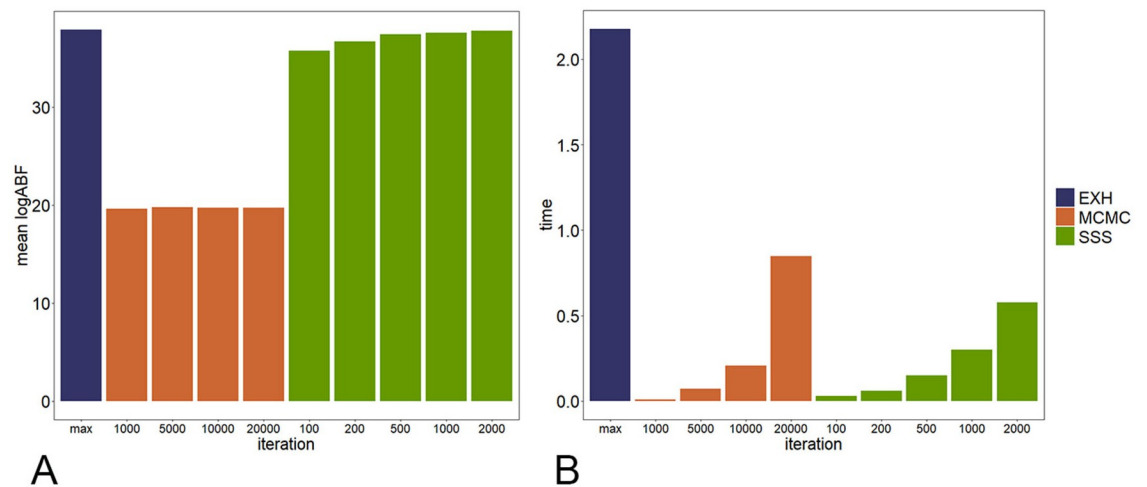


Fig 3. The comparison in accuracy and speed of the three algorithms under different iterations.

<https://doi.org/10.1371/journal.pcbi.1009948.g003>

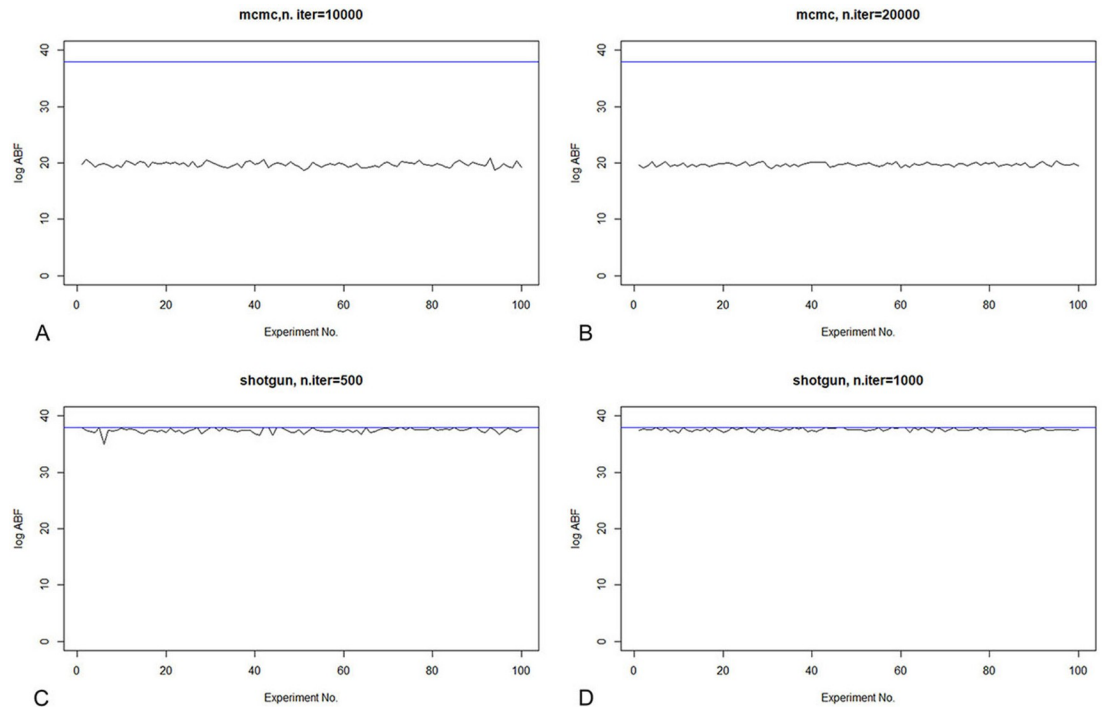


Fig 4. The comparison in stability of MCMC and SSS.

<https://doi.org/10.1371/journal.pcbi.1009948.g004>

common clinical manifestations include tremors, slow movement, and disorders in balance and movement posture. PD has been reported to be associated with both genetic variations [17] and environmental factors like personal lifestyles such as smoking and drinking [18, 19], but the detailed mechanism remains unclear. Recent studies discuss the potential relationship between PD and autoimmune disorders [20]. To explore the underlying relationships, we conduct a GWAS meta-analysis across PD and three common autoimmune disorders: inflammatory bowel disease, multiple sclerosis, and systemic sclerosis.

Through the websites accommodating GWAS datasets, including DistiLD [21] (<http://distild.jensenlab.org>), Open Targets Genetics [22] (<https://genetics.opentargets.org/>) and GWAS Catalog [23] (<https://www.ebi.ac.uk/gwas/>), 59 studies in which the summary statistics ($\hat{\beta}$ and $SE_{\hat{\beta}}$) are provided or can be calculated are included in this application. Tables 2 and 3 show detailed information about the studies included. A pure meta-analysis across 29 studies on PD is conducted firstly, and then all 59 studies are analyzed jointly to obtain a mixed association pattern. The effects of over 10 million variants are assessed through parallel computing. The Manhattan plots for both the pure pattern and the mixed pattern are shown in Fig 5.

We can find PD is highly associated with several loci located within gene *SCNA* on chromosomes 4. A peak also appears on chromosome 17, around gene *MAPT*, *KANSL1*, and *NSF*. When other autoimmune disorders are included in the meta-analysis, the peaks appear on chromosomes 1 and 6 in the mixed pattern. Some significant variants can also be found on chromosome 4. Table 4 shows several SNPs identified in the analysis. Detailed results are available at https://figshare.com/articles/dataset/Table_S_zip/19179179.

The results supports the previous reports on several essential loci related to PD, such as *SCNA*, *MAPT* and *KANSL1* [17]. Additionally, some degree of underlying relationships between PD and other autoimmune disorders are revealed by comparing the mixed pattern to

Table 2. The information of studies on PD included in the application.

First author	Published year	Ancestry	Sample Size (cases/controls)
Maraganore DM [24]	2005	European (US)	332/332
Pankratz N [25]	2008	European (US)	857/867
Satake W [26]	2009	East Asian (Japan)	2,011/18,381
Simn-Snchez J [27]	2009	European	1,713/3,978
Sutherland GT [28]	2009	European (Australia)	331/296
Edwards TL [29]	2010	European (US)	1,752/1,745
Hamza TH [30]	2010	European (US)	2,000/1,986
Saad M [31]	2010	European	4,271/9,048
Do CB [32]	2011	European	3,426/29,624
Liu X [33]	2011	Ashkenazi Jewish	2,050/1,836
Nalls MA [34]	2011	European	5,333/12,019
Spencer C [35]	2011	European (UK)	1,705/5,175
Simn-Snchez J [36]	2011	European (Dutch)	772/2,024
Lill CM [37]	2012	World	16,452/48,810
Nalls MA [38]	2014	European (US)	13,708/95,282
Hill-Burns EM [39]	2014	European (US)	1,986/2,000
Foo JN [40]	2017	East Asian	5,125/17,604
Chang D [41]	2017	World	26,035/403,190
Bandres-Ciga S [42]	2019	European (Spain)	4,639/2,949
Blauwendraat C [43]	2019	World	17,996 cases
Nalls MA [44]	2019	European	33,674/449,056
Blauwendraat C [45]	2020	European	1,588/7,584
Alfradique-Dunham I [46]	2021	European	1,570/1,259
Backman JD [47]	2021	European (UK)	828/330,926
Jiang L [48]	2021	European (UK)	294/456,054
Rodrigo LM [49]	2021	European	5,167/5,366
Smeland OB [50]	2021	European	20,184/975,838
Sakaue S [51]	2021	European & East Asian (Japan)	2,978/653,168
CIDR dataset ¹	-	World	1,048/943

¹ Details of the dataset can be found at https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000126.v2.p1

<https://doi.org/10.1371/journal.pcbi.1009948.t002>

the pure pattern, for they have some shared risk variants. For example, variants on *BTNL2* have high ABF in both the pure pattern ($\sim 10^{17}$) and the mixed pattern ($\sim 10^{127}$), and the subsets indicates that *BTNL2* is associated with all four disorders. For the top 4 variants on *BTNL2* in the mixed pattern, on average 66.7% studies on PD, 65.7% studies on inflammatory bowel disease, 100% studies on multiple sclerosis, and 62.5% studies on systemic sclerosis are included in the final subsets to calculate ABF. However, although variants on *SCNA* also have a high ABF value in both pure pattern and mixed pattern, we found most of the studies in the final subsets to calculate ABF are from those studies on PD and the ABF values remain similar ($\sim 10^{135}$) in both patterns. In other words, *SCNA* has weaker associations with other autoimmune disorders. The same is also true for those variants on *MAPT*. These two loci may relate more to the diseases in nervous system instead of autoimmune disorders. On the contrary, variants on *KANSL1*, reported as a factor in the immune system [71], show associations with both PD and other autoimmune disorders. In brief, the results of the meta-analysis indicate the presence of potential biological pathways and functional interactions between PD and

Table 3. The information of studies on other autoimmune disorders included in the application.

First author	Published year	Ancestry	Sample Size (cases/controls)
inflammatory bowel disease			
Anderson CA [52]	2011	European	6,687/19,718
Jostins L [53]	2012	European	12,924/21,442
Juli A [54]	2014	European	7,483/21,211
Liu JZ [55]	2015	European	25,273/26,715
Liu JZ [55]	2015	Iranian	548/342
Liu JZ [55]	2015	Indian	1,423/990
Liu JZ [55]	2015	East Asian	2,824/3,719
Ostrowski J [56]	2016	European (Poland)	1,118/582
Yang SK [57]	2016	East Asian (Korea)	1,505/4,041
de Lange KM [58]	2017	European	25,042/34,915
Backman JD [47]	2021	European (UK)	5,650/298,738
Dnerta HM [59]	2021	UK	4,101/480,497
Glanville KP [60]	2021	European (UK)	5,105/324,074
Jiang L [48] (Crohn's disease)	2021	European (UK)	1,342/455,006
Jiang L [48] (ulcerative colitis)	2021	European (UK)	2,569/453,779
Sakaue S [51]	2021	European & East Asian (Japan)	5,685/590,936
Wu Y [61]	2021	European	7,045/449,282
multiple sclerosis			
Hafler DA [62]	2007	European	931/2,431
De Jager PL [63]	2009	European	2,624/7,220
Patsopoulos NA [64]	2011	European	5,545/12,153
Sawcer S [65]	2011	European	9,772/16,849
Beecham AH [66]	2013	European	14,498/24,091
Andlauer TF [67]	2016	European (German)	4,888/10,395
IMSGC [68]	2019	World	14,802/26,703
Backman JD [47]	2021	European (UK)	1,596/330,158
Glanville KP [60]	2021	European (UK)	1,683/324,074
Jiang L [48]	2021	European (UK)	775/455,573
systemic sclerosis			
Mayes MD [69]	2014	European	1,833/3,466
Lpez-Isac E [70]	2019	European	9,095/17,584
Jiang L [48]	2021	European (UK)	104/456,244

IMSGC: International Multiple Sclerosis Genetics Consortium

<https://doi.org/10.1371/journal.pcbi.1009948.t003>

autoimmune disorders. Tools like GESLM can use the shared variants to further identify causal variants [72].

4.2 Software

We implement all the algorithms as an R package named GWASmeta. Besides, to help researchers to use SMetABF to quickly find key SNPs, we develop a web tool based on R Shiny as well. The requirements of the file uploaded can be found in the website. Multiple variants can be analyzed at once. This tool is accessible at <https://sunjianle-sjtu.shinyapps.io/analycode>.

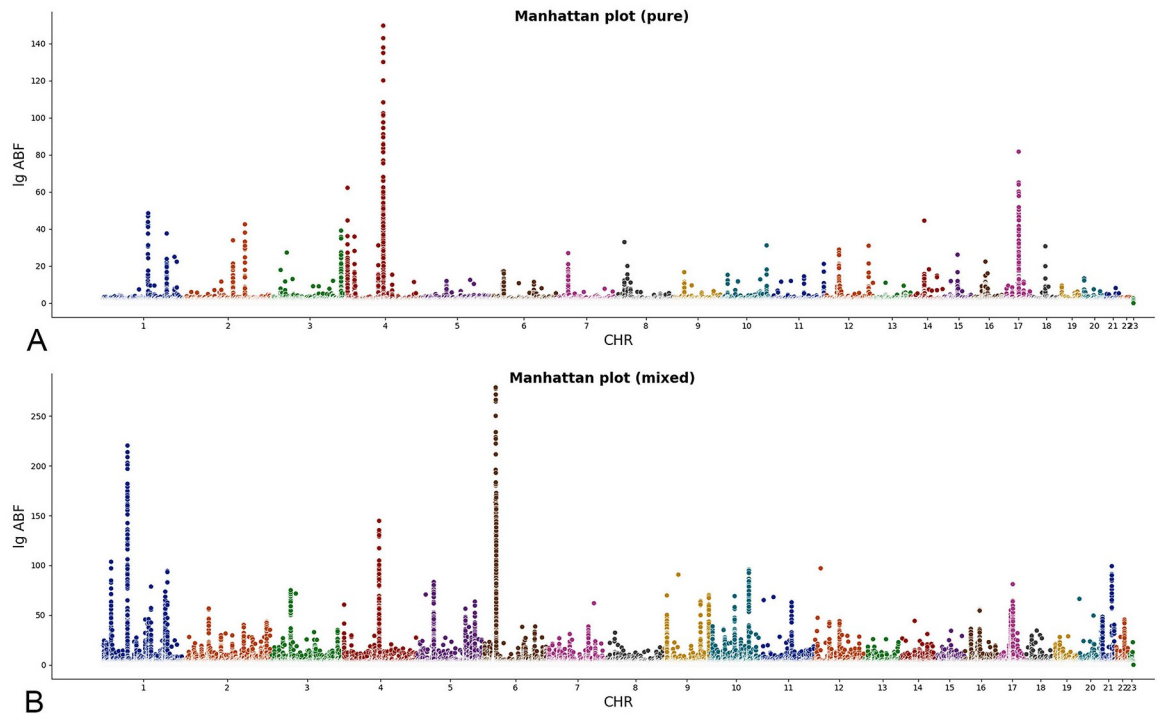


Fig 5. Manhattan plots. A. Results of pure meta-analysis, which includes 29 studies on PD. B. Results of mixed meta-analysis, which includes 59 studies on PD, inflammatory bowel disease (including its two subtypes: Crohn’s disease and ulcerative colitis), multiple sclerosis, and systemic sclerosis. \lg ABF: \log_{10} ABF.

<https://doi.org/10.1371/journal.pcbi.1009948.g005>

5 Discussion

Meta-analysis has been widely conducted on GWAS data to discover essential loci associated with some complex genetic diseases during recent years [73–75], satisfying the requirements of large sample size in GWAS. However, the traditional p -value method used in meta-analysis is facing increasing criticisms. For example, it is not proper to use a single threshold since p -value is dependent on the MAF and sample size [10]. Moreover, the sophisticated relationships

Table 4. Some key variants identified in analysis.

	CHR	Gene	SNP
pure pattern	4	SCNA	rs2736990, rs356165, rs356203, rs356168, rs356200, rs2737029
	17	MAPT	rs17649553, rs62056850, rs8070723, rs62062279
	17	KANSL1	rs2532276, rs2696664, rs56406462, rs2532275, rs2532278, rs2532281, rs2696658
	17	NSF	rs199447, rs199451, rs169201
	1	ASH1L	rs71628662, rs145330152, rs12734374, rs145331499
mixed pattern	6	BTNL2	rs3763309, rs3763312, rs3793127, rs9268491, rs3817963
	6	TSBP1-ASI	rs3130320, rs926070
	6	TSBP1	rs2395150, rs6904636, rs3129908, rs502626, rs477005
	6	BTNL2	rs3129954, rs3129955, rs2076530, rs3817963
	6	BAG6	rs3130050, rs2242656, rs3130617, rs1077394
	1	IL23R	rs11465804, rs80174646, rs75328060, rs11805303, rs7539625, rs1004819
4	SCNA	rs2736990, rs356165, rs356203, rs356168, rs356200, rs2737029	

<https://doi.org/10.1371/journal.pcbi.1009948.t004>

between different studies are tricky to deal with in traditional methods. FEM relies on the assumption that all studies in the meta-analysis share a common true effect size. The true effect size is allowed to vary in different studies in REM, but the detail information is hard to be included in the model. And the test for heterogeneity to determine whether FEM or REM should be used is often regarded poor in power [11]. Instead, the structure among different studies can be easily integrated into the Bayesian model as a prior. The BF compares the relative size between $P(H_0|Y)$ and $P(H_1|Y)$, and therefore is a better alternative to the p -value. Based on the Bayesian framework, a useful statistical model named MetABF has been proposed, which could easily measure the associations between multiple phenotypes and variants at the same time using GWAS summary statistics but confronts challenges in computation.

In this article, we propose SMetABF, an improved tool to attain the optimal ABF in a large-scale meta-analysis efficiently. Through simulation, we confirm that SSS is superior to MCMC in terms of speed, accuracy, and stability. To a certain extent, our improvements effectively overcome the calculation problems due to the increase in the number of studies included. We perform an application to PD and other autoimmune disorders, illustrating the effectiveness of SMetABF. With more research conducted on various traits among a larger population and the increasing accumulations of GWAS summary statistics, the large-scale multi-phenotypic meta-analyses will be possible through SMetABF. Another possible application is to analyze the effect size across different variants in one study, where σ_i represents the prior variation of the i -th variant on the outcome, and $\rho_{i,j}$ stands for the linkages between different variants. Furthermore, since many traits related to some complex diseases are correlated, it is necessary to consider the effect of multiple loci on the outcome across a large number of studies simultaneously [76]. In this case, the prior correlation matrix Σ will transform to a three-dimensional array, which will bring more challenges in computation.

The method still confronts many challenges. The choice of prior parameters is an example. Sensitive analysis reveals that different values of σ and ρ will affect the ABF values but seem not to change the relative effect size between different variants. Besides, the considerable size of human genome still brings challenges in computation.

The pooled statistics derived through meta-analysis can be further used for other post-GWAS analysis, for example, to identify causal genes through statistical fine-mapping [77] or to infer the causal relationships between traits by Mendelian randomization [78]. GWAS summary statistics from different studies can be conveniently integrated to a powerful pooled statistic by SMetABF. We believe the method will benefit to the integration of previous studies and help to reveal the genetic mechanisms of complex diseases.

Author Contributions

Conceptualization: Jianle Sun, Yue Zhang.

Data curation: Jianle Sun.

Formal analysis: Jianle Sun.

Funding acquisition: Yue Zhang.

Methodology: Jianle Sun.

Software: Jianle Sun.

Supervision: Yang Zhao, Yue Zhang.

Visualization: Jianle Sun.

Writing – original draft: Jianle Sun.

Writing – review & editing: Ruiqi Lyu, Luoja Deng, Qianwen Li, Yang Zhao, Yue Zhang.

References

1. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 years of GWAS discovery: biology, function, and translation. *The American Journal of Human Genetics*. 2017; 101(1):5–22. <https://doi.org/10.1016/j.ajhg.2017.06.005> PMID: 28686856
2. Evangelou E, Ioannidis JP. Meta-analysis methods for genome-wide association studies and beyond. *Nature Reviews Genetics*. 2013; 14(6):379–389. <https://doi.org/10.1038/nrg3472> PMID: 23657481
3. McClellan J, King MC. Genetic heterogeneity in human disease. *Cell*. 2010; 141(2):210–217. <https://doi.org/10.1016/j.cell.2010.03.032> PMID: 20403315
4. Pasaniuc B, Price AL. Dissecting the genetics of complex traits using summary association statistics. *Nature Reviews Genetics*. 2017; 18(2):117–127. <https://doi.org/10.1038/nrg.2016.142> PMID: 27840428
5. Almadhoun N, Ayday E, Ulusoy Ö. Differential privacy under dependent tuples the case of genomic privacy. *Bioinformatics*. 2020; 36(6):1696–1703. PMID: 31702787
6. Mohammed Yakubu A, Chen YPP. Ensuring privacy and security of genomic data and functionalities. *Briefings in Bioinformatics*. 2019; 21(2):511–526. <https://doi.org/10.1093/bib/bbz013>
7. Pfeiffer R, Gail M, Pee D. On Combining Data From Genome-Wide Association Studies to Discover Disease-Associated SNPs. *Statistical Science*. 2009; 24:547–560. <https://doi.org/10.1214/09-STS286>
8. Han B, Eskin E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *American Journal of Human Genetics*. 2011; 88 5:586–98. <https://doi.org/10.1016/j.ajhg.2011.04.014> PMID: 21565292
9. Evangelou E, Ioannidis J. Meta-analysis methods for genome-wide association studies and beyond. *Nature Reviews Genetics*. 2013; 14:379–389. <https://doi.org/10.1038/nrg3472> PMID: 23657481
10. Wakefield J. Bayes factors for genome-wide association studies: comparison with P-values. *Genetic Epidemiology*. 2009; 33(1):79–86. <https://doi.org/10.1002/gepi.20359> PMID: 18642345
11. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*. 2010; 1(2):97–111. <https://doi.org/10.1002/jrsm.12> PMID: 26061376
12. Kruschke J, Liddell T. The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*. 2018; 25:178–206. <https://doi.org/10.3758/s13423-016-1221-4> PMID: 28176294
13. Trochet H, Pirinen M, Band G, Jostins L, McVean G, Spencer CCA. Bayesian meta-analysis across genome-wide association studies of diverse phenotypes. *Genetic Epidemiology*. 2019; 43(5):532–547. <https://doi.org/10.1002/gepi.22202> PMID: 30920090
14. Hans C, Dobra A, West M. Shotgun Stochastic search for “Large p” regression. *Journal of the American Statistical Association*. 2007; 102(478):507–516. <https://doi.org/10.1198/016214507000000121>
15. Raftery A, Madigan D, Hoeting J. Bayesian Model Averaging for Linear Regression Models. *Journal of the American Statistical Association*. 1997; 92:179–191. <https://doi.org/10.1080/01621459.1997.10473615>
16. Lamniso D, Griffin J, Steel M. Adaptive MC3 and Gibbs algorithms for Bayesian Model Averaging in Linear Regression Models. *arXiv: Computation*. 2013;.
17. Deng H, Wang P, Jankovic J. The genetics of Parkinson disease. *Ageing research reviews*. 2018; 42:72–85. <https://doi.org/10.1016/j.arr.2017.12.007> PMID: 29288112
18. Kim R, Yoo D, Jung YJ, Han K, Lee JY. Sex differences in smoking, alcohol consumption, and risk of Parkinson’s disease: A nationwide cohort study. *Parkinsonism & Related Disorders*. 2020; 71:60–65. <https://doi.org/10.1016/j.parkreldis.2019.12.006> PMID: 31882374
19. Paul KC, Chuang YH, Shih IF, Keener A, Bordelon Y, Bronstein JM, et al. The association between life-style factors and Parkinson’s disease progression and mortality. *Movement Disorders*. 2019; 34(1):58–66. <https://doi.org/10.1002/mds.27577> PMID: 30653734
20. McFarland NR, McFarland KN, Golde TE. Parkinson Disease and Autoimmune Disorders What Can We Learn From Genome-wide Pleiotropy? *JAMA neurology*. 2017; 74(7):769–770. <https://doi.org/10.1001/jamaneurol.2017.0843> PMID: 28586798
21. Pallejà A, Horn H, Eliasson S, Jensen L. DistalD Database: diseases and traits in linkage disequilibrium blocks. *Nucleic Acids Research*. 2012; 40:D1036–D1040. <https://doi.org/10.1093/nar/gkr899> PMID: 22058129

22. Ghousaini M, Mountjoy E, Carmona M, Peat G, Schmidt EM, Hercules A, et al. Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic acids research*. 2021; 49(D1):D1311–D1320. <https://doi.org/10.1093/nar/gkaa840> PMID: 33045747
23. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic acids research*. 2017; 45(D1):D896–D901. <https://doi.org/10.1093/nar/gkw1133> PMID: 27899670
24. Maraganore DM, De Andrade M, Lesnick TG, Strain KJ, Farrer MJ, Rocca WA, et al. High-resolution whole-genome association study of Parkinson disease. *The American Journal of Human Genetics*. 2005; 77(5):685–693. <https://doi.org/10.1086/496902> PMID: 16252231
25. Pankratz N, Wilk JB, Latourelle JC, DeStefano AL, Halter C, Pugh EW, et al. Genomewide association study for susceptibility genes contributing to familial Parkinson disease. *Human Genetics*. 2009; 124(6):593–605. <https://doi.org/10.1007/s00439-008-0582-9> PMID: 18985386
26. Satake W, Nakabayashi Y, Mizuta I, Hirota Y, Ito C, Kubo M, et al. Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. *Nature Genetics*. 2009; 41(12):1303–1307. <https://doi.org/10.1038/ng.485> PMID: 19915576
27. Simon-Sanchez J, Schulte C, Bras JM, Sharma M, Gibbs JR, Berg D, et al. Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nature Genetics*. 2009; 41(12):1308–1312. <https://doi.org/10.1038/ng.487> PMID: 19915575
28. Sutherland GT, Halliday GM, Silburn PA, Mastaglia FL, Rowe DB, Boyle RS, et al. Do polymorphisms in the familial Parkinsonism genes contribute to risk for sporadic Parkinson's disease? *Movement disorders: official journal of the Movement Disorder Society*. 2009; 24(6):833–838. <https://doi.org/10.1002/mds.22214> PMID: 19224617
29. Edwards TL, Scott WK, Almonte C, Burt A, Powell EH, Beecham GW, et al. Genome-wide association study confirms SNPs in SNCA and the MAPT region as common risk factors for Parkinson disease. *Annals of Human Genetics*. 2010; 74(2):97–109. <https://doi.org/10.1111/j.1469-1809.2009.00560.x> PMID: 20070850
30. Hamza TH, Zabetian CP, Tenesa A, Laederach A, Montimurro J, Yearout D, et al. Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease. *Nature Genetics*. 2010; 42(9):781–785. <https://doi.org/10.1038/ng.642> PMID: 20711177
31. Saad M, Lesage S, Saint-Pierre A, Corvol JC, Zelenika D, Lambert JC, et al. Genome-wide association study confirms BST1 and suggests a locus on 12q24 as the risk loci for Parkinson's disease in the European population. *Human Molecular Genetics*. 2011; 20(3):615–627. <https://doi.org/10.1093/hmg/ddq497> PMID: 21084426
32. Do CB, Tung JY, Dorfman E, Kiefer AK, Drabant EM, Francke U, et al. Web-based genome-wide association study identifies two novel loci and a substantial genetic component for Parkinson's disease. *PLoS Genetics*. 2011; 7(6):e1002141. <https://doi.org/10.1371/journal.pgen.1002141> PMID: 21738487
33. Liu X, Cheng R, Verbitsky M, Kisselev S, Browne A, Mejia-Sanatana H, et al. Genome-wide association study identifies candidate genes for Parkinson's disease in an Ashkenazi Jewish population. *BMC Medical Genetics*. 2011; 12(1):1–16. <https://doi.org/10.1186/1471-2350-12-104>
34. International Parkinson Disease Genomics Consortium, Nalls MA, Plagnol V, Hernandez DG, Sharma M, Sheerin UM, et al. Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet (London, England)*. 2011; 377(9766):641–649. [https://doi.org/10.1016/S0140-6736\(10\)62345-8](https://doi.org/10.1016/S0140-6736(10)62345-8) PMID: 21292315
35. Spencer C, Plagnol V, Strange A, Gardner M, Paisán-Ruiz C, Band G, et al. Dissection of the genetics of Parkinson's disease identifies an additional association 5' of SNCA and multiple associated haplotypes at 17q21. *Human Molecular Genetics*. 2011; 20:345–353. <https://doi.org/10.1093/hmg/ddq469> PMID: 21044948
36. Simón-Sánchez J, Van Hilten JJ, Van De Warrenburg B, Post B, Berendse HW, Arepalli S, et al. Genome-wide association study confirms extant PD risk loci among the Dutch. *European Journal of Human Genetics*. 2011; 19(6):655–661. <https://doi.org/10.1038/ejhg.2010.254> PMID: 21248740
37. Lill CM, Roehr JT, McQueen MB, Kavvoura FK, Bagade S, Schjeide BMM, et al. Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: The PDGene database. *PLoS Genetics*. 2012; 8(3):e1002548. <https://doi.org/10.1371/journal.pgen.1002548> PMID: 22438815
38. Nalls MA, Pankratz N, Lill CM, Do CB, Hernandez DG, Saad M, et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nature genetics*. 2014; 46(9):989–993. <https://doi.org/10.1038/ng.3043> PMID: 25064009

39. Hill-Burns EM, Wissemann WT, Hamza TH, Factor SA, Zabetian CP, Payami H. Identification of a novel Parkinson's disease locus via stratified genome-wide association study. *BMC genomics*. 2014; 15:118. <https://doi.org/10.1186/1471-2164-15-118> PMID: 24511991
40. Foo JN, Tan LC, Irwan ID, Au WL, Low HQ, Prakash KM, et al. Genome-wide association study of Parkinson's disease in East Asians. *Human Molecular Genetics*. 2017; 26(1):226–232. PMID: 28011712
41. Chang D, Nalls M, Hallgrímsdóttir I, Hunkapiller J, van der Brug MP, Cai F, et al. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nature Genetics*. 2017; 49:1511–1516. <https://doi.org/10.1038/ng.3955> PMID: 28892059
42. Bandres-Ciga S, Ahmed S, Sabir MS, Blauwendraat C, Adames-Gmez AD, Bernal-Bernal I, et al. The Genetic Architecture of Parkinson Disease in Spain: Characterizing Population-Specific Risk, Differential Haplotype Structures, and Providing Etiologic Insight. *Movement disorders: official journal of the Movement Disorder Society*. 2019; 34(12):1851–1863. <https://doi.org/10.1002/mds.27864> PMID: 31660654
43. Blauwendraat C, Heilbron K, Vallerga CL, Bandres-Ciga S, Von Coelln R, Pihlstrøm L, et al. Parkinson's disease age at onset genome-wide association study: defining heritability, genetic loci, and α -synuclein mechanisms. *Movement Disorders*. 2019; 34(6):866–875. <https://doi.org/10.1002/mds.27659> PMID: 30957308
44. Nalls MA, Blauwendraat C, Vallerga CL, Heilbron K, Bandres-Ciga S, Chang D, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *The Lancet Neurology*. 2019; 18(12):1091–1102. [https://doi.org/10.1016/S1474-4422\(19\)30320-5](https://doi.org/10.1016/S1474-4422(19)30320-5) PMID: 31701892
45. Blauwendraat C, Reed X, Krohn L, Heilbron K, Bandres-Ciga S, Tan M, et al. Genetic modifiers of risk and age at onset in GBA associated Parkinson's disease and Lewy body dementia. *Brain*. 2020; 143(1):234–248. <https://doi.org/10.1093/brain/awz350> PMID: 31755958
46. Alfradique-Dunham I, Al-Ouran R, von Coelln R, Blauwendraat C, Hill E, Luo L, et al. Genome-wide association study Meta-analysis for Parkinson disease motor subtypes. *Neurology Genetics*. 2021; 7(2). <https://doi.org/10.1212/NXG.0000000000000557> PMID: 33987465
47. Backman JD, Li AH, Marcketta A, Sun D, Mbatchou J, Kessler MD, et al. Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature*. 2021; 599(7886):628–634. <https://doi.org/10.1038/s41586-021-04103-z> PMID: 34662886
48. Jiang L, Zheng Z, Fang H, Yang J. A generalized linear mixed model association tool for biobank-scale data. *Nature genetics*. 2021; 53(11):1616–1621. <https://doi.org/10.1038/s41588-021-00954-4> PMID: 34737426
49. Rodrigo LM, Nyholt DR. Imputation and Reanalysis of ExomeChip Data Identifies Novel, Conditional and Joint Genetic Effects on Parkinson's Disease Risk. *Genes*. 2021; 12(5):689. <https://doi.org/10.3390/genes12050689> PMID: 34064523
50. Smeland OB, Shadrin A, Bahrami S, Broce I, Tesli M, Frei O, et al. Genome-wide Association Analysis of Parkinson's Disease and Schizophrenia Reveals Shared Genetic Architecture and Identifies Novel Risk Loci. *Biological psychiatry*. 2021; 89(3):227–235. <https://doi.org/10.1016/j.biopsych.2020.01.026> PMID: 32201043
51. Sakae S, Kanai M, Tanigawa Y, Karjalainen J, Kurki M, Koshiba S, et al. A cross-population atlas of genetic associations for 220 human phenotypes. *Nature genetics*. 2021; 53(10):1415–1424. <https://doi.org/10.1038/s41588-021-00931-x> PMID: 34594039
52. Anderson CA, Boucher G, Lees CW, Franke A, D'Amato M, Taylor KD, et al. Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nature genetics*. 2011; 43(3):246–252. <https://doi.org/10.1038/ng.764> PMID: 21297633
53. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*. 2012; 491(7422):119–124. <https://doi.org/10.1038/nature11582> PMID: 23128233
54. Julià A, Domènech E, Chaparro M, García-Sánchez V, Gomollón F, Panés J, et al. A genome-wide association study identifies a novel locus at 6q22. 1 associated with ulcerative colitis. *Human molecular genetics*. 2014; 23(25):6927–6934. <https://doi.org/10.1093/hmg/ddu398> PMID: 25082827
55. Liu JZ, Van Sommeren S, Huang H, Ng SC, Alberts R, Takahashi A, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nature genetics*. 2015; 47(9):979–986. <https://doi.org/10.1038/ng.3359> PMID: 26192919
56. Ostrowski J, Paziewska A, Lazowska I, Ambrozkiwicz F, Goryca K, Kulecka M, et al. Genetic architecture differences between pediatric and adult-onset inflammatory bowel diseases in the Polish population. *Scientific reports*. 2016; 6(1):1–10. <https://doi.org/10.1038/srep39831> PMID: 28008999

57. Yang SK, Hong M, Oh H, Low HQ, Jung S, Ahn S, et al. Identification of loci at 1q21 and 16q23 that affect susceptibility to inflammatory bowel disease in Koreans. *Gastroenterology*. 2016; 151(6):1096–1099. <https://doi.org/10.1053/j.gastro.2016.08.025> PMID: 27569725
58. De Lange KM, Moutsianas L, Lee JC, Lamb CA, Luo Y, Kennedy NA, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nature genetics*. 2017; 49(2):256–261. <https://doi.org/10.1038/ng.3760> PMID: 28067908
59. Dönertaş HM, Fabian DK, Fuentealba M, Partridge L, Thornton JM. Common genetic associations between age-related diseases. *Nature aging*. 2021; 1(4):400–412. <https://doi.org/10.1038/s43587-021-00051-5> PMID: 33959723
60. Glanville KP, Coleman JR, O'Reilly PF, Galloway J, Lewis CM. Investigating pleiotropy between depression and autoimmune diseases using the UK Biobank. *Biological psychiatry global open science*. 2021; 1(1):48–58. <https://doi.org/10.1016/j.bpsgos.2021.03.002> PMID: 34278373
61. Wu Y, Murray GK, Byrne EM, Sidorenko J, Visscher PM, Wray NR. GWAS of peptic ulcer disease implicates *Helicobacter pylori* infection, other gastrointestinal disorders and depression. *Nature communications*. 2021; 12(1):1–17. <https://doi.org/10.1038/s41467-021-21280-7> PMID: 33608531
62. Consortium IMSG. Risk alleles for multiple sclerosis identified by a genomewide study. *New England Journal of Medicine*. 2007; 357(9):851–862. <https://doi.org/10.1056/NEJMoa073493>
63. De Jager PL, Jia X, Wang J, De Bakker PI, Ottoboni L, Aggarwal NT, et al. Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci. *Nature genetics*. 2009; 41(7):776–782. <https://doi.org/10.1038/ng.401> PMID: 19525953
64. Patsopoulos Nikolaos A, Bayer Pharma MS Genetics Working Group, the Steering Committees of Studies Evaluating IFN β -1b & a CCR1-Antagonist, Consortium A, GeneMSA, International Multiple Sclerosis Genetics Consortium, et al. Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. *Annals of neurology*. 2011; 70(6):897–912. <https://doi.org/10.1002/ana.22609> PMID: 22190364
65. Sawcer S, Hellenthal G, Pirinen M, Spencer CC, Patsopoulos NA, Moutsianas L, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature*. 2011; 476(7359):214. <https://doi.org/10.1038/nature10251> PMID: 21833088
66. Beecham AH, Patsopoulos NA, Xifara DK, Davis MF, Kempainen A, Cotsapas C, et al. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nature genetics*. 2013; 45(11):1353. <https://doi.org/10.1038/ng.2770> PMID: 24076602
67. Andlauer TF, Buck D, Antony G, Bayas A, Bechmann L, Berthele A, et al. Novel multiple sclerosis susceptibility loci implicated in epigenetic regulation. *Science advances*. 2016; 2(6):e1501678. <https://doi.org/10.1126/sciadv.1501678> PMID: 27386562
68. International Multiple Sclerosis Genetics Consortium, ANZgene, IIBDGC, WTCCC2, Patsopoulos N, Baranzini S, Santaniello A, Shoostari P, et al. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science*. 2019; 365(6460).
69. Mayes MD, Bossini-Castillo L, Gorlova O, Martin JE, Zhou X, Chen WV, et al. Immunochip analysis identifies multiple susceptibility loci for systemic sclerosis. *The American Journal of Human Genetics*. 2014; 94(1):47–61. <https://doi.org/10.1016/j.ajhg.2013.12.002> PMID: 24387989
70. López-Isac E, Acosta-Herrera M, Kerick M, Assassi S, Satpathy AT, Granja J, et al. GWAS for systemic sclerosis identifies multiple risk loci and highlights fibrotic and vasculopathy pathways. *Nature communications*. 2019; 10(1):1–14. <https://doi.org/10.1038/s41467-019-12760-y> PMID: 31672989
71. Fejzo M, Chen H, Anderson L, McDermott M, Karlan B, Konecny G, et al. Analysis in epithelial ovarian cancer identifies KANSL1 as a biomarker and target gene for immune response and HDAC inhibition. *Gynecologic Oncology*. 2020;. PMID: 33229045
72. Lyu R, Sun J, Xu D, Jiang Q, Wei C, Zhang Y. GESLM algorithm for detecting causal SNPs in GWAS with multiple phenotypes. *Briefings in Bioinformatics*. 2021; 22(6):bbab276. <https://doi.org/10.1093/bib/bbab276> PMID: 34323927
73. Graff M, Scott RA, Justice AE, Young KL, Feitosa MF, Barata L, et al. Genome-wide physical activity interactions in adiposity-A meta-analysis of 200,452 adults. *PLoS Genetics*. 2017; 13(4):e1006528. <https://doi.org/10.1371/journal.pgen.1006528> PMID: 28448500
74. Day F, Karaderi T, Jones MR, Meun C, He C, Drong A, et al. Large-scale genome-wide meta-analysis of polycystic ovary syndrome suggests shared genetic architecture for different diagnosis criteria. *PLoS Genetics*. 2018; 14(12):e1007813. <https://doi.org/10.1371/journal.pgen.1007813> PMID: 30566500
75. Kalra G, Milon B, Casella AM, Herb BR, Humphries E, Song Y, et al. Biological insights from multi-omic analysis of 31 genomic risk loci for adult hearing difficulty. *PLoS Genetics*. 2020; 16(9):e1009025. <https://doi.org/10.1371/journal.pgen.1009025> PMID: 32986727

76. Ray D, Boehnke M. Methods for meta-analysis of multiple traits using GWAS summary statistics. *Genetic Epidemiology*. 2018; 42:134–145. <https://doi.org/10.1002/gepi.22105> PMID: 29226385
77. Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, et al. Genome-wide association studies. *Nature Reviews Methods Primers*. 2021; 1(1):1–21. <https://doi.org/10.1038/s43586-021-00056-9>
78. Pingault JB, O'reilly PF, Schoeler T, Ploubidis GB, Rijdsdijk F, Dudbridge F. Using genetic data to strengthen causal inference in observational research. *Nature Reviews Genetics*. 2018; 19(9):566–580. <https://doi.org/10.1038/s41576-018-0020-3> PMID: 29872216