The American Journal of Human Genetics, Volume *101*

Supplemental Data

A Scalable Bayesian Method for Integrating Functional Information in Genome-wide Association Studies

Jingjing Yang, Lars G. Fritsche, Xiang Zhou, Gonçalo Abecasis, and International Age-Related Macular Degeneration Genomics Consortium

Supplemental Data

Supplemental Note

Technical Details about bfGWAS

Supplemental Figures

Figure S1: Flowcharts of bfGWAS.

Figure S2: Plots of the potential scale reduction factors (PSRF).

Figure S3: Prioritization ranks of the true causal SNP1 (pink) and SNP2 (cyan).

Figure S4: Estimates of the log-relative-risk $\ln(\pi_0/\pi_1)$ by bfGWAS and the enrich-parameter by fGWAS, along with 95% confidence intervals.

Figure S5: Estimates of the log-ratio of effect-size variances $\ln(\sigma_0^2/\sigma_1^2)$ by bfGWAS, along with 95% confidence intervals.

Figure S6: Sorted top bfGWAS PPs versus sorted top -log10(P-values) of single variant tests.

Figure S7: Manhattan plot highlighting AMD GWAS signals with BVSR PP >0.1068.

Figure S8: Manhattan plots highlighting AMD GWAS signals by accounting for gene-based annotations.

Figure S9: ZoomLocus plots of region *CHR19:6218146-7218146*.

Figure S10: Enrichment analysis results with varying prior means as well as starting values $(10^{-6}, 5\times10^{-6}, 10^{-5})$ for π_q, and varying starting values $(10, 5, 1)$ for σ_q^2.

Figure S11: fGWAS enrichment estimates with 95% error bars.

Figure S12: Ratios of enrich parameters versus the respective genome-wide averages, along with 95% confidence intervals.

Figure S13: Enrichment analysis results for the AMD GWAS data with chromatin states profiled with respect to the epigenome of fetal thymus (E093).

Figure S14: Manhattan plot highlighting MGI GWAS signals of skin cancer with BVSR PP >0.1068.

Figure S15: Manhattan plots highlighting MGI GWAS signals of skin cancer by accounting for gene-based annotations.

Figure S16: Enrichment analysis results of the MGI GWAS of skin cancer, accounting for gene-based annotations.

Figure S17: LocusZoom plots of region of *CHR16:89686117-90172696*.

Supplemental Tables

Table S1: Classification of gene-based functional annotations.

Table S2: Compare results by P-value, fGWAS, and bfGWAS in the 34 known AMD loci, accounting for gene-based functional annotations.

Table S3: AMD risk variants identified by bfGWAS in the 34 known loci, accounting for gene-based annotations.

Table S4: AMD risk variants by fGWAS in the 34 known loci, accounting for gene-based annotations.

Table S5: Candidate AMD loci identified by bfGWAS, accounting for gene-based annotations.

Table S6: Candidate AMD loci identified by fGWAS, accounting for gene-based annotations.

Table S7: AMD risk variants by bfGWAS in the 34 known loci, accounting for summarized regulatory annotations.

Table S8: AMD risk variants by fGWAS in the 34 known loci, accounting for summarized regulatory annotations.

Table S9: Candidate AMD loci identified by bfGWAS, accounting for summarized regulatory annotations.

Table S10: Candidate AMD loci identified by fGWAS, accounting for summarized regulatory annotations.

Table S11: AMD risk variants by bfGWAS in the 34 known loci, accounting for chromatin states profiled with the epigenome of fetal thymus.

Table S12: AMD risk variants by fGWAS in the 34 known loci, accounting for chromatin states profiled with the epigenome of fetal thymus.

Table S13: Candidate AMD loci identified by bfGWAS, accounting for chromatin states profiled with the epigenome of fetal thymus.

Table S14: Candidate AMD loci identified by fGWAS, accounting for chromatin states profiled with the epigenome of fetal thymus.

Table S15: Haplotype analysis in locus C2/CFB/SKIV2L.

Table S16: Model comparison.

Supplemental References

Supplemental Note

Technical Details about bfGWAS

1 Bayesian Hierarchical Model

1.1 Standard Bayesian Variable Selection Regression Model

Consider the following standard Bayesian variable selection regression (BVSR) model

$$\boldsymbol{y}_{n\times 1} = \boldsymbol{X}_{n\times p}\boldsymbol{\beta}_{p\times 1} + \boldsymbol{\epsilon}_{n\times 1},\ \beta_i \sim \pi_i N(0, \tau^{-1}\sigma_i^2) + (1-\pi_i)\delta_0(\beta_i),\ \epsilon_i \sim N(0, \tau^{-1}), \quad (1)$$

where $\boldsymbol{y}_{n\times 1}$ denotes the centered phenotype vector of n samples; $\boldsymbol{X}_{n\times p}$ denotes the centered genotype matrix of p genetic variants; ϵ_i denotes the residual error independently and identically distributed (i.i.d.) with normal distribution $N(0, \tau^{-1})$; and β_i follows a spike-and-slab prior distribution [5, 6, 7] — that is, β_i follows the normal distribution $N(0, \tau^{-1}\sigma_i^2)$ with probability π_i and the point-mass density function $\delta_0(\cdot)$ at 0 with probability $(1-\pi_i)$ ($\delta_0(\beta_i) = 1$ if $\beta_i = 0$, otherwise $\delta_0(\beta_i) = 0$).

Here, the genotype matrix contains either dosage data within range $[0, 2]$ or genotype data with values $\{0, 1, 2\}$ denoting the number of minor alleles. The assumption of the spike-and-slab prior for β_i enforces variable selection in the regression model (1). We drop the intercept term here for assuming both $\boldsymbol{y}_{n\times 1}$ and columns of $\boldsymbol{X}_{n\times p}$ are centered. Although this model is developed for quantitative trait, we can treat dichotomous traits (e.g., cases and controls) as quantitative with values of 1 and 0 (e.g., 1 for cases and 0 for controls), which was proven to be equivalent as using the logistic or probit model by previous approaches [6, 7].

1.2 Integrating Functional Information

In this paper, we only consider non-overlapped categorical annotations. Let $\boldsymbol{A_i} = (A_{i1}, \cdots, A_{iQ})^T$ denotes the vector of Q annotations for the ith variant, where A_{iq} takes binary values $(1/0)$ to denote whether the ith variant is of the qth annotation. In order to integrate functional annotations into the standard BVSR model (1), we assume all variants

of annotation q have the same spike-and-slab prior with parameters (π_q, σ_q^2). We further assume the following independent and conjugate hyper priors (Figure S 1(A)):

$$\pi_q \text{ i.i.d.} \sim Beta(a_q, b_q),\ \sigma_q^2 \text{ i.i.d.} \sim IG(k_1, k_2),\ \tau \sim G(k_3, k_4), \tag{2}$$

where $Beta(a_q, b_q)$ denotes a Beta distribution with positive shape parameters a_q and b_q, $IG(k_1, k_2)$ denotes an Inverse-Gamma distribution with shape parameter k_1 and scale parameter k_2, and $G(k_3, k_4)$ denotes a Gamma distribution with shape parameter k_3 and scale parameter k_4 (Figure S1(A)). Note that parameters (a_q, b_q) could be different with respect to different annotations. This hierarchical BVSR model is equivalent to the standard BVSR model when modeling no functional information (i.e., assuming the same π_q and σ_q^2 for all variants).

In order to adjust for the unbalance distribution of functional annotations among all variants and encourage for a sparse model, we choose values for a_q and b_q such that the mean of the Beta distribution $\frac{a_q}{a_q+b_q} = 10^{-6}$ with $(a_q + b_q) = m_q = \sum_{i=1,j=q}^{p} A_{ij}$ (the total number of variants of annotation q). Here, the mean 10^{-6} of $Beta(a_q, b_q)$ helps enforce a sparse initial model that is desired for controlling false positives (assuming one signal per1M variants). We take $k_1 = k_2 = k_3 = k_4 = 0.1$ to induce non-informative priors on σ_q^2 and τ. Thus, the posterior estimates of π_q and σ_q^2 will mainly depend on the data likelihood. However, when there are few association signals in the qth category, the posterior estimates of π_q and σ_q^2) will be set as their respective prior modes. Note that although the hyper priors are assumed to be independent, the posterior distributions of π_q and σ_q^2 are no longer independent.

1.3 Latent Indicator Variable

To facilitate computation, we introduce a latent indicator vector $\boldsymbol{\gamma}_{p\times 1}$ [5] into the model, where each element $\gamma_i \in \{0, 1\}$ indicates whether the corresponding ith effect β_i equals to 0 with $\gamma_i = 0$ or follows the $N(0, \tau^{-1}\sigma_i^2)$ distribution with $\gamma_i = 1$. Equivalently,

$$\gamma_i \sim Bernoulli(\pi_i),\ \boldsymbol{\beta}_{-\boldsymbol{\gamma}} \sim \delta_0(\cdot),\ \boldsymbol{\beta}_{\boldsymbol{\gamma}} \sim MVN_{|\boldsymbol{\gamma}|}(0, \tau^{-1}\boldsymbol{V}_{\boldsymbol{\gamma}}),$$

where $|\boldsymbol{\gamma}|$ denotes the number of non-zero entries in $\boldsymbol{\gamma}$; $\boldsymbol{\beta}_{-\boldsymbol{\gamma}}$ denotes the sub-vector of $\boldsymbol{\beta}_{p\times 1}$ corresponding to variants with $\gamma_i = 0$; $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ denotes the sub-vector of $\boldsymbol{\beta}_{p\times 1}$ corresponding to the variants with $\{\gamma_j = 1; j = 1, \cdots, |\boldsymbol{\gamma}|\}$; and $\boldsymbol{V}_{|\boldsymbol{\gamma}|}$ is the corresponding sub-matrix (with $\gamma_j = 1$) of $\boldsymbol{V}_{p\times p} = diag(\sigma_1^2, \cdots, \sigma_p^2)$.

1.4 Bayesian Inference

With the above Bayesian hierarchical model, the posterior joint distribution of $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\sigma^2}, \boldsymbol{\pi}, \tau)$ is proportional to the product of likelihood and prior density functions,

$$P(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\sigma^2}, \boldsymbol{\pi}, \tau | \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{A}) \propto P(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \tau) P(\boldsymbol{\beta}|\boldsymbol{A}, \boldsymbol{\pi}, \boldsymbol{\sigma^2}, \tau) P(\boldsymbol{\gamma}|\boldsymbol{\pi}) P(\boldsymbol{\pi}) P(\boldsymbol{\sigma^2}) P(\tau), \quad (3)$$

where $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_Q)$, $\boldsymbol{\sigma^2} = (\sigma_1^2, \ldots, \sigma_Q^2)$, and $\boldsymbol{A}$ is the $p \times Q$ annotation matrix with binary values.

Now our goal is to make inference on the category-specific parameters $(\boldsymbol{\pi}, \boldsymbol{\sigma^2})$ and the variable-specific parameters $(\boldsymbol{\beta}, E[\boldsymbol{\gamma}])$ from their respective marginal posterior distributions, conditioning on the data $(\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{A})$. The category-specific parameters $(\boldsymbol{\pi}, \boldsymbol{\sigma^2})$ denote the shared characteristics of variants with the same annotation, which are also referred as enrichment parameters in this paper. Specifically, π_q denotes the causality for variants of annotation q, and σ_q^2 denotes the effect-size variance for associated variants (with nonzero β_j) of annotation q.

To make the Bayesian inference of our model applicable for genome-wide analysis, we pair it with a novel Expectation-Maximization Markov chain Monte Carlo (EM-MCMC) algorithm. Because of the block-wise linkage disequilibrium (LD) structure of human genome, we can segment the genotype data $\boldsymbol{X}$ into K approximately independent blocks, i.e., $\boldsymbol{X} = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_K\}$, where each submatrix $\boldsymbol{X}_k$ has dimension $n \times p_k$ (genotypes of p_k variants for n samples). Thus, we can write the likelihood function in (3) as a product of a series likelihood functions for $\boldsymbol{X_k}$,

$$P(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \tau) = \prod_{k=1}^{K} P_k(\boldsymbol{y}|\boldsymbol{X}_k, \boldsymbol{\beta}_k, \boldsymbol{\gamma}_k, \tau), \quad (4)$$

where $(\boldsymbol{y}|\boldsymbol{X}_k, \boldsymbol{\beta}_k, \boldsymbol{\gamma}_k, \tau) \sim MVN_{|\boldsymbol{\gamma_k}|}(\boldsymbol{X}_k\boldsymbol{\beta}_k, \tau^{-1}\boldsymbol{I}_{|\boldsymbol{\gamma_k}|})$.

To avoid adjusting for the residual variance with respect to each genome-block, we fix τ^{-1} as the phenotype variance. This assumption is reasonable because most genome-blocks explain little phenotype variance in practice. Although fixing τ^{-1} as the phenotype variance seems conservative for genome-blocks with true signals, our analysis showed that it barely affect identifying true signals.

In the Expectation step (E-step), $(\boldsymbol{\beta_k}, E[\boldsymbol{\gamma_k}])$ are estimated by implementing MCMC per block, conditioning on the given values of $(\boldsymbol{\pi}, \boldsymbol{\sigma})$; in the Maximization step (M-step), $(\boldsymbol{\pi}, \boldsymbol{\sigma})$ are updated, conditioning on genome-wide estimates of $(\boldsymbol{\beta}, E[\boldsymbol{\gamma}])$ from the E-step. In general, $\sim$5 EM iterations will lead to convergent estimates of $(\boldsymbol{\pi}, \boldsymbol{\sigma})$, and the estimates of $(\boldsymbol{\beta_k}, E[\boldsymbol{\gamma_k}])$ from the last E-step will be used to identify association signals (details are provided in Section 2; Figure S 1(B)).

1.4.1 Conditional Posterior Distribution for $\boldsymbol{\beta_k}$

Conditioning on the values of $(\boldsymbol{\pi}, \boldsymbol{\sigma^2}, \tau)$, the posterior distribution for the variant-specific parameters $(\boldsymbol{\beta}_k, \boldsymbol{\gamma}_k)$ of block k is

$$P(\boldsymbol{\beta}_k, \boldsymbol{\gamma}_k | \boldsymbol{X}_k, \boldsymbol{y}, \boldsymbol{\pi}, \boldsymbol{\sigma^2}, \tau) \propto P(\boldsymbol{y} | \boldsymbol{X}_k, \boldsymbol{\beta}_k, \boldsymbol{\gamma}_k, \tau) P(\boldsymbol{\beta}_k | \boldsymbol{\gamma}_k, \boldsymbol{\sigma^2}, \tau) P(\boldsymbol{\gamma}_k | \boldsymbol{\pi}). \tag{5}$$

Conditioning on the indicator vector $\boldsymbol{\gamma_k}$, the effect-sizes associated with zero indicator variables are 0, while the posterior distribution for $\boldsymbol{\beta}_{|\boldsymbol{\gamma_k}|}$ is given by

$$\begin{aligned}
&P(\boldsymbol{\beta}_{|\boldsymbol{\gamma_k}|} | \boldsymbol{X}_{|\boldsymbol{\gamma_k}|}, \boldsymbol{y}, \boldsymbol{\gamma}_k, \boldsymbol{\sigma^2}, \tau) \propto P_k(\boldsymbol{y} | \boldsymbol{X}_{|\boldsymbol{\gamma_k}|}, \boldsymbol{\beta}_{|\boldsymbol{\gamma_k}|}, \boldsymbol{\gamma}_k, \tau) P(\boldsymbol{\beta}_{|\boldsymbol{\gamma_k}|} | \boldsymbol{\gamma}_k, \boldsymbol{\sigma^2}, \tau) \\
&\propto \exp\left\{-\tfrac{\tau}{2}(\boldsymbol{y} - \boldsymbol{X}_{|\boldsymbol{\gamma_k}|}\boldsymbol{\beta}_{|\boldsymbol{\gamma_k}|})^T(\boldsymbol{y} - \boldsymbol{X}_{|\boldsymbol{\gamma_k}|}\boldsymbol{\beta}_{|\boldsymbol{\gamma_k}|})\right\} \exp\left\{-\tfrac{\tau}{2}\boldsymbol{\beta}_{|\boldsymbol{\gamma_k}|}^T \boldsymbol{V}_{|\boldsymbol{\gamma_k}|}^{-1} \boldsymbol{\beta}_{|\boldsymbol{\gamma_k}|}\right\} \\
&\propto \exp\left\{-\tfrac{\tau}{2}\left(\boldsymbol{\beta}_{|\boldsymbol{\gamma_k}|}^T \boldsymbol{X}_{|\boldsymbol{\gamma_k}|}^T \boldsymbol{X}_{|\boldsymbol{\gamma_k}|} \boldsymbol{\beta}_{|\boldsymbol{\gamma_k}|} - 2\boldsymbol{\beta}_{|\boldsymbol{\gamma_k}|}^T \boldsymbol{X}_{|\boldsymbol{\gamma_k}|} \boldsymbol{y} + \boldsymbol{\beta}_{|\boldsymbol{\gamma_k}|}^T \boldsymbol{V}_{|\boldsymbol{\gamma_k}|}^{-1} \boldsymbol{\beta}_{|\boldsymbol{\gamma_k}|}\right)\right\} \\
&\propto \exp\left\{-\tfrac{\tau}{2}\left(\boldsymbol{\beta}_{|\boldsymbol{\gamma_k}|}^T (\boldsymbol{X}_{|\boldsymbol{\gamma_k}|}^T \boldsymbol{X}_{|\boldsymbol{\gamma_k}|} + \boldsymbol{V}_{|\boldsymbol{\gamma_k}|}^{-1}) \boldsymbol{\beta}_{|\boldsymbol{\gamma_k}|} - 2\boldsymbol{\beta}_{|\boldsymbol{\gamma_k}|}^T \boldsymbol{X}_{|\boldsymbol{\gamma_k}|}^T \boldsymbol{y}\right)\right\}.
\end{aligned} \tag{6}$$

From (6), it is easy to see that

$$\begin{aligned}
&(\boldsymbol{\beta}_{|\boldsymbol{\gamma_k}|} | \boldsymbol{X}_{|\boldsymbol{\gamma_k}|}, \boldsymbol{y}, \boldsymbol{\gamma}_k, \boldsymbol{\sigma^2}, \tau) \sim \\
&MVN_{|\boldsymbol{\gamma_k}|}\left((\boldsymbol{X}_{|\boldsymbol{\gamma_k}|}^T \boldsymbol{X}_{|\boldsymbol{\gamma_k}|} + \boldsymbol{V}_{|\boldsymbol{\gamma_k}|}^{-1})^{-1} \boldsymbol{X}_{|\boldsymbol{\gamma_k}|}^T \boldsymbol{y},\ \tau^{-1}(\boldsymbol{X}_{|\boldsymbol{\gamma_k}|}^T \boldsymbol{X}_{|\boldsymbol{\gamma_k}|} + \boldsymbol{V}_{|\boldsymbol{\gamma_k}|}^{-1})^{-1}\right).
\end{aligned} \tag{7}$$

Here, the subscript $|\gamma_k|$ indicates sub-matrices or sub-vectors corresponding to variants with nonzero indicator variables, and $\boldsymbol{V}_{|\boldsymbol{\gamma_k}|}$ is a diagonal matrix with $(\boldsymbol{V}_{|\boldsymbol{\gamma_k}|})_{jj} = \sigma_q^2$ if the jth variant is of annotation q.

1.4.2 Conditional Posterior Distribution for $\boldsymbol{\gamma_k}$

Because of the conditional conjugate prior for $\boldsymbol{\beta_k}$, we can easily integrate $\boldsymbol{\beta}_k$ out from the joint conditional posterior distribution (5) to obtain the marginal conditional posterior distribution for $\boldsymbol{\gamma}_k$,

$$\begin{aligned}
P(\boldsymbol{\gamma}_k | \boldsymbol{X}_k, \boldsymbol{y}, \boldsymbol{\pi}, \boldsymbol{\sigma^2}, \tau) &\propto \int_{\boldsymbol{\beta}_k} P_k(\boldsymbol{y} | \boldsymbol{X}_k, \boldsymbol{\beta}_k, \boldsymbol{\gamma}_k, \tau) P(\boldsymbol{\beta}_k | \boldsymbol{\gamma}_k, \boldsymbol{\sigma^2}, \tau) P(\boldsymbol{\gamma}_k | \boldsymbol{\pi}) d\boldsymbol{\beta}_k \\
&\propto |\boldsymbol{\Omega}_{|\boldsymbol{\gamma_k}|}|^{-1/2} \exp\left\{\frac{\tau}{2} \boldsymbol{y}^T \boldsymbol{X}_{|\boldsymbol{\gamma_k}|} \boldsymbol{V}_{|\boldsymbol{\gamma_k}|} \boldsymbol{\Omega}_{|\boldsymbol{\gamma_k}|}^{-1} \boldsymbol{X}_{|\boldsymbol{\gamma_k}|}^T \boldsymbol{y}\right\} P(\boldsymbol{\gamma}_k | \boldsymbol{\pi}),
\end{aligned} \tag{8}$$

where $\boldsymbol{\Omega}_{|\boldsymbol{\gamma_k}|} = \boldsymbol{V}_{|\boldsymbol{\gamma_k}|} \boldsymbol{X}_{|\boldsymbol{\gamma_k}|}^T \boldsymbol{X}_{|\boldsymbol{\gamma_k}|} + \boldsymbol{I}_{|\boldsymbol{\gamma_k}|}$.

2 EM-MCMC Algorithm

The steps of the EM-MCMC algorithm are as follows:

(i) Fix τ at the value of phenotype variance;

(ii) Set initial values for the category-specific parameters $(\boldsymbol{\pi}, \boldsymbol{\sigma^2})$;

(iii) E-step: Conditioning on the most recent values of $(\boldsymbol{\pi}, \boldsymbol{\sigma^2})$, estimate variant-specific parameters $(\boldsymbol{\beta}, E[\boldsymbol{\gamma}])$ by implementing MCMC per block;

(iv) M-step: Conditioning on the genome-wide estimates of $(\boldsymbol{\beta}, E[\boldsymbol{\gamma}])$ from the previous E-step, update $(\boldsymbol{\pi}, \boldsymbol{\sigma^2})$ by their MAPs (maximum a posteriori estimates), maximizing the expected log-posterior-likelihood functions [2];

(v) Repeat the EM-steps (iii) and (iv) for a few times until the MAPs of $(\boldsymbol{\pi}, \boldsymbol{\sigma^2})$ converge.

2.1 Setup Initial Values

In this paper, we fix τ at the value of phenotype variance, equivalent to assuming no phenotype variance explained by the genetic variants. This assumption is true for most blocks and slightly conservative for blocks with true signals. However, our analysis showed that this assumption barely affects identifying true signals. We take initial values $\pi_q = 1 \times 10^{-6}$ to initial a sparse and conservative model, and $\sigma_q^2 = 10$ to start with a large effect-size variance for all associated variants.

2.2 MCMC Sampling Scheme

The MCMC sampling is implemented per block for estimating $(\boldsymbol{\beta_k}, E[\boldsymbol{\gamma_k}])$, conditioning on category-specific parameters $(\boldsymbol{\pi}, \boldsymbol{\sigma^2})$:

(i) First, sort all variants in the block by their base positions, perform single variant tests, and rank variants based on their marginal association evidence (e.g., P-values) from strong to weak.

(ii) Second, select an initial model with independent significant signals. We first include the variant with the smallest P-value into the model (i.e., set the corresponding indicator value as 1). Then, conditioning on the currently selected variant(s), select the next most significant variant with P-value $< 5 \times 10^{-8}$. Stop selection when no other independent genome-wide signal exists. Generally, most of the blocks with $\sim$10K variants will start with only one variant.

(iii) Third, repeat the MCMC sampling for a large number of iterations (e.g., 50K iterations with 50K burnins), in which the Metropolis-Hastings algorithm is used

to draw posterior samples for $\boldsymbol{\gamma_k}$ based on (8). With indicator vector $\boldsymbol{\gamma'_k}$ and corresponding effect-size vector $\boldsymbol{\beta}_{|\boldsymbol{\gamma'_k}|}$ from previous iteration, each MCMC iteration is as follows:

(a) Randomly propose a new indicator vector $\boldsymbol{\gamma''_k}$ by:

- Including an extra variant into the model with probability $1/3$: generate a rank r from a proposal distribution P_{γ_k} such that the variant with rank r is not included in the current model (change the corresponding indicator variable from 0 to 1). Here, P_{γ_k} is constructed as the mixture distribution $0.9{*}U_{top}{+}0.1U_{rest}$, where U_{top} denotes the uniform distribution on top ranks $(1, \ldots, t_k)$ and U_{rest} denotes the uniform distribution on the remain ranks $(t_{k+1}, \cdots, p_k)$ (t_k is an arbitrary number). That is, we assume a variant whose P-value is ranked in the top association group will be proposed with probability $0.9/(t_k)$, while a variant in the remaining group will be proposed with probability $0.1/(p_k - t_k)$. A rank will keep being proposed from P_{γ_k} until the corresponding variant is absent in the current model. We take $t_k = min(p_k, 300)$ in our software.
- Deleting a variant from the current model with probability $1/3$: randomly delete a variant from the current model (change the corresponding indicator variable from 1 to 0), i.e., each variant in the current model has probability $1/|\boldsymbol{\gamma'_k}|$ to be deleted.
- Switching a variant in the current model with an un-included variant in the neighborhood of the switch candidate (switch the corresponding indicator variable values): randomly select a variant in the current model as a switch candidate; propose a variant within its neighborhood from the proposal distribution P_{neib}. In order to improve the MCMC mixing property, we calibrate P_{neib} based on the conditional association evidence of all un-included variants in the neighborhood, conditioning on all variants in the current model except the switch candidate. For example, if there are 20 un-included variants in the neighborhood with conditional likelihood ratio test (LRT) statistic values $\{s_1, \cdots, s_{20}\}$, we first subtract the largest statistic value s_{max} from all values, then take $P_{neib}(s_j) = \exp(s_j - s_{max}) / \sum_{b=1}^{20} \exp(s_b - s_{max})$ as the probability for the corresponding jth variant to be proposed. The neighborhood size can be tuned by users (we set the neighborhood window as 100 variants near the switch candidate in our analyses).

(b) Conditioning on the indicator vector $\boldsymbol{\gamma}_{\boldsymbol{k}}''$, the effect-size vector $\boldsymbol{\beta}_{|\boldsymbol{\gamma}_{\boldsymbol{k}}''|}$ is estimated by its conditional posterior mean in (7).

(c) Calculate the Metropolis-Hastings acceptance ratio, and then decide whether to accept or reject $\boldsymbol{\gamma}_{\boldsymbol{k}}''$ by the Metropolis-Hastings algorithm.

(iv) Finally, $E[\gamma_{kj}]$ is estimated by u_{kj}/M, where u_{kj} is the number of times when the jth variant in block k is included into the model and M is the total MCMC iterations. Note that $E[\gamma_{kj}]$ is also referred as the Bayesian posterior inclusion probability (PP), evidence for the ith variant in block k to be an association signal. The Bayesian estimate of the corresponding β_{kj} is given by the posterior mean $\sum_{l=1}^{u_{kj}} \beta_{kjl}/u_{kj}$, where β_{kjl} is the effect-size estimate for the jth variant (in block k) when it is included into the model for the lth time.

Within the MCMC sampling, we also record the number of iterations M_{active} when the linear regression model includes at least one variant by the Metropolis-Hastings algorithm. Then the proportion of such MCMC iterations M_{active}/M gives us the regional posterior inclusion probability (regional-PP) of the study block, which is the probability of existing at least one signal in the block. Because variants in high LD and the same annotation category have the same chance to be included into the linear model (splitting the posterior probability for a single signal), the regional-PP is more appropriate than the single variant Bayesian PP for claiming a risk locus.

2.3 EM Algorithm

In the EM algorithm, values of $(\boldsymbol{\pi}, \boldsymbol{\sigma^2})$ are updated by their respective maximum a posteriori estimates (MAPs), maximizing expected log-posterior-likelihood functions. With the Bayesian estimates of $(\boldsymbol{\beta}, E[\boldsymbol{\gamma}])$ from the E-step, the expected log-posterior-likelihood functions and MAPs can be derived with closed-form expressions.

2.3.1 MAP for $\boldsymbol{\sigma^2}$

From the joint posterior distribution (3), the conditional posterior density function (posterior likelihood) of $\boldsymbol{\sigma^2}$ becomes

$$P(\boldsymbol{\sigma^2}|\boldsymbol{\beta}, \boldsymbol{\gamma}, \tau) \propto P(\boldsymbol{\beta}|\boldsymbol{\gamma}, \boldsymbol{\sigma^2}, \tau) P(\boldsymbol{\sigma^2}), \tag{9}$$

where $P(\boldsymbol{\sigma^2}) = \prod_{q=1}^{Q} P(\sigma_q^2)$ with $\sigma_q^2 \sim IG(k_1, k_2)$, i.e. $P(\sigma_q^2) \propto \left(\sigma_q^2\right)^{-(k_1+1)} \exp\left(-\frac{k_2}{\sigma_q^2}\right)$; $P(\boldsymbol{\beta}|\boldsymbol{\gamma}, \boldsymbol{\sigma^2}, \tau) = \prod_{i=1}^{p} P(\beta_i|\sigma_i^2, \gamma_i, \tau)$ with $P(\beta_i|\sigma_i^2, \gamma_i, \tau) =$

$(\gamma_i N(\beta_i; 0, \tau^{-1}\sigma_i^2) + (1-\gamma_i)\delta_0(\beta_i))$; and $\sigma_i^2 = \sigma_q^2$ if the ith variant is of annotation q.

The expected log-posterior-likelihood of $\boldsymbol{\sigma^2}$ is given by

$$\begin{aligned}
& l(\boldsymbol{\sigma^2}) = E_{\boldsymbol{\gamma}}\left[ln(P(\boldsymbol{\sigma^2}|\boldsymbol{\beta},\boldsymbol{\gamma},\tau))\right] \\
= \;& E_{\boldsymbol{\gamma}}\left[\sum_{i=1}^{p} ln\left(P(\beta_i|\sigma_i^2,\gamma_i,\tau)\right)\right] + \sum_{q=1}^{Q} ln\left(P(\sigma_q^2)\right) + C \\
= \;& \sum_{i=1}^{p} E_{\boldsymbol{\gamma}}\left[ln\left(P(\beta_i|\sigma_i^2,\gamma_i,\tau)\right)\right] + \sum_{q=1}^{Q} ln\left(P(\sigma_q^2)\right) + C \\
\approx \;& \sum_{i=1}^{p}\left[\widehat{\gamma_i} ln\left(P(\beta_i|\gamma_i=1,\sigma_i^2)\right) + (1-\widehat{\gamma_i}) ln\left(P(\beta_i|\gamma_i=0)\right)\right] + \\
& \sum_{q=1}^{Q}\left[(k_1+1)ln\left(\frac{1}{\sigma_q^2}\right) - k_2\frac{1}{\sigma_q^2}\right] + C \\
= \;& \sum_{i=1}^{p}\left[\widehat{\gamma_i}\left(\frac{1}{2}ln\left(\frac{\tau}{\sigma_i^2}\right) - \frac{\tau\widehat{\beta_i}^2}{2\sigma_i^2}\right)\right] + \sum_{q=1}^{Q}\left[(k_1+1)ln\left(\frac{1}{\sigma_q^2}\right) - k_2\frac{1}{\sigma_q^2}\right] + C, \qquad (10)
\end{aligned}$$

where $\{\widehat{\gamma_i} = E[\gamma_i]\}, \{\widehat{\beta_i}\}$ are Bayesian estimates by MCMC in the E-step, and C is a constant free of $\boldsymbol{\sigma^2}$.

From (10), we can see that the posterior distributions of $\{\sigma_q^2;\, q = 1, \ldots, Q\}$ are disjoint, because of independent priors and non-overlapped annotations. Thus, the expected log-posterior-likelihood function for each σ_q^2 is

$$l_{\sigma_q^2} = \sum_{j_q=1}^{m_q}\left[\widehat{\gamma_{j_q}}\left(\frac{1}{2}ln\left(\frac{\tau}{\sigma_q^2}\right) - \frac{\tau\widehat{\beta_{j_q}}^2}{2\sigma_q^2}\right)\right] + (k_1+1)ln\left(\frac{1}{\sigma_q^2}\right) - \frac{k_2}{\sigma_q^2} + C, \qquad (11)$$

where $\{\widehat{\gamma_{j_q}}, \widehat{\beta_{j_q}};\, j_q = 1, \ldots, n_q\}$ are the Bayesian estimates for variants of annotation q, and m_q is the total number of variants with annotation q. The MAP of σ_q^2 can be solved from

$$\frac{dl_{\sigma_q^2}}{d(1/\sigma_q^2)} = \sum_{j_q=1}^{m_q}\left[\widehat{\gamma_{j_q}}\frac{\sigma_q^2}{2} - \widehat{\gamma_{j_q}}\frac{\tau\beta_{j_q}^2}{2}\right] + (k_1+1)\sigma_q^2 - k_2 = 0,$$

which is

$$\widehat{\sigma_q^2} = \frac{\tau\sum_{j_q=1}^{m_q}(\widehat{\gamma_{j_q}}\widehat{\beta_{j_q}^2}) + 2k_2}{\sum_{j_q=1}^{m_q}\widehat{\gamma_{j_q}} + 2(k_1+1)}.$$

2.3.2 MAP for $\boldsymbol{\pi}$

From the joint posterior distribution (3), the conditional posterior density function (posterior likelihood) of $\boldsymbol{\pi}$ becomes

$$P(\boldsymbol{\pi}|\boldsymbol{\gamma}) \propto P(\boldsymbol{\gamma}|\boldsymbol{\pi})P(\boldsymbol{\pi}), \tag{12}$$

where $P(\boldsymbol{\gamma}|\boldsymbol{\pi}) = \prod_{i=1}^{p} P(\gamma_i|\pi_i) \propto \prod_{i=1}^{p} \pi_i^{\gamma_i}(1-\pi_i)^{1-\gamma_i}$; $\pi_i = \pi_q$ if the ith variant is of annotation q; and $P(\boldsymbol{\pi}) = \prod_{q=1}^{Q} P(\pi_q)$ with π_q i.i.d. $\sim Beta(a_q, b_q)$.

The expected log-posterior-likelihood of $\boldsymbol{\pi}$ can be derived as

$$\begin{aligned}
& l(\boldsymbol{\pi}) = E_{\boldsymbol{\gamma}}\left[ln(P(\boldsymbol{\pi}|\boldsymbol{\gamma}))\right] \\
= \;& E_{\boldsymbol{\gamma}}\left[\sum_{i=1}^{p} ln(P(\gamma_i|\pi_i))\right] + ln\left(P(\boldsymbol{\pi})\right) + C \\
= \;& \sum_{i=1}^{p} E_{\boldsymbol{\gamma}}\left[ln(P(\gamma_i|\pi_i))\right] + ln\left(P(\boldsymbol{\pi})\right) + C \\
= \;& \sum_{i=1}^{p} \left(Prob(\gamma_i = 1)ln(\pi_i) + Prob(\gamma_i = 0)ln(1-\pi_i)\right) + \\
& \sum_{q=1}^{Q} \left((a_q - 1)ln(\pi_q) + (b_q - 1)ln(1-\pi_q)\right) + C \\
\approx \;& \sum_{i=1}^{p} \left(\widehat{\gamma_i} ln(\pi_i)) + (1-\widehat{\gamma_i})ln(1-\pi_i)\right) + \sum_{q=1}^{Q} \left((a_q - 1)ln(\pi_q) + (b_q - 1)ln(1-\pi_q)\right) + C,
\end{aligned} \tag{13}$$

where $\{\widehat{\gamma_i} = E[\gamma_i]\}$ are estimated by MCMC, and C is a constant free of $\boldsymbol{\pi}$.

Similarly, because the posterior distributions of $\{\pi_q;\, q = 1, \ldots, Q\}$ are also disjoint, the expected log-posterior-likelihood function for π_q is given by

$$l_{\pi_q} = \sum_{j_q=1}^{m_q} \left[\widehat{\gamma_{j_q}} ln(\pi_q) + (1 - \widehat{\gamma_{j_q}})ln(1-\pi_q)\right] + (a_q - 1)ln(\pi_q) + (b_q - 1)ln(1-\pi_q) + C, \tag{14}$$

and the MAP for π_q is solved as

$$\widehat{\pi_q} = \frac{\sum_{j_q=1}^{m_q} \widehat{\gamma_{j_q}} + a_q - 1}{m_q + a_q + b_q - 2}.$$

3 Construct Confidence Intervals by Fisher Information

Fisher information of $(\boldsymbol{\pi}, \boldsymbol{\sigma^2})$ can be derived from the second derivatives of the respective expected log-posterior-likelihood functions as in (11) and (13). By the asymptotic-normality of MAP, as $n \to \infty$, the distribution of a MAP estimate $\widehat{\boldsymbol{\theta}}$ converges to a multivariate normal (MVN) distribution with mean equal to the true parameter value $\boldsymbol{\theta_0}$ and covariance matrix equal to the inverse of the Fisher information.

Therefore, the MAPs $\widehat{\boldsymbol{\sigma^2}}$ and $\widehat{\boldsymbol{\pi}}$ are converging to the following MVN distributions as $n \to \infty$,

$$\widehat{\boldsymbol{\sigma^2}} \longrightarrow MVN(\boldsymbol{\sigma^2_*}, \boldsymbol{I_{\sigma^2}}(\widehat{\boldsymbol{\sigma^2}})^{-1}),\ \widehat{\boldsymbol{\pi}} \longrightarrow MVN(\boldsymbol{\pi_*}, \boldsymbol{I_{\pi}}(\widehat{\boldsymbol{\pi}})^{-1}), \tag{15}$$

where $\boldsymbol{\sigma^2_*}$ and $\boldsymbol{\pi_*}$ are the true parameter values; $\boldsymbol{I_{\sigma^2}}(\widehat{\boldsymbol{\sigma^2}}) \approx -\frac{\partial^2 l(\boldsymbol{\sigma^2})}{\partial \boldsymbol{\sigma^2} (\partial \boldsymbol{\sigma^2})^T}|_{\widehat{\boldsymbol{\sigma^2}}}$; and $\boldsymbol{I_{\pi}}(\widehat{\boldsymbol{\pi}}) \approx -\frac{\partial^2 l(\boldsymbol{\pi})}{\partial \boldsymbol{\pi} \partial \boldsymbol{\pi}^T}|_{\widehat{\boldsymbol{\pi}}}$. Because of the mutual independence among $\{\sigma_q^2, \pi_q;\ q = 1, \ldots, Q\}$ (conditioning on the estimates of $\boldsymbol{\beta}$ and $E[\boldsymbol{\gamma}]$), the analytical forms for the second derivatives of $l_{\sigma_q^2}, l_{\pi_q}$ are

$$\frac{dl_{\sigma_q^2}}{d^2\sigma_q^2} = \sum_{j_q=1}^{m_q} \left(\frac{\widehat{\gamma_{j_q}}}{2(\sigma^2)^2} - \frac{\widehat{\gamma_{j_q}} \tau \widehat{\beta_{j_q}}^2}{(\sigma^2)^3} \right) + \frac{k_1 + 1}{(\sigma^2)^2} - \frac{2k_2}{(\sigma^2)^3},$$

$$\frac{dl_{\pi_q}}{d^2\pi_q} = -\frac{\sum_{j_q=1}^{m_q} \widehat{\gamma_{j_q}} + a_q - 1}{\pi_q^2} - \frac{n_q - \sum_{j_q=1}^{m_q} \widehat{\gamma_{j_q}} + b_q - 1}{(1 - \pi_q)^2}.$$

Then the Fisher informations of σ_q^2, π_q are given by

$$I(\sigma_q^2) = \frac{1}{(\sigma_q^2)^2} \left(\sum_{j_q=1}^{m_q} \widehat{\gamma_{j_q}} (\tau - 0.5) - (k_1 + 1) + \frac{2k_2}{\sigma_q^2} \right),$$

$$I(\pi_q) = \frac{\sum_{j_q=1}^{m_q} \widehat{\gamma_{j_q}} + a_q - 1}{\pi_q^2} + \frac{n_q - \sum_{j_q=1}^{m_q} \widehat{\gamma_{j_q}} + b_q - 1}{(1 - \pi_q)^2}.$$

The $(1 - \alpha)\%$ confidence intervals of σ_q^2, π_q can be constructed by

$$\widehat{\sigma_q^2} \pm Z_{\alpha/2} \sqrt{I(\widehat{\sigma_q^2})^{-1}},\ \widehat{\pi_q} \pm Z_{\alpha/2} \sqrt{I(\widehat{\pi_q})^{-1}}, \tag{16}$$

where $Z_{\alpha/2}$ is the upper $\alpha/2$ quantile of the standard normal distribution $N(0, 1)$.

4 Compare Enrichment among Multiple Groups

With the MAPs of (π_q, σ_q^2) and corresponding standard errors, we can easily compare the enrichment among multiple groups. Take the case with two annotation groups for an example, the 95% confidence intervals of the quantities $ln(\pi_1/\pi_2)$, $ln(\sigma_1^2/\sigma_2^2)$ can be easily approximated by Fieller's theorem [3] (if variables $a \sim N(a_0, \sigma_a^2)$, $b \sim N(b_0, \sigma_b^2)$, then $ln\left(a/b\right) \sim N\left(ln\left(a_0/b_0\right), \sigma_a^2/a_0^2 + \sigma_b^2/b_0^2\right)$), and then can be used to test whether or not the enrichment is significantly different between two groups (i.e. whether or not the 95% confidence intervals of $ln(\pi_1/\pi_2)$, $ln(\sigma_1^2/\sigma_2^2)$ overlap 0). Moreover, with the approximated variance of the log-ratio by Fieller's theorem, we can calculate a P-value for the null hypothesis that the log-ratio equals 0. For example, the P-value for testing the null hypothesis $ln(\pi_1/\pi_2) = 0$ vs. the alternative hypothesis $ln(\pi_1/\pi_2) \neq 0$ can be calculated by

$$2\left(1 - \Psi\left(\frac{|ln(\widehat{\pi_1}/\widehat{\pi_2})|}{sd(ln(\pi_1/\pi_2))}\right)\right),$$

where Ψ is the probability distribution function of $N(0,1)$, $(\widehat{\pi_1}, \widehat{\pi_2})$ are MAPs, and $sd(ln(\pi_1/\pi_2))$ is the standard deviation of $ln(\pi_1/\pi_2)$.

For the case with multiple annotation groups, we can calculate similar quantities to compare the estimates by each group vs. the genome-wide average. That is, for causal probability, $ln(\pi_q/\pi_{avg})$ is used to test whether or not the causal probability of group q is significantly different from the overall average, where $\pi_{avg} = \sum_{q=1}^{Q} w_q\pi_q$, $w_q = \frac{m_q}{\sum_{q=1}^{Q} m_q}$ (m_q is the number of variants of annotation q). For the effect-size variance, a similar quantity $ln(\sigma_q^2/\sigma_{avg}^2)$ is used, where $\sigma_{avg}^2 = \sum_{q=1}^{Q} f_q\sigma_q^2$ is the weighted average of effect-size variances with weights given by $f_q = \frac{m_q\pi_q}{\sum_{q=1}^{Q} m_q\pi_q}$ ($m_q\pi_q$ is the expected number of associations in annotation category q). Again, the hypothesis tests for comparing enrichment among multiple groups can be easily performed, because the approximated 95% confidence intervals of these log-ratios can be easily obtained by Fieller's theorem [3].

In addition, we can approximate the enrichment-fold π_1/π_2 by $\exp(ln(\pi_1/\pi_2))$, and σ_1^2/σ_2^2 by and $\exp(ln(\sigma_1^2/\sigma_2^2))$.

5 Convergence Diagnosis

We used the potential scale reduction factor (PSRF) [4] to quantify the mixing property of MCMC algorithms. With multiple MCMC chains, the PSRF for a parameter is basically the ratio between the overall estimated parameter variance and the within-chain variance. A PSRF value within $(0.9, 1.2)$ suggests that the MCMC algorithm has good mixing property

and posterior samples converge. For example, in Figure S2, we present the PSRFs for the $E[\gamma_i]$ of top 58 variants with P-values $< 5 \times 10^{-8}$ in the WTCCC GWAS of Chrohn's disease [1]. We can see that about half of the 58 variants had PSRFs > 1.2 by the standard MCMC algorithm as used in GEMMA [7], while all PSRFs by our MCMC algorithm all fall within $(0.9, 1.2)$, suggesting greatly improved mixing property due to the refined proposal distribution and relatively small block-sizes.

6 Challenges for Extending bfGWAS for Overlapped and Quantitative Annotations

Theoretically, this Bayesian hierarchical model can be easily extended for analyzing overlapped categorical and quantitative annotations, by assuming the following logistic model for the π_i in model (1),

$$logit(\pi_i) = \alpha_0 + \boldsymbol{A_i}^T \boldsymbol{\alpha}. \tag{17}$$

In the logistic model (17), $\boldsymbol{A_i}$ is the quantitative annotation vector (with binary values for categorical annotations) for the ith variant, and $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_Q)$ is the vector of log-odds for all considered annotations. Independent normal distributions can be assumed as the hyper priors for the category-specific (enrichment) parameters $(\alpha_0, \boldsymbol{\alpha})$. With a large number of annotations, variable selection of annotations might even be integrated by assuming independent point-normal priors for $\boldsymbol{\alpha}$.

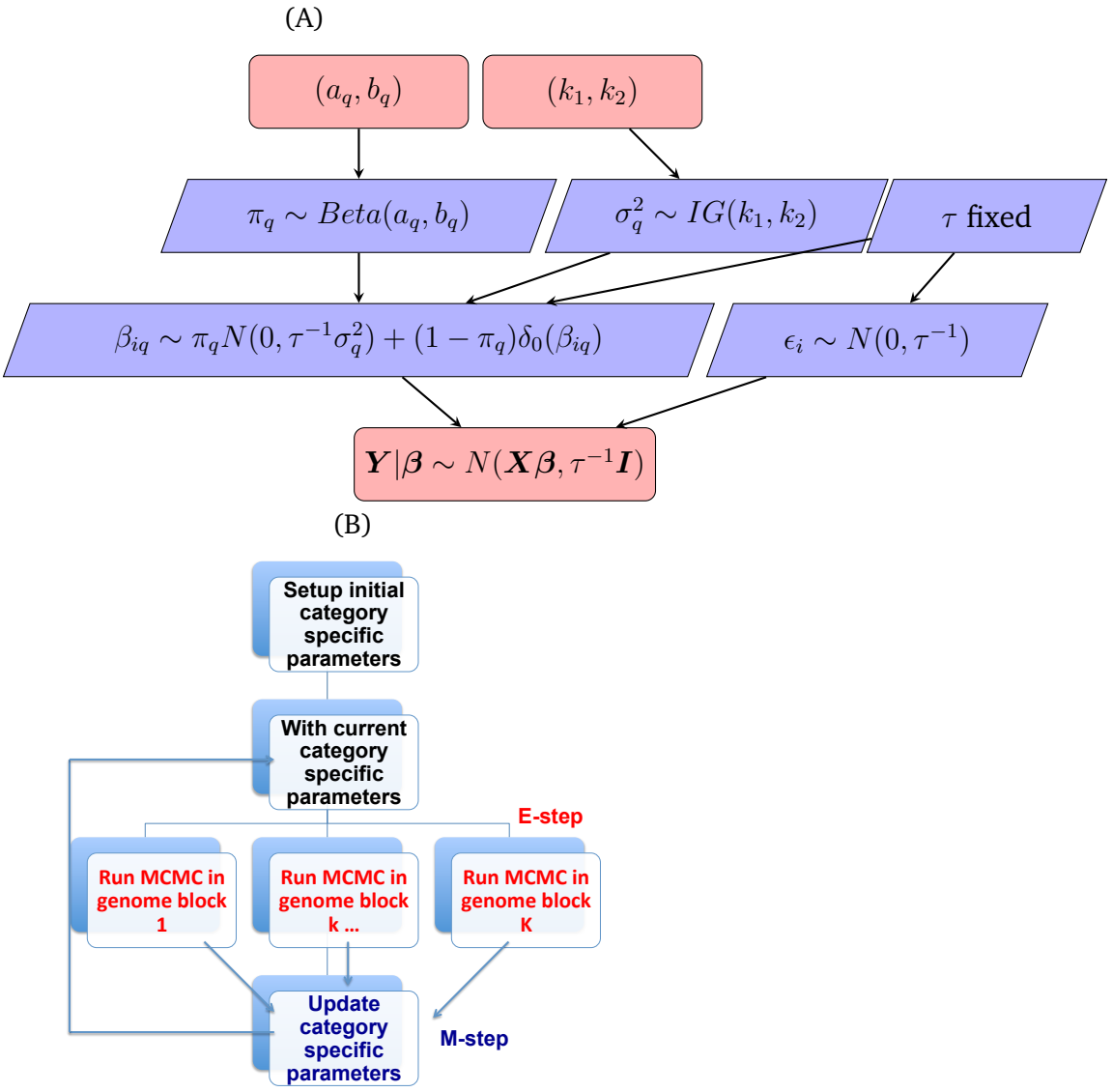
Conditioning on values for $(\alpha_0, \boldsymbol{\alpha})$, the MCMC algorithm (Section 2.2) can be implemented similarly per block in the E-step. However, in the M-step, analytical formulas are no longer available for the posterior MAPs of $(\alpha_0, \boldsymbol{\alpha})$. In preliminary analysis, we found that the false positive rate was inflated due to over estimated π_i, which is due to the difficulties of estimating $(\alpha_0, \boldsymbol{\alpha})$. We are still exploring an appropriate approach to effectively control the false positive rate for this extension.

7 Software

Software implementing this Bayesian hierarchical model with the EM-MCMC algorithm, referred as Bayesian Functional Genome-wide Association Study (bfGWAS), is now available at GitHub (`https://github.com/yjingj/bfGWAS`). Within the software, the E-step (MCMC algorithm) is written in C++ language; the M-step is written in an R script; and both steps are wrapped together (enabling parallel computation) through submitting jobs by a Makefile that is generated by a Perl script.

Supplemental Figures

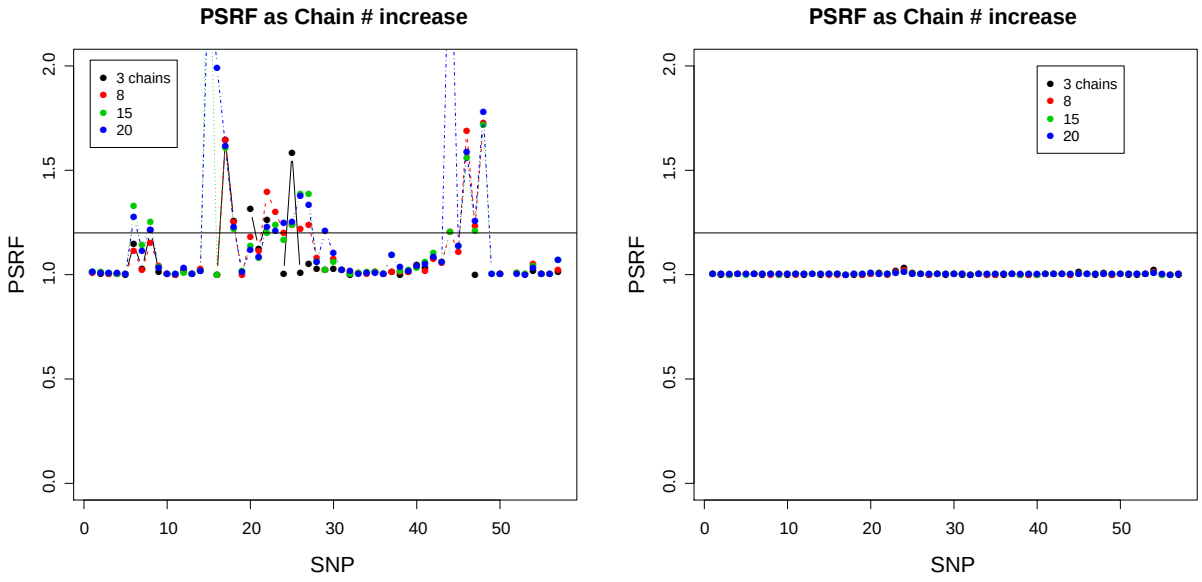
Figure S 1: Flowcharts of bfGWAS.



(A) Hierarchical Bayesian variable selection model; (B) EM-MCMC algorithm.

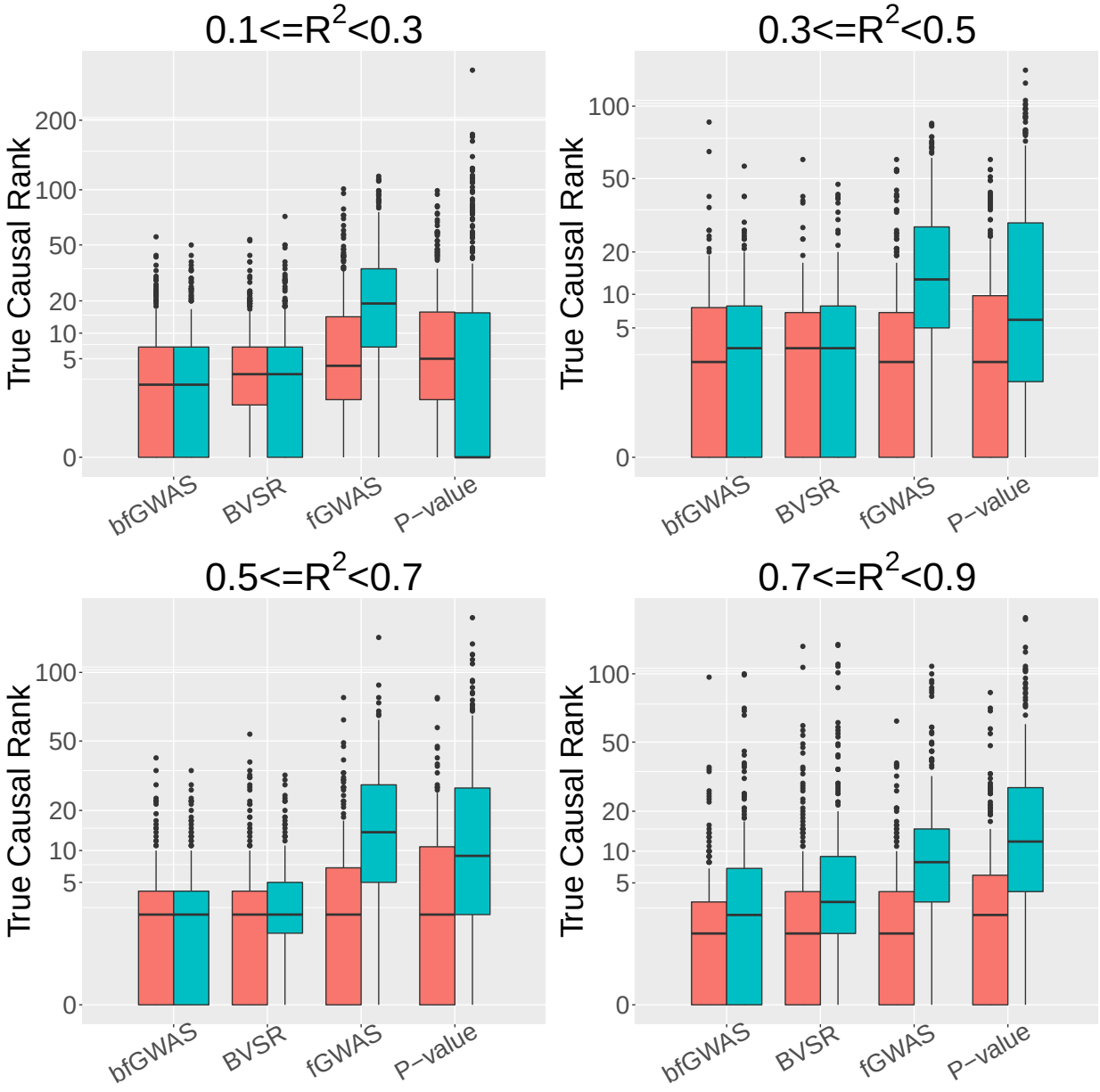
Figure S 2: Plots of the potential scale reduction factors (PSRF).

(A) (B)



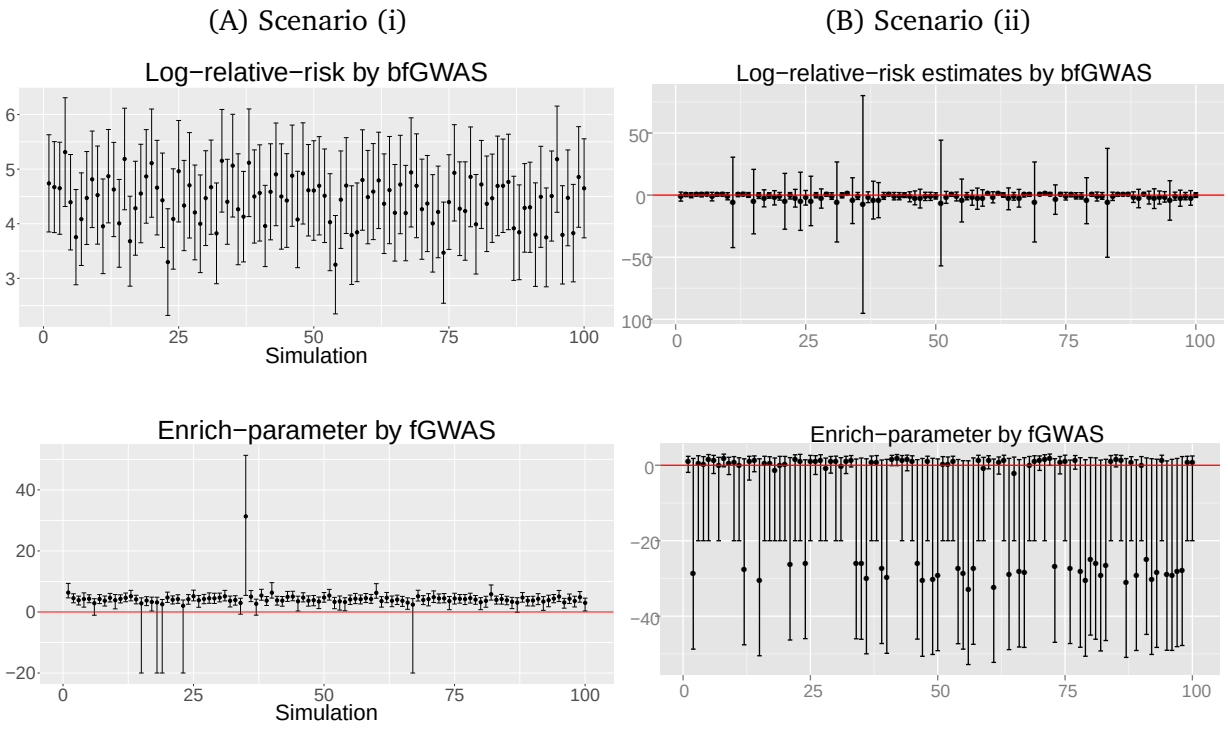
Potential scale reduction factors (PSRF) of the Bayesian posterior inclusion probabilities of 58 top marginally significant SNPs (WTCCC GWAS of Crohn's disease) with 3, 8, 15, and 20 MCMC chains, where PSRF within $(0.9, 1.2)$ suggests good mixing property. (A) Standard MCMC algorithm as used in GEMMA; (B) Our MCMC algorithm.

Figure S 3: Prioritization ranks of the true causal SNP1 (pink) and SNP2 (cyan).



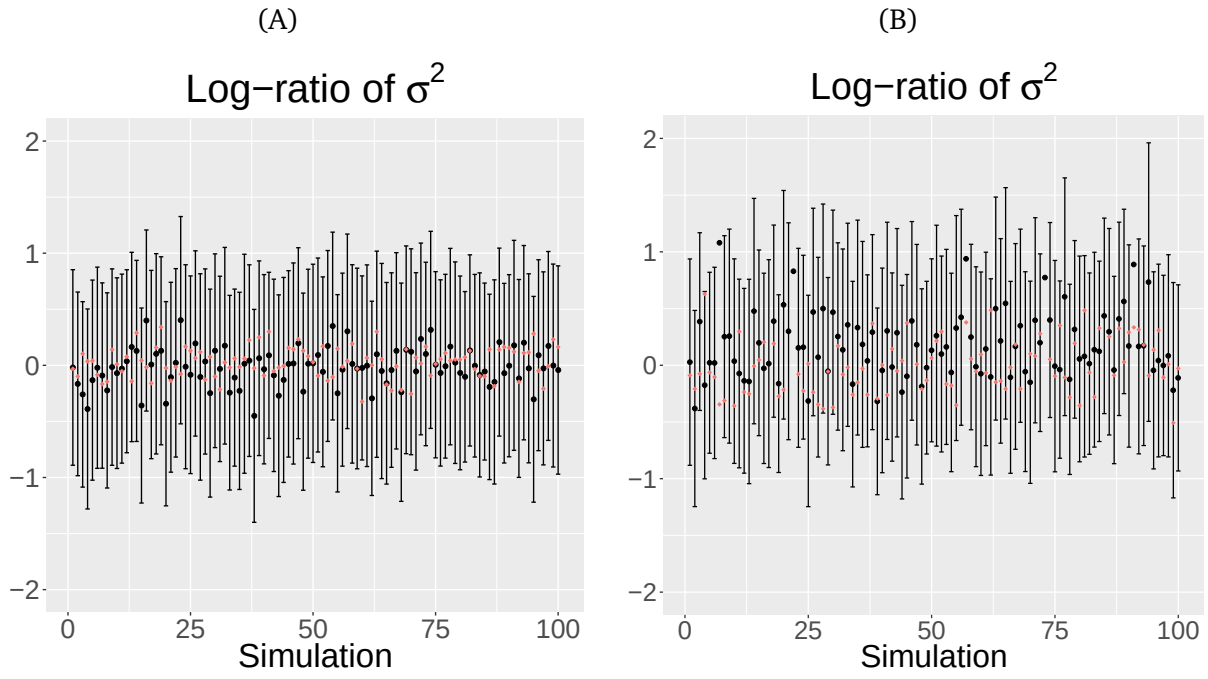
Ranks prioritized by bfGWAS, the standard Bayesian variable selection regression model (BVSR), fGWAS, P-value (single variant test for SNP1 and conditional analysis for SNP2), stratified by the R2 (LD) between SNP1 and SNP2. Here higher ranks (smaller numeric values) suggest higher power.

Figure S 4: Estimates of the log-relative-risk $ln(\pi_0/\pi_1)$ by bfGWAS and the enrich-parameter by fGWAS, along with 95% confidence intervals.



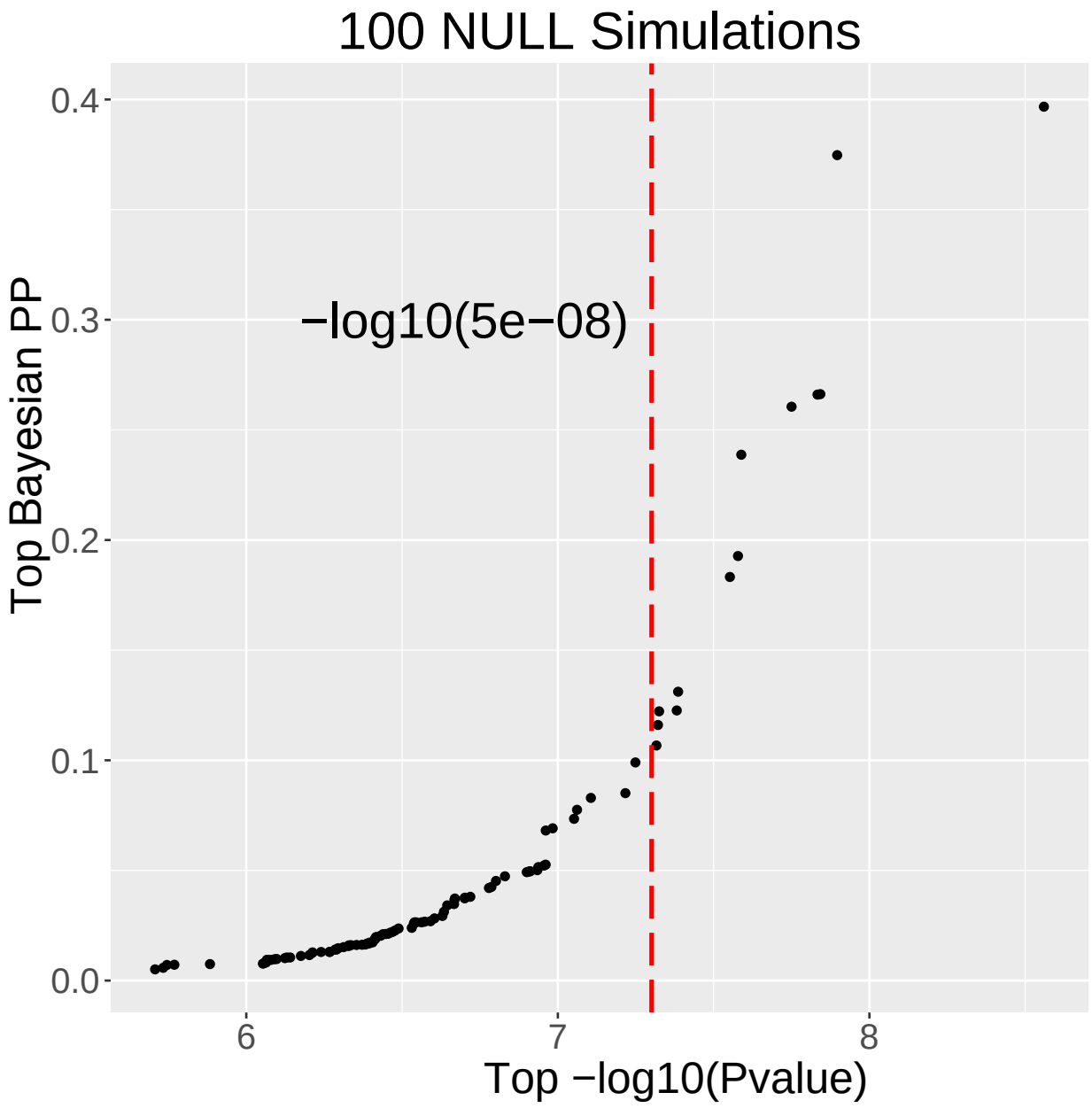
(A) Simulation scenario (i) with enrichment in coding and Scenario; (B) Simulation scenario (ii) with no enrichment. No enrichment is estimated when the 95% confidence interval covers 0, while enrichment for coding is estimated with the 95% confidence interval above 0.

Figure S 5: Estimates of the log-ratio of effect-size variances $ln(\sigma_0^2/\sigma_1^2)$ by bfGWAS, along with 95% confidence intervals.



(A) Simulation scenario (i) with enrichment in coding; (B) Simulation scenario (ii) with no enrichment. Note that the effect-sizes of both groups in scenarios (i) and (ii) were simulated from the same normal distribution, thus the 95% confidence intervals covering 0 suggest that bfGWAS estimates similar effect-size variances between two categories.

Figure S 6: Sorted top bfGWAS PPs versus sorted top $-log10$(P-values) of single variant tests.



Results of 100 GWASs with AMD genotype data and permuted phenotypes. Note that the P-value 5×10^{-8} roughly corresponds to bfGWAS posterior inclusion probability (PP) 0.1068.

Figure S 7: Manhattan plot highlighting AMD GWAS signals with BVSR $PP > 0.1068$.

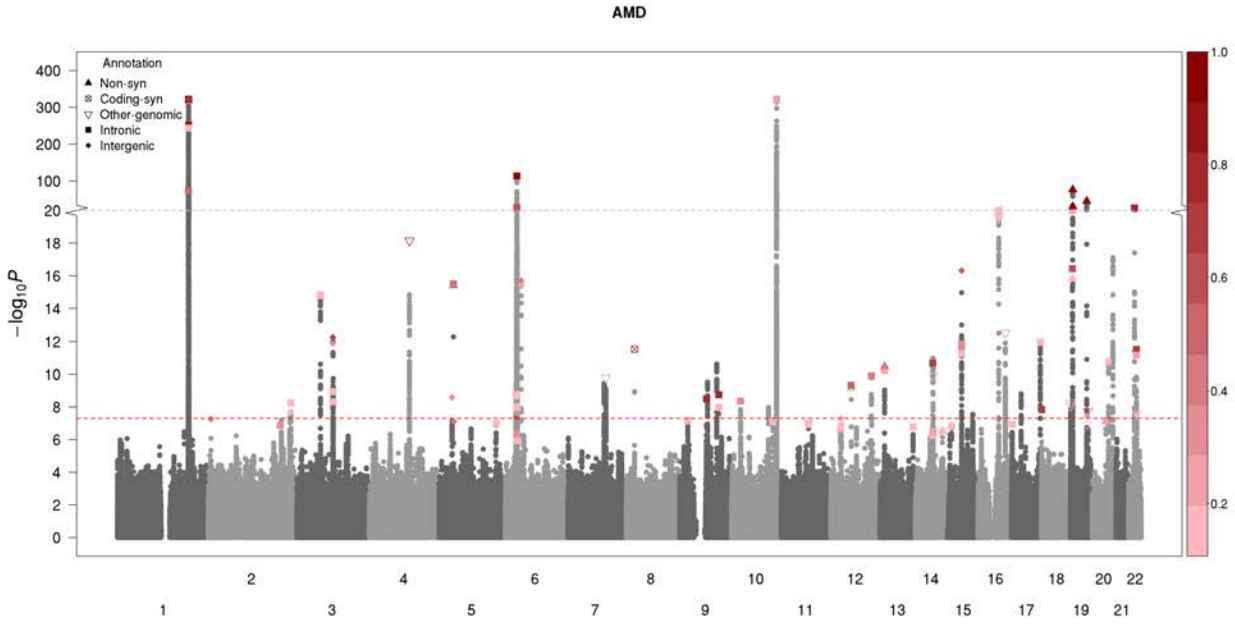
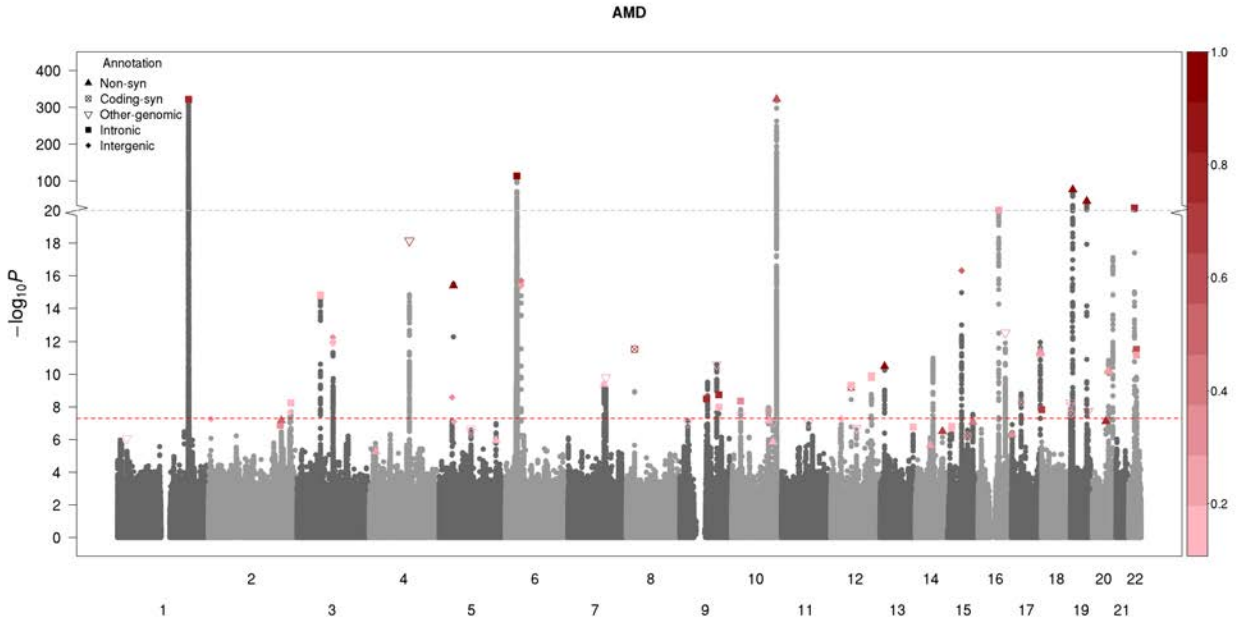
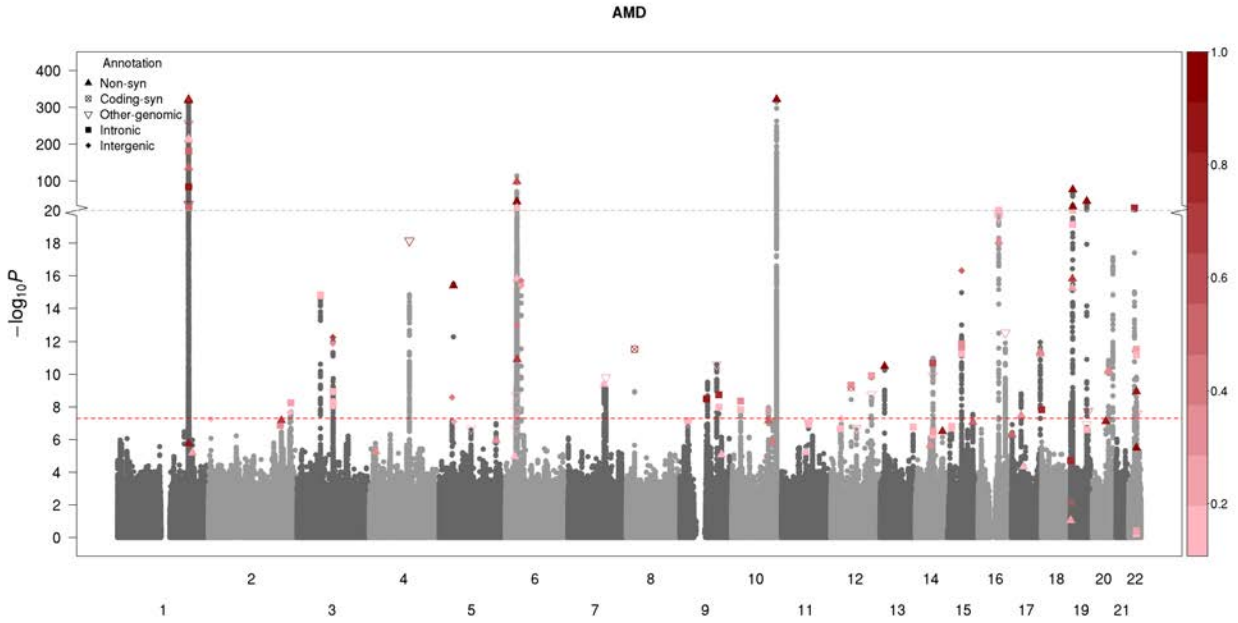


Figure S 8: Manhattan plots highlighting AMD GWAS signals by accounting for gene-based annotations.

(A)

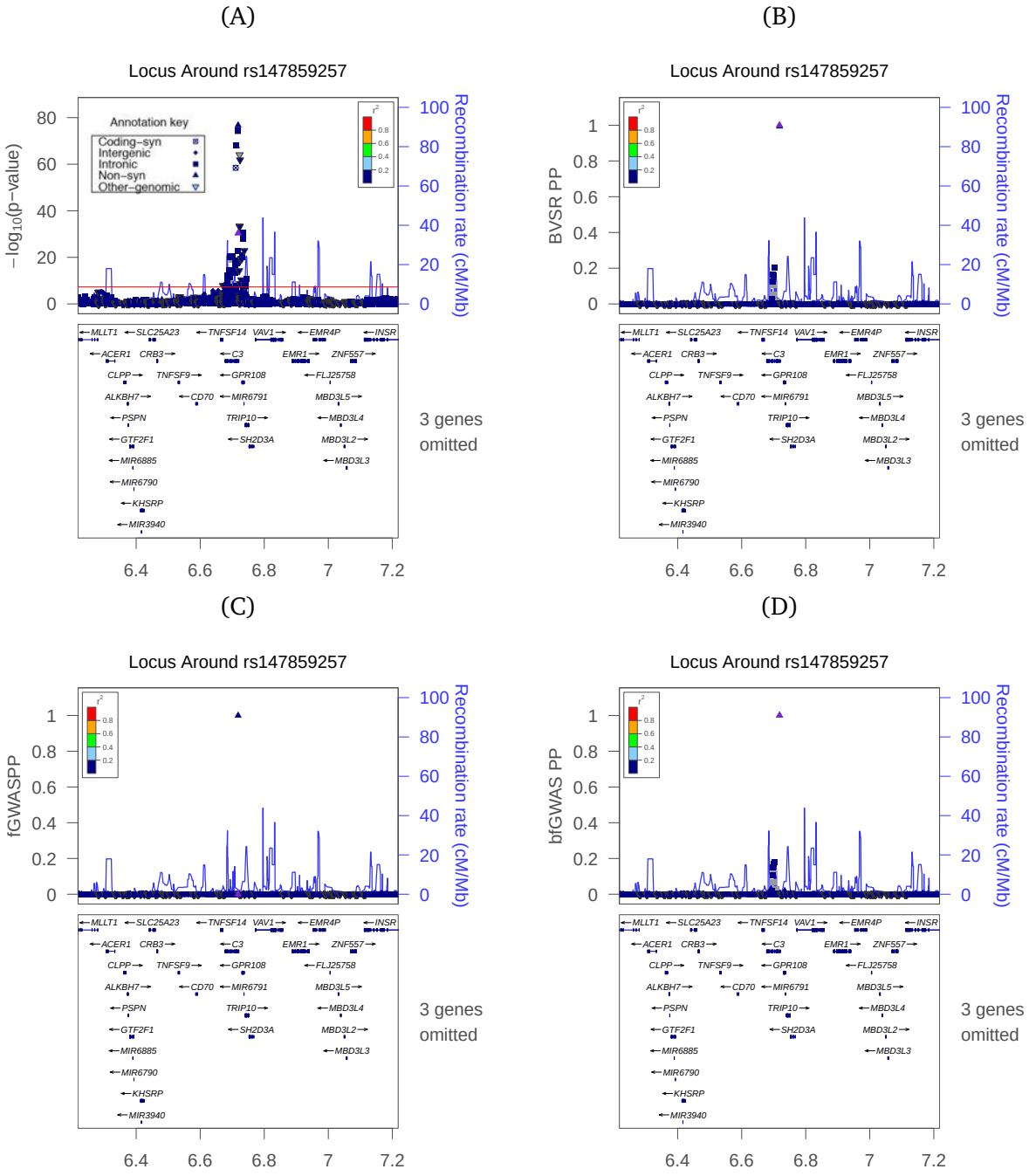


(B)



(A) Highlighting signals with fGWAS posterior association probability (PP) > 0.1068 are colored; (B) Highlighting signals with bfGWAS PP > 0.1068.

Figure S 9: LocusZoom plots of region *CHR19:6218146-7218146*.



(A) P-values by single variant tests; (B) BVSR PPs; (C) fGWAS PPs; (D) bfGWAS PPs. The purple triangle in (B, D) denotes the variant *rs147859257*; the blue triangle in (A, C) denotes the top significant variant by single variant tests *rs2230199*.

Figure S 10: Enrichment analysis results with varying prior means as well as starting values $(10^{-6}, 5 \times 10^{-6}, 10^{-5})$ for π_q, and varying starting values $(10, 5, 1)$ for σ_q^2.

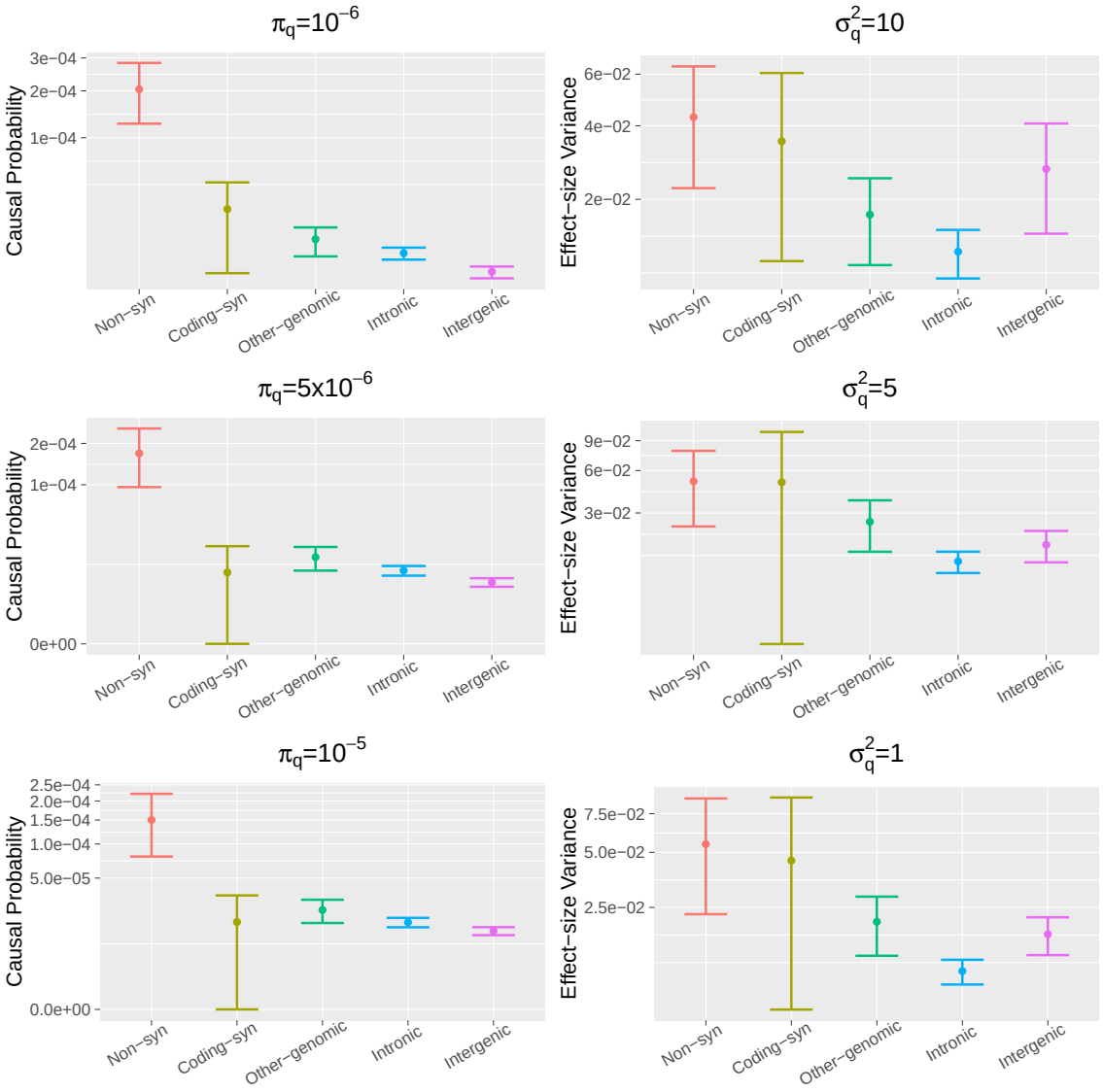


Figure S 11: fGWAS enrichment estimates with 95% error bars.

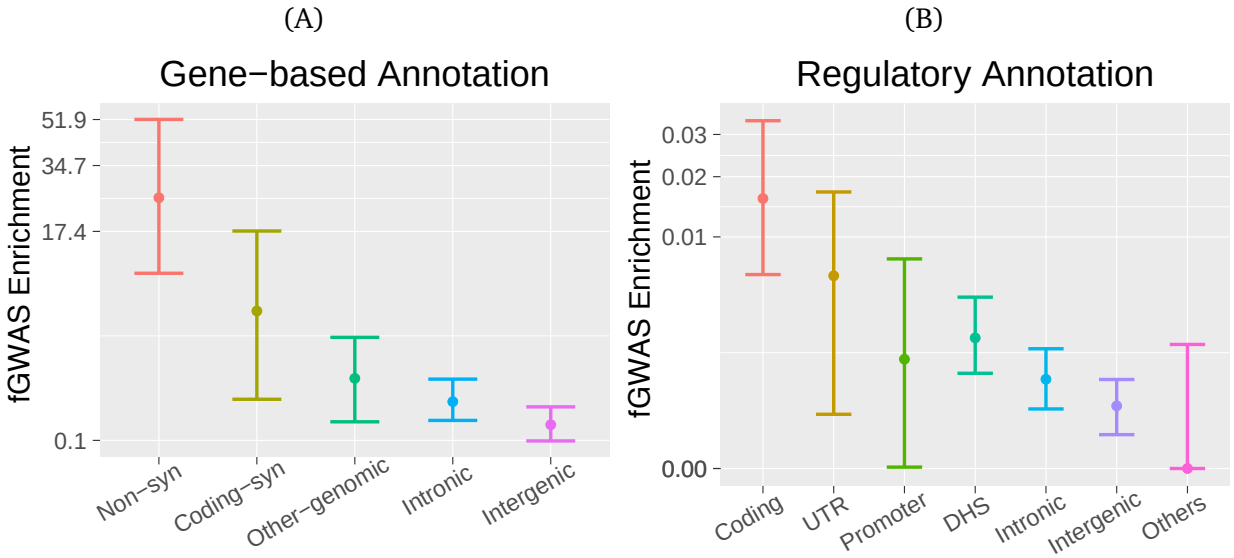
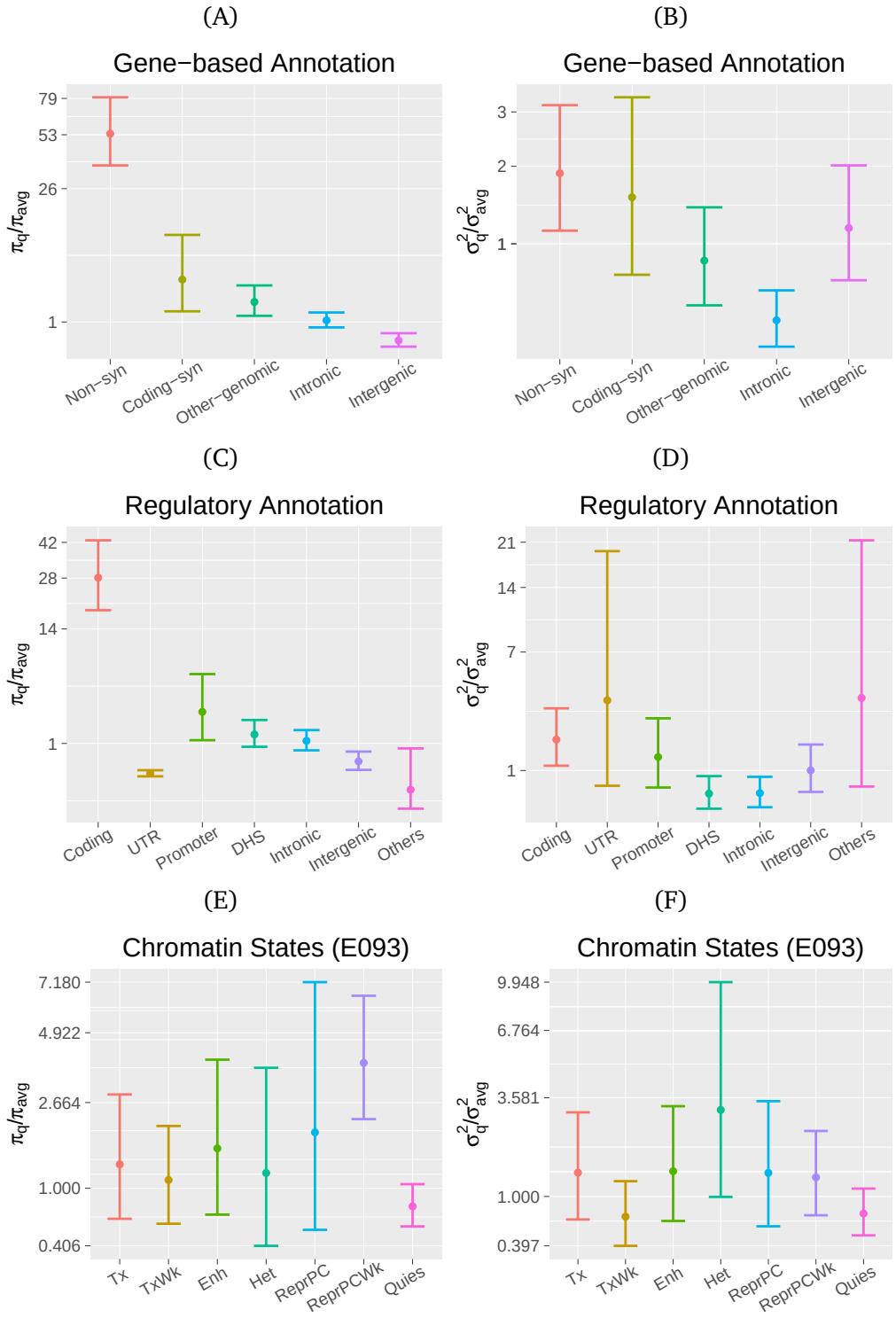


Figure S 12: Ratios of enrich parameters versus the respective genome-wide averages, along with 95% confidence intervals.



(A, C, E) Causal probability ratios (π_q/π_{avg}); (B, D, F) Effect-size variance ratios ($\sigma^2_q/\sigma^2_{avg}$).

Figure S 13: Enrichment analysis results for the AMD GWAS data with chromatin states profiled with respect to the epigenome of fetal thymus (E093).

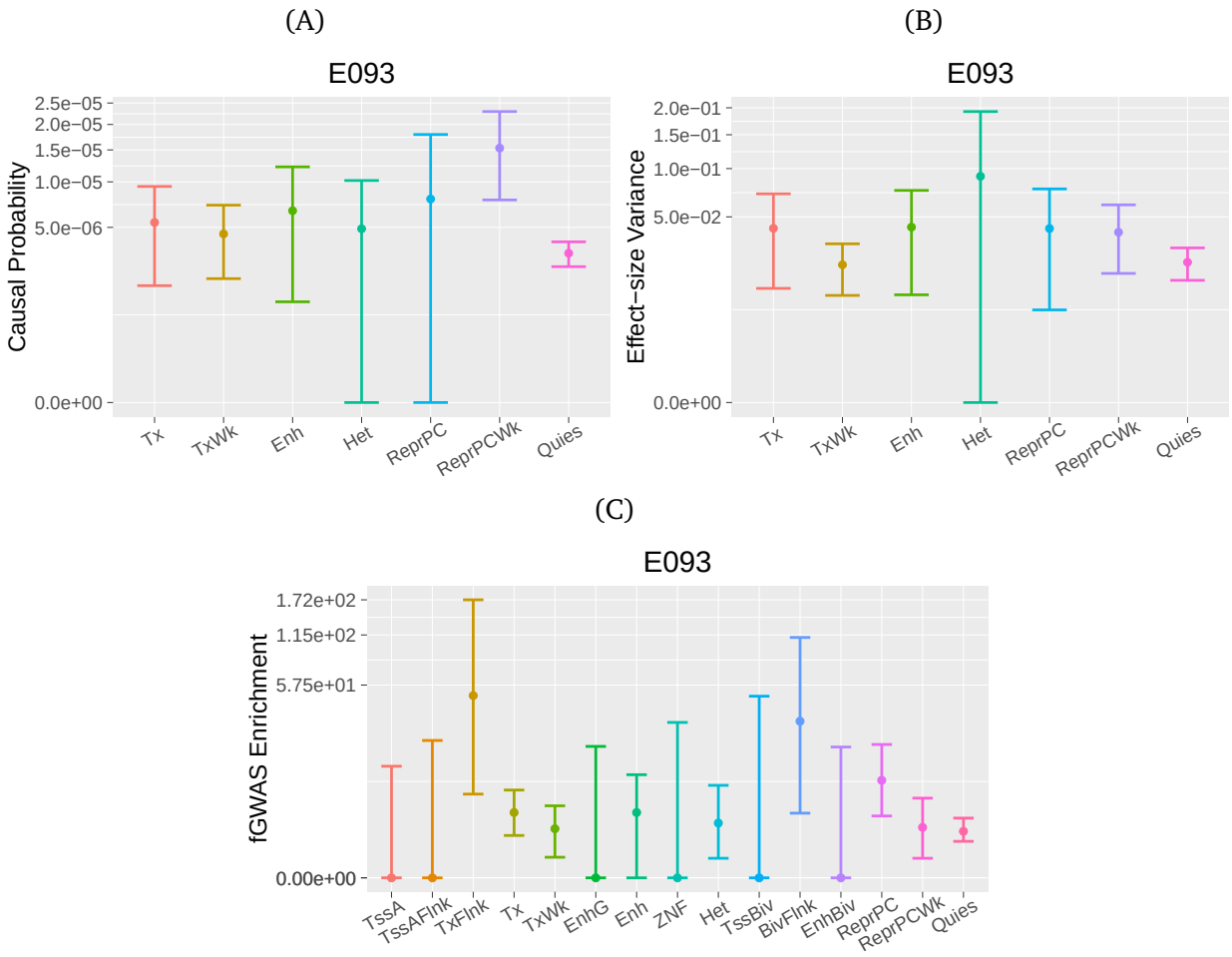


Figure S 14: Manhattan plot highlighting MGI GWAS signals of skin cancer with BVSR PP > 0.1068.

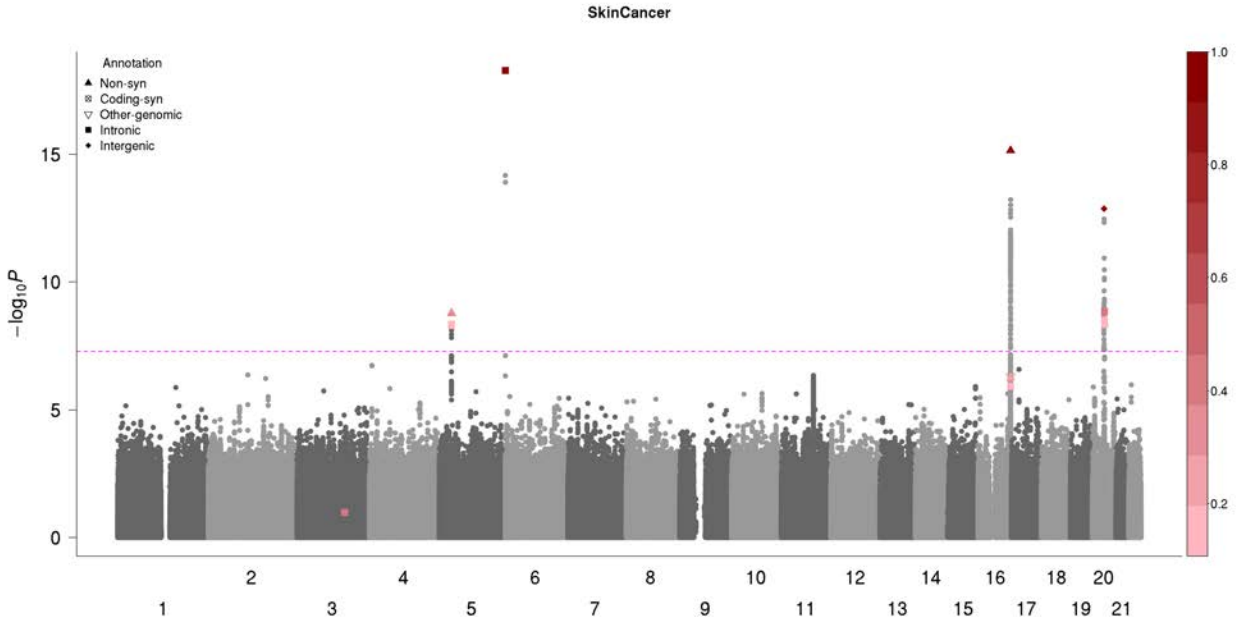
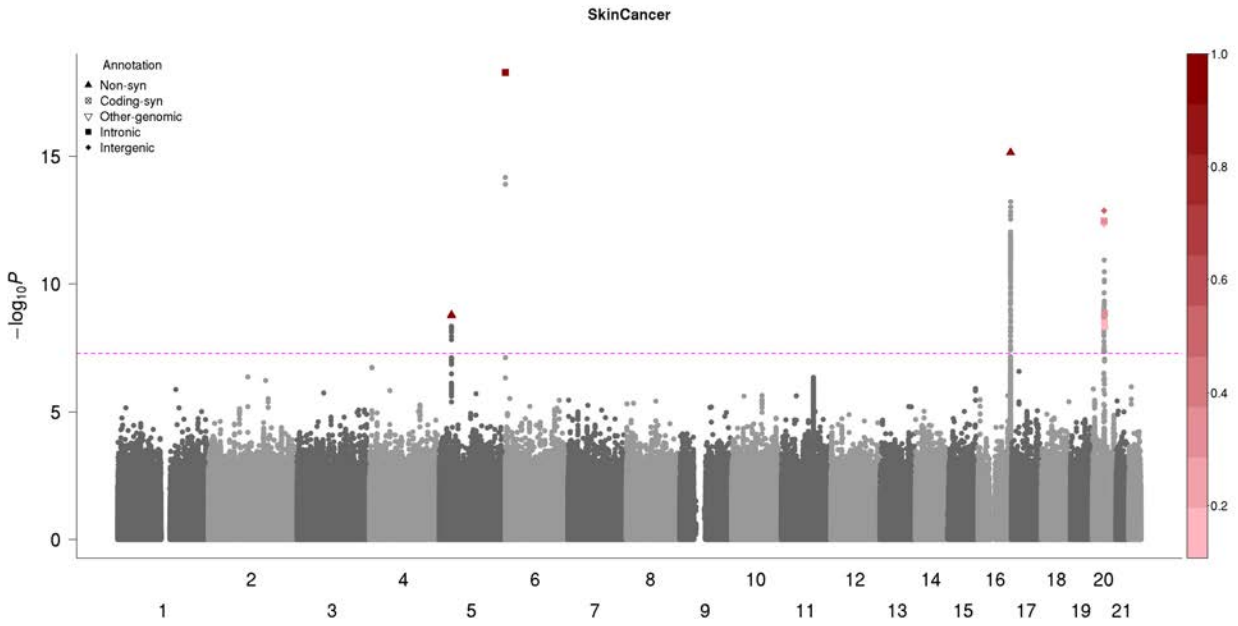
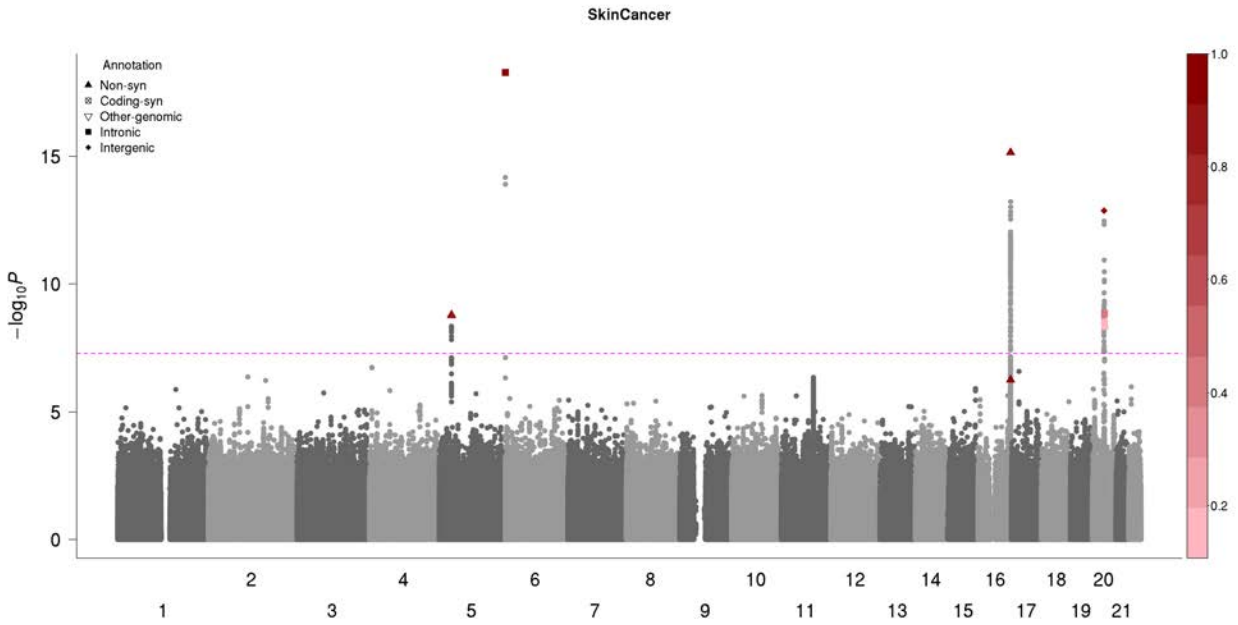


Figure S 15: Manhattan plots highlighting MGI GWAS signals of skin cancer by accounting for gene-based annotations.

(A)



(B)



(A) Highlighting signals with with fGWAS PP > 0.1068; (B) Highlighting signals with bfGWAS PP > 0.1068. Variants with PP > 0.1068 are plotted in different shapes with respect to gene-based annotations.

Figure S 16: Enrichment analysis results of the MGI GWAS of skin cancer, accounting for gene-based annotations.

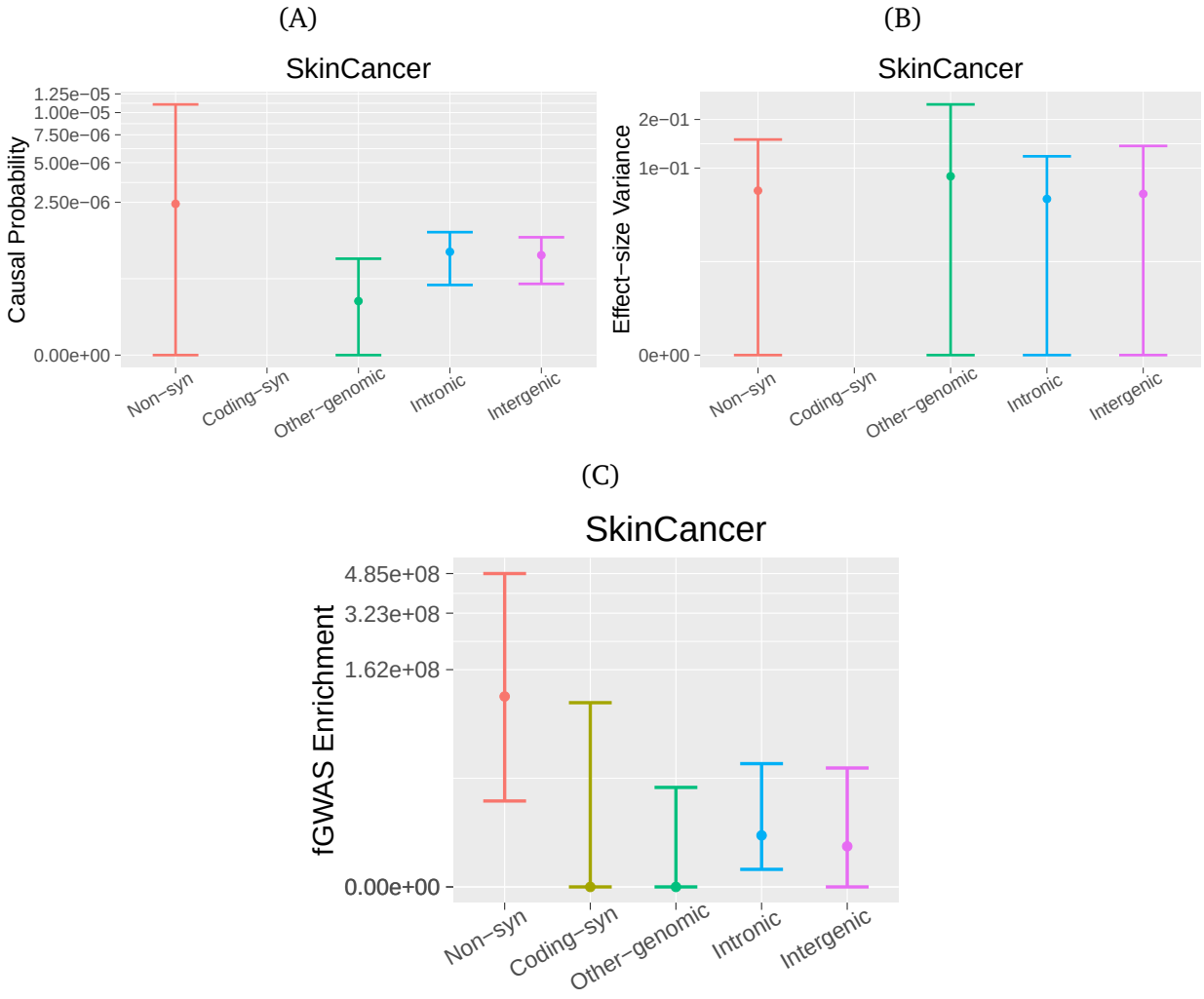
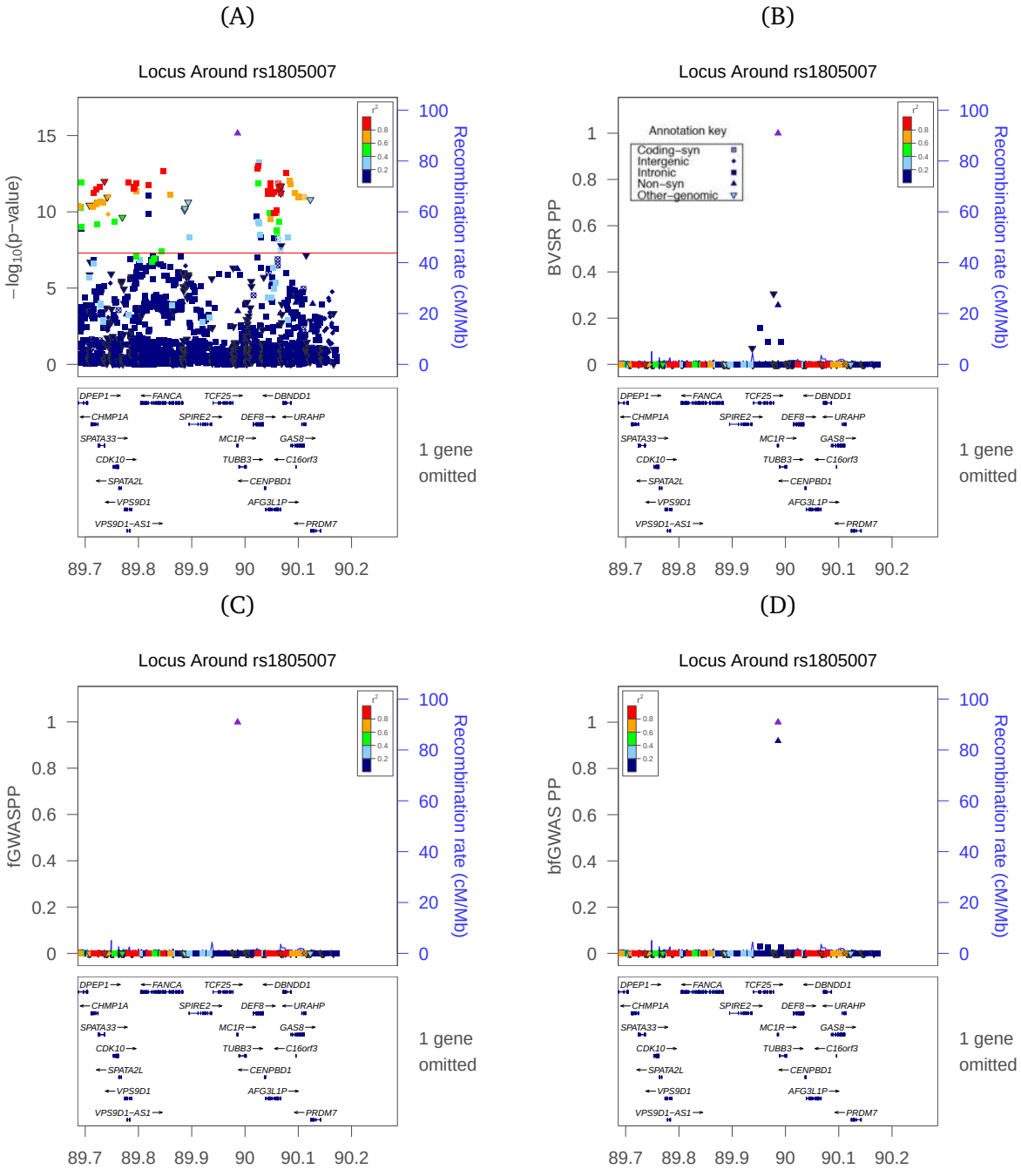


Figure S 17: LocusZoom plots in the region of *CHR16:89686117-90172696*.



(A) P-values by single variant tests; (B) BVSR PPs; (C) fGWAS PPs; (D) bfGWAS PPs. The purple triangle denotes the variant *rs1805007*.

Supplemental Tables

Table S1: Classification of gene-based functional annotations.

Native gene-based functional annotations	Annotation categories considered in the analysis
frameshift, frameshift-near-splice	Non-synonymous
splice-acceptor, splice-donor,	
stop-gained, stop-gained-near-splice, stop-lost	
missense, missense-near-splice	
synonymous-near-splice, non-coding-exon-near-splice, coding-near-splice, coding-unknown-near-splice, intron-near-splice	
coding, coding-unknown, synonymous, nc-transcript-variant	Coding-synonymous
intronic	Intronic
intergenic, NAs	Intergenic
3-prime-UTR, 5-prime-UTR,	Other-genomic
downstream-gene, upstream-gene, non-coding-exon	

Table S2: Compare results by P-value, fGWAS, and bfGWAS in the 34 known AMD loci, accounting for gene-based annotations.

Known 34 Loci				Top significant variant by P-value					Bayesian Regional-PP	fGWAS Regional-PP
Locus name	Chr	Start	End	dbSNPID	Chr:Position	MAF	P-value	Anno		
CFH	1	195,679,832	197,768,053	rs10922109	1:196,704,632	0.329	$<9\times10^{-321}$	intronic	1.000	1.000
COL4A3	2	227,573,015	228,592,110	rs11884770	2:228,086,920	0.731	5.6×10^{-9}	intronic	0.984	0.986
ADAMTS9-AS	3	64,199,445	65,230,121	rs62247658	3:64,715,155	0.551	1.4×10^{-15}	intronic	0.978	1.000
COL8A1	3	98,551,114	100,381,567	rs140647181	3:99,180,668	0.019	5.4×10^{-13}	intergenic	1.000	0.999
CFI	4	110,126,506	111,185,820	rs10033900	4:110,659,067	0.506	7.1×10^{-19}	downstream	1.000	1.000
C9	5	38,699,134	39,831,894	rs62358361	5:39,327,888	0.012	3.1×10^{-16}	intronic	1.000	1.000
PRLR/SPEF2	5	34,769,332	36,493,378	rs114092250	5:35,494,448	0.018	2.5×10^{-9}	intergenic	0.961	0.987
C2/CFB/SKIV2L	6	30,505,490	33,238,589	rs116503776	6:31,930,462	0.120	2.1×10^{-114}	intronic	1.000	1.000
VEGFA	6	43,305,296	44,329,629	rs943080	6:43,826,627	0.518	2.0×10^{-16}	intergenic	1.000	1.000
KMT2E/SRPK2	7	104,081,402	105,563,372	rs1142	7:104,756,326	0.357	1.5×10^{-10}	downstream	0.999	0.999
PILRB/PILRA	7	99,394,940	100,611,776	rs7803454	7:99,991,548	0.199	3.6×10^{-10}	intronic	0.999	0.999
TNFRSF10B	8	22,582,971	23,588,984	rs79037040	8:23,080,971	0.534	2.9×10^{-12}	nc-transcript	1.000	0.999
MIR6130/RORB	9	75,935,160	77,189,752	rs10781180	9:76,615,662	0.683	3.0×10^{-10}	intergenic	0.997	0.999
TRPM3	9	72,938,605	73,946,180	rs7150714	9:73,438,605	0.584	3.2×10^{-9}	intronic	0.929	0.999
TGFBR1	9	101,358,102	102,431,769	rs1626340	9:101,923,372	0.199	2.3×10^{-11}	intergenic	1.000	0.999
ABCA1	9	107,139,414	108,167,147	rs2740488	9:107,661,742	0.265	1.7×10^{-9}	intronic	0.963	0.985
ARHGAP21	10	24,360,361	25,556,538	rs12357257	10:24,999,593	0.232	4.3×10^{-9}	intronic	0.962	0.986
ARMS2/HTRA1	10	123,702,126	124,735,355	rs3750846	10:124,215,565	0.316	$<9\times10^{-321}$	intronic	1.000	1.000
RDH5/CD63	12	55,615,585	56,713,297	rs3138141	12:56,115,778	0.214	4.7×10^{-10}	intronic	0.034	0.999
ACAD10	12	110,919,995	113,502,935	rs73205633	12:112,357,085	0.019	1.2×10^{-10}	intergenic	0.997	0.999

Known 34 Loci				Top significant variant by P-value					Bayesian Regional-PP	fGWAS Regional-PP
Locus name	Chr	Start	End	dbSNPID	Chr:Position	MAF	P-value	Anno		
B3GALTL	13	31,242,232	32,339,274	rs9564692	13:31,821,240	0.288	3.2×10^{-11}	splice	1.000	0.999
RAD51B	14	68,227,506	69,550,783	rs1956526	14:68,799,787	0.650	$1..0\times10^{-11}$	intronic	1.000	0.999
LIPC	15	58,171,721	59,242,418	rs2414577	15:58,680,638	0.365	4.8×10^{-17}	nc-transcript	1.000	1.000
CETP	16	56,485,514	57,506,829	rs5817082	16:56,997,349	0.248	1.7×10^{-21}	intronic	1.000	1.000
CTRB2/CTRB1	16	74,732,528	76,017,115	rs72802342	16:75,234,872	0.073	2.8×10^{-13}	downstream	1.000	1.000
TMEM97/VTN	17	26,092,946	27,240,139	rs11080055	17:26,649,724	0.524	1.5×10^{-9}	intronic	0.996	0.998
NPLOC4/TSPAN10	17	79,015,509	80,186,552	rs6565597	17:79,526,821	0.390	1.0×10^{-12}	intronic	1.000	0.999
C3	19	5,311,717	7,224,340	rs2230199	19:6,718,387	0.764	1.7×10^{-77}	missense	1.000	1.000
CNN2	19	523,867	1,533,360	rs10422209	19:1,026,318	0.132	5.5×10^{-9}	upstream	0.970	0.993
APOE	19	44,892,254	46,313,830	rs429358	19:45,411,941	0.118	3.3×10^{-46}	missense	1.000	1.000
MMP9	20	44,114,991	45,160,699	rs142450006	20:44,614,991	0.132	1.4×10^{-11}	intergenic	1.000	0.999
C20orf85	20	56,084,276	57,174,034	rs117739907	20:56,652,781	0.062	7.8×10^{-18}	intergenic	1.000	1.000
SYN3/TIMP3	22	32,546,536	33,613,375	rs5754227	22:33,105,817	0.123	2.0×10^{-27}	intronic	1.000	1.000
SLC16A8	22	37,795,271	39,003,972	rs8135665	22:38,476,276	0.205	2.9×10^{-12}	intronic	1.000	0.999

Table S3: AMD risk variants identified by bfGWAS in the 34 known loci, accounting for gene-based annotations.

Signal number	Reside/Nearby Gene	dbSNPID	Chr:Position	Anno	MAF	bfGWAS PP	Effect-size	P-value
1.1	*CFH*	rs800292	1:196,642,233	missense	0.183	0.997	-0.312	2.4×10^{-319}
1.2	*CFH*	rs10922094	1:196,661,505	intronic	0.530	1.000	-0.214	$< 9.0\times10^{-321}$
1.3	*CFHR1*	rs605082	1:196,801,917	downstream	0.353	0.518	-0.092	7.5×10^{-257}
1.4	*CFHR4*	rs58175074	1:196,820,080	intronic	0.158	0.792	-0.314	$< 9.0\times10^{-321}$
1.5	*CFHR4*	rs149032610	1:196,857,150	5'-UTR	0.015	1.000	0.195	6.6×10^{-38}
1.6	*CFHR4*	rs10494745	1:196,887,457	missense	0.134	0.526	0.092	7.4×10^{-137}
1.7	*CFHR2*	rs138579109	1:196,923,955	intronic	0.043	0.893	0.167	8.4×10^{-85}
1.8	*CFHR5*	rs35662416	1:196,967,354	missense	0.022	0.889	-0.122	5.8×10^{-6}
2	*COL4A3*	rs11884770	2:228,086,920	intronic	0.731	0.269	0.052	5.6×10^{-9}
3	*ADAMTS9-AS2*	rs7428936	3:64,710,850	intronic	0.448	0.167	-0.061	1.5×10^{-15}
4	*COL8A1*	rs140647181	3:99,180,668	intergenic	0.019	0.687	0.224	54×10^{-13}
5	*CFI*	rs10033900	4:110,659,067	downstream	0.506	0.999	-0.067	7.2×10^{-19}
6	*C9*	rs34882957	5:39,331,894	missense	0.012	0.998	0.278	4.0×10^{-16}
7	*PRLR/SPEF2*	rs114092250	5:35,494,448	intergenic	0.019	0.403	-0.174	2.5×10^{-9}
8.1	*C2/CFB*	rs4151667	6:31,914,024	missense	0.036	0.917	-0.279	1.4×10^{-44}
8.2	*SKIV2L/NELFE*	rs115270436	6:31,928,306	missense	0.071	0.633	-0.321	2.8×10^{-99}
8.3	*HLA-DQB1*	rs3891176	6:32,634,318	missense	0.159	0.726	0.153	1.2×10^{-11}
9	*VEGFA*	rs943080	6:43,826,627	intergenic	0.518	0.435	0.063	2.0×10^{-16}
10	*KMT2E/SRPK2*	rs1142	7:104,756,326	downstream	0.357	0.125	0.052	1.5×10^{-10}
11	*PILRB*	rs35986051	7:99,956,439	missense	0.139	0.193	0.075	4.0×10^{-10}
12	*TNFRSF10A*	rs79037040	8:23,082,971	nc-transcript	0.534	0.996	0.053	2.9×10^{-12}
13	*MIR6130/RORB*	rs10781182	9:76,617,720	intergenic	0.684	0.070	-0.052	3.0×10^{-10}
14	*TRPM3*	rs71507014	9:73,438,605	intronic	0.584	0.822	-0.046	3.2×10^{-9}
15	*TGFBR1*	rs10819635	9:101,864,510	upstream	0.186	0.137	-0.066	2.4×10^{-11}
16	*ABCA1*	rs2740488	9:107,661,742	intronic	0.266	0.756	-0.053	1.7×10^{-9}
17	*ARHGAP21*	rs12357257	10:24,999,593	intronic	0.232	0.318	0.053	4.3×10^{-9}
18	*ARMS2*	rs10490924	10:124,214,448	missense	0.316	0.996	0.474	$< 9.0\times10^{-321}$
19	*RDH5/CD63*	rs3138142	12:56,115,585	coding-syn	0.213	0.706	0.074	6.1×10^{-10}
20	*MAPKAPK5*	rs61941287	12:112,330,305	intronic	0.019	0.309	0.191	1.2×10^{-10}
21	*B3GLCT*	rs9564692	13:31,821,240	splice	0.288	0.942	-0.056	3.2×10^{-11}
22	*RAD51B*	rs2842339	14:68,986,999	intronic	0.899	0.243	-0.082	3.1×10^{-7}
23	*ALDH1A2*	rs2414577	15:58,680,638	intronic	0.366	0.501	-0.067	4.8×10^{-17}
24	*CETP*	rs1532625	16:57,005,301	splice	0.448	0.358	0.044	7.9×10^{-19}
25	*CTRB2*	rs72802342	16:75,234,872	downstream	0.360	0.297	-0.114	2.8×10^{-13}
26	*CTB-96E2.2/VTN*	rs704	17:26,694,861	missense	0.483	0.325	0.042	3.3×10^{-8}
27	*NPLOC4/TSPAN10*	rs6420484	17:79,612,397	missense	0.622	0.402	-0.055	4.0×10^{-12}
28.1	*FUT6/NRTN*	rs17855739	19:5,831,840	missense	0.044	0.681	-0.159	1.5×10^{-16}
28.2	*C3/CTD-3128G10.7*	rs147859257	19:6,718,146	missense	0.008	1.000	0.501	4.3×10^{-31}
28.3	*C3/CTD-3128G10.7*	rs2230199	19:6,718,387	missense	0.764	1.000	-0.172	1.7×10^{-77}

Signal number	Reside/Nearby Gene	dbSNPID	Chr:Position	Anno	MAF	bfGWAS PP	Effect-size	P-value
29.1	*ABCA7*	rs3752237	19:1,047,161	coding-syn	0.644	0.544	-0.065	6.7×10^{-3}
29.2	*ABCA7*	rs12151021	19:1,050,874	intronic	0.708	1.000	0.091	1.9×10^{-5}
30	*APOE/TOMM40/ CTB-129P6.7*	rs429358	19:45,411,941	missense	0.118	1.000	-0.173	3.3×10^{-46}
31	*MMP9/RP11-465L10.10*	rs2274755	20:44,639,692	splice	0.138	0.435	-0.073	5.4×10^{-11}
32	*C20orf85*	rs201459901	20:56,653,724	intergenic	0.063	0.078	-0.135	7.9×10^{-18}
33	*SYN3*	rs5754227	22:33,105,817	intronic	0.124	0.764	-0.128	2.0×10^{-27}
34.1	*SLC16A8/BAIAP2L2*	rs4289289	22:38,477,342	missense	0.485	0.824	0.056	1.1×10^{-09}
34.2	*SLC16A8/BAIAP2L2*	rs77968014	22:38,478,666	splice	0.009	0.973	0.212	3.1×10^{-6}

Variants with Bayesian PPs >0.5 or the highest bfGWAS PPs in the loci are listed. Shown are reside/nearby genes, dbSNPIDs, positions, functional annotations, MAFs (unfolded, corresponding to the direction of effect-sizes), P-values, and Bayesian PPs/effect-sizes.

Table S4: AMD risk variants identified by fGWAS in the 34 known loci, accounting for gene-based annotations.

Signal number	Reside/Nearby Gene	dbSNPID	Chr:Position	Anno	MAF	fGWAS PP	P-value
1	*CFH*	rs10922109	1:196,704,632	intronic	0.329	0.802	$< 9.0\times10^{-321}$
2	*COL4A3*	rs11884770	2:228,086,920	intronic	0.731	0.181	5.7×10^{-9}
3	*ADAMTS9-AS2*	rs62247658	3:64,715,155	intronic	0.551	0.167	1.5×10^{-15}
4	*COL8A1*	rs140647181	3:99,180,668	intergenic	0.019	0.999	5.4×10^{-13}
5	*CFI*	rs10033900	4:110,659,067	downstream	0.506	0.996	7.2×10^{-19}
6	*C9*	rs34882957	5:39,331,894	missense	0.012	0.900	4.0×10^{-16}
7	*PRLR/SPEF2*	rs114092250	5:35,494,448	intergenic	0.019	0.626	2.5×10^{-9}
8	*NELFE/SKIV2L*	rs116503776	6:31,930,462	intronic	0.120	0.912	2.1×10^{-114}
9	*VEGFA*	rs943080	6:43,826,627	intergenic	0.518	0.437	2.0×10^{-16}
10	*KMT2E/SRPK2*	rs1142	7:104,756,326	downstream	0.357	0.182	1.5×10^{-10}
11	*PILRB*	rs72615157	7:99,956,444	missense	0.139	0.118	4.0×10^{-10}
12	*TNFRSF10A*	rs79037040	8:23,082,971	nc-transcript	0.534	0.996	2.9×10^{-12}
13	*MIR6130/RORB*	rs10781180	9:76,615,662	intergenic	0.683	0.068	3.0×10^{-10}
14	*TRPM3*	rs71507014	9:73,438,605	intronic	0.584	0.860	3.2×10^{-9}
15	*TGFBR1*	rs10819635	9:101,864,510	upstream	0.186	0.188	2.4×10^{-11}
16	*ABCA1*	rs2740488	9:107,661,742	intronic	0.266	0.760	1.7×10^{-9}
17	*ARHGAP21*	rs12357257	10:24,999,593	intronic	0.232	0.280	4.3×10^{-9}
18	*ARMS2*	rs10490924	10:124,214,448	missense	0.316	0.626	$< 9.0\times10^{-321}$
19	*RDH5/CD63*	rs3138142	12:56,115,585	coding-syn	0.213	0.847	6.1×10^{-10}
20	*MAPKAPK5*	rs61941287	12:112,330,305	intronic	0.019	0.503	1.2×10^{-10}
21	*B3GALTL*	rs9564692	13:31,821,240	splice	0.288	0.889	3.2×10^{-11}
22	*RAD51B*	rs1956526	14:68,799,787	intronic	0.650	0.039	1.0×10^{-11}
23	*ALDH1A2*	rs2414577	15:58,680,638	intronic	0.366	0.495	4.8×10^{-17}
24	*CETP*	rs5817082	16:56,997,349	intronic	0.248	0.193	1.7×10^{-21}
25	*BCAR1*	rs72802395	16:75,286,484	intronic	0.068	0.605	2.1×10^{-11}
26	*POLDIP2/TNFAIP1*	rs13469	17:26,676,135	coding-syn	0.523	0.168	5.1×10^{-9}
27	*NPLOC4/TSPAN10*	rs6420484	17:79,612,397	missense	0.622	0.351	4.0×10^{-12}
28	*C3*	rs2230199	19:6,718,387	missense	0.764	0.999	1.7×10^{-77}
29	*CNN2*	rs10422209	19:1,026,318	upstream	0.132	0.229	5.2×10^{-9}
30	*APOE/TOMM40*	rs429358	19:45,411,941	missense	0.118	1.000	3.3×10^{-46}
31	*MMP9*	rs2274755	20:44,639,692	splice	0.138	0.194	5.4×10^{-11}
32	*C20orf85*	rs117739907	20:56,652,781	intergenic	0.063	0.079	7.8×10^{-18}
33	*SYN3*	rs5754227	22:33,105,817	intronic	0.124	0.781	2.0×10^{-27}
34	*SLC16A8/PICK1*	rs8135665	22:38,476,276	intronic	0.205	0.596	2.9×10^{-12}

Variants with fGWAS PPs >0.5 or the highest fGWAS PPs in the loci are listed in this table. Shown are reside/nearby genes, dbSNPIDs, positions, functional annotations, MAFs (unfolded), fGWAS PPs, and P-values.

Table S5: Candidate AMD loci identified by bfGWAS, accounting for gene-based annotations.

Locus	Reside gene	*dbSNPID*	Chr:Position	Anno	MAF	P-value	Regional-PP	bfGWAS PP	Effect-size
1	*PPIL3*	*rs7562391*	2:201,736,166	missense	0.127	4.8×10^{-7}	0.989	0.666	-0.061
2	*ZNRD1ASP*	*rs114318558*	6:29,966,787	downstream	0.175	2.3×10^{-7}	0.993	0.135	0.058
3	*CPN1*	*rs61751507*	10:101,829,514	missense	0.043	6.7×10^{-8}	0.994	0.598	-0.106
4	*ABHD2*	*rs6496562*	15:89,736,558	splice	0.417	8.4×10^{-8}	0.974	0.517	0.042
5	*LBP*	*rs2232613*	20:36,997,655	missense	0.073	4.3×10^{-7}	0.955	0.881	-0.079

Variants with the highest bfGWAS single variant PP in the candidate loci are listed in this table. Shown are reside genes, dbSNPIDs, positions, functional annotations, MAFs, P-values, Bayesian regional-PPs, and Bayesian PPs/effect-sizes.

Table S6: Candidate AMD loci identified by fGWAS, accounting for gene-based annotations.

Locus	Reside gene	dbSNPID	Chr:Position	Anno	MAF	P-value	Regional-PP	fGWAS PP	Effect-size
1	*PPIL3*	*rs7562391*	2:201,736,166	missense	0.127	4.8×10^{-7}	0.986	0.475	-0.061
2	*HLA-K*	*rs116803720*	6:29,889,989	upstream	0.691	9.3×10^{-10}	0.998	0.101	0.056
3	*CPN1*	*rs61733667*	10:101,802,262	coding-syn	0.036	1.0×10^{-7}	0.994	0.254	-0.118
4	*ABHD2*	*rs6496562*	15:89,736,558	splice	0.417	8.4×10^{-8}	0.978	0.405	0.042
5	*LBP*	*rs2232613*	20:36,997,655	missense	0.073	4.3×10^{-7}	0.973	0.796	-0.079

Variants with the highest fGWAS single variant PP in the candidate loci are listed in this table. Shown are reside genes, dbSNPIDs, positions, functional annotations, MAFs, P-values, fGWAS regional-PPs, fGWAS PPs, and Bayesian effect-sizes

Table S7: AMD risk variants by bfGWAS in the 34 known loci, accounting for summarized regulatory annotations.

Signal number	Reside/nearby gene	dbSNPID	Chr:Position	Anno	MAF	bfGWAS PP	Effect-size	P-value
1.1	*KCNT2*	rs144520124	1:196,371,908	DHS	0.005	1.000	-0.383	1.9×10^{-23}
1.2	*CFH*	rs74979069	1:196,588,463	intergenic	0.049	1.000	0.181	8.1×10^{-92}
1.3	*CFH*	rs1089033	1:196,666,793	intronic	0.412	1.000	-0.117	$< 9.0\times10^{-321}$
1.4	*CFH*	rs2133143	1:196,718,099	intergenic	0.165	0.736	-0.358	5.7×10^{-246}
1.5	*CFH*	esv2672010	1:196,733,401	others	0.157	1.000	-0.283	3.3×10^{-314}
1.6	*CFHR3*	rs188826801	1:196,762,123	intronic	0.014	0.993	0.176	1.2×10^{-39}
1.7	*CFH*	rs79251424	1:196,782,416	intergenic	0.030	0.998	0.144	2.1×10^{-6}
1.8	*RP4-608O15.3*	rs146093852	1:196,811,860	intergenic	0.277	0.994	-0.143	5.7×10^{-254}
2	*COL4A3*	rs11884770	2:228,086,920	intronic	0.731	0.213	0.050	5.6×10^{-9}
3	*ADAMTS9-AS2*	rs11914351	3:64,723,441	intronic	0.240	0.950	-0.064	8.7×10^{-7}
4	*COL8A1*	rs140647181	3:99,180,668	intergenic	0.019	0.575	0.221	5.4×10^{-13}
5	*CFI*	rs10033900	4:110,659,067	intergenic	0.506	0.994	-0.067	7.2×10^{-19}
6	*C9*	rs34882957	5:39,331,894	coding	0.012	0.982	0.278	4.0×10^{-9}
7	*PRLR/SPEF2*	rs114092250	5:35,494,448	intergenic	0.019	0.346	-0.172	2.5×10^{-9}
8.1	*C2/CFB*	rs4151667	6:31,914,024	coding	0.035	0.579	-0.284	1.3×10^{-44}
8.2	*SKIV2/NELFE*	rs115270436	6:31,928,306	coding	0.071	0.566	-0.321	2.8×10^{-99}
9	*VEGFA*	rs943080	6:43,826,627	DHS	0.518	0.678	0.063	2.0×10^{-16}
10	*LINC01004/KMT2E-AS1*	rs6950894	7:104,652,671	promoter	0.511	0.063	-0.047	9.8×10^{-10}
11	*PILRB*	rs7783159	7:100,017,454	coding	0.203	0.115	0.059	5.1×10^{-10}
12	*TNFRSF10A*	rs79037040	8:23,082,971	DHS	0.534	0.995	0.053	2.9×10^{-12}
13	*MIR6130/RORB*	rs10781180	9:76,615,662	intergenic	0.684	0.070	-0.052	3.0×10^{-10}
14	*TRPM3*	rs71507014	9:73,438,605	intronic	0.584	0.763	-0.046	3.2×10^{-9}
15	*TGFBR1*	rs401186	9:101,925,077	promoter	0.200	0.109	-0.063	2.5×10^{-11}
16	*ABCA1*	rs2740488	9:107,661,742	intronic	0.266	0.727	-0.053	1.7×10^{-9}
17	*ARHGAP21*	rs12357257	10:24,999,593	intronic	0.232	0.297	0.053	4.3×10^{-9}
18.1	*ARMS2*	rs7068411	10:124,202,878	intergenic	0.621	1.000	0.252	2.4×10^{-212}
18.2	*ARMS2*	rs7898343	10:124,212,887	promoter	0.083	0.868	-0.311	2.0×10^{-51}
18.3	*ARMS2*	rs10490923	10:124,214,251	coding	0.109	0.962	-0.272	1.7×10^{-53}
18.4	*ARMS2*	rs2736911	10:124,214,355	coding	0.137	0.781	-0.350	1.8×10^{-53}
18.5	*HTRA1*	rs2672601	10:124,220,023	promoter	0.136	0.524	-0.321	4.8×10^{-53}
18.6	*HTRA1*	rs74895474	10:124,230,397	intronic	0.094	1.000	-0.199	1.3×10^{-42}
18.7	*HTRA1*	rs12252027	10:124,234,988	intronic	0.099	1.000	-0.189	1.4×10^{-51}
18.8	*HTRA1*	rs2672589	10:124,234988	DHS	0.653	1.000	0.220	8.9×10^{-180}
19	*RDH5/CD63*	rs143673140	12:56,514,414	coding	0.009	0.001	-0.096	1.3×10^{-2}
20	*MAPKAPK5*	rs61941287	12:112,330,305	intronic	0.019	0.318	0.199	1.2×10^{-10}
21	*B3GALTL*	rs9564692	13:31,821,240	DHS	0.288	0.429	-0.056	3.2×10^{-11}
22	*RAD51B*	rs2842344	14:68,976,971	DHS	0.899	0.215	-0.082	3.7×10^{-7}
23	*ALDH1A2*	rs2414577	15:58,680,638	DHS	0.366	0.508	-0.067	1.5×10^{-9}
24	*CETP*	rs5883	16:57,007,353	promoter	0.060	0.415	0.085	1.4×10^{-20}

Signal number	Reside/nearby gene	dbSNPID	Chr:Position	Anno	MAF	bfGWAS PP	Effect-size	P-value
25	*CTRB2*	rs55993634	16:75,236,763	promoter	0.082	0.321	-0.104	4.6×10^{-5}
26	*POLDIP2/TNFAIP1*	rs13469	17:26,676,135	coding	0.524	0.280	0.044	5.2×10^{-9}
27	*NPLOC4/TSPAN10*	rs9894429	17:79,596,811	coding	0.441	0.261	-0.045	4.0×10^{-12}
28.1	*FUT6/NRTN*	rs17855739	19:5,831,840	coding	0.044	0.549	-0.159	1.5×10^{-16}
28.2	*C3/CTD-3128G10.7*	rs147859257	19:6,718,146	coding	0.008	1.000	0.501	4.3×10^{-31}
28.3	*C3/CTD-3128G10.7*	rs2230199	19:6,718,387	coding	0.764	0.999	-0.173	1.7×10^{-77}
29	*ABCA7*	rs3752241	19:1,053,524	coding	0.160	0.268	0.055	3.2×10^{-7}
30	*APOE(EXOC3L2/MARK4)*	rs429358	19:45,411,941	coding	0.118	1.000	-0.173	3.3×10^{-46}
31	*MMP9/RP11-465L10.10*	rs17577	20:44,643,111	coding	0.138	0.377	-0.072	6.8×10^{-11}
32	*RP13-379L11.1*	rs7266392	20:56,651,542	DHS	0.063	0.115	-0.134	9.2×10^{-18}
33	*SYN3*	rs5754227	22:33,105,817	intronic	0.124	0.524	-0.129	2.0×10^{-27}
34	*SLC16A8/BAIAP2L2*	rs77968014	22:38,478,666	coding	0.009	0.842	0.207	3.1×10^{-6}

Variants with Bayesian PPs >0.5 or the highest bfGWAS PPs in the loci are listed (horizontal lines separate loci). Shown are reside/nearby genes, dbSNPIDs, positions, functional annotations, MAFs (unfolded, corresponding to the direction of effect-sizes), Bayesian PPs/effect-sizes, and P-values.

Table S8: AMD risk variants by fGWAS in the 34 known loci, accounting for summarized regulatory annotations.

Signal number	Reside/nearby gene	dbSNPID	Chr:Position	Anno	MAF	fGWAS PP	P-value
1	*CFH*	rs1089033	1:196,666,793	Intronic	0.412	0.522	$< 9.0\times10^{-321}$
2	*COL4A3*	rs112103000	2:228,072,336	intronic	0.163	0.135	2.0×10^{-8}
3	*ADAMTS9-AS2*	rs6793431	3:64,729,510	intronic	0.891	0.001	6.4×10^{-7}
4	*Intergenic*	rs115407994	3:99,268,860	intergenic	0.018	0.367	9.4×10^{-13}
5	*CFI*	rs10033900	4:110,659,067	intergenic	0.506	0.996	7.2×10^{-19}
6	*C9*	rs34882957	5:39,331,894	coding	0.012	0.757	4.0×10^{-16}
7	*Intergenic*	rs114092250	5:35,494,448	intergenic	0.019	0.617	2.5×10^{-9}
8	*NELFE/SKIV2L*	rs116503776	6:31,930,462	intronic	0.120	0.789	2.1×10^{-114}
9	*Intergenic*	rs943080	6:43,826,627	DHS	0.518	0.557	2.0×10^{-16}
10	*KMT2E/SRPK2*	rs1142	7:104,756,326	UTR	0.357	0.215	1.5×10^{-10}
11	*ZCWPW1*	rs7783159	7:100,017,454	coding	0.203	0.047	5.1×10^{-10}
12	*TNFRSF10A*	rs79037040	8:23,082,971	DHS	0.534	0.995	2.9×10^{-12}
13	*Intergenic*	rs10781180	9:76,615,662	intergenic	0.683	0.067	3.0×10^{-10}
14	*TRPM3*	rs71507014	9:73,438,605	intronic	0.584	0.837	3.2×10^{-9}
15	*TGFBR1*	rs10760667	9:101,864,607	DHS	0.105	0.186	2.5×10^{-11}
16	*ABCA1*	rs2740488	9:107,661,742	intronic	0.266	0.667	1.7×10^{-9}
17	*ARHGAP21*	rs142336524	10:24,879,784	intronic	0.215	0.255	3.2×10^{-8}
18	*ATE1-AS1*	rs11594070	10:123,702,736	nc-transcript	0.334	0.003	1.7×10^{-1}
19	*RDH5/CD63*	rs3138136	12:56,117,570	intronic	0.098	0.001	3.9×10^{-4}
20	*MAPKAPK5*	rs61941287	12:112,330,305	nc-transcript	0.019	0.153	1.2×10^{-10}
21	*B3GALTL*	rs9564692	13:31,821,240	DHS	0.288	0.543	3.2×10^{-11}
22	*RAD51B*	rs11158728	14:68,762,205	DHS	0.641	0.040	1.2×10^{-11}
23	*ALDH1A2*	rs2414577	15:58,680,638	DHS	0.366	0.500	4.8×10^{-17}
24	*CETP*	rs7499892	16:57,006,590	intronic	0.169	0.182	5.3×10^{-21}
25	*BCAR1*	rs72802395	16:75,286,484	intronic	0.068	0.623	2.1×10^{-11}
26	*POLDIP2/NFAIP1*	rs13469	17:26,676,135	coding	0.523	0.134	5.1×10^{-12}
27	*NPLOC4*	rs8070929	17:79,530,993	intronic	0.378	0.176	1.1×10^{-12}
28	*C3*	rs2230199	19:6,718,387	coding	0.764	0.999	1.7×10^{-77}
29	*CNN2/ABCA7*	rs58369307	19:1,038,290	UTR	0.109	0.207	8.5×10^{-9}
30	*APOE/TOMM40*	rs429358	19:45,411,941	coding	0.118	1.000	3.3×10^{-46}
31	*MMP9*	rs17577	20:44,643,111	coding	0.138	0.131	6.8×10^{-11}
32	*RP13-379L11.1*	rs141945849	20:56,650,604	DHS	0.063	0.092	9.3×10^{-18}
33	*SYN3*	rs5754227	22:33,105,817	intronic	0.124	0.681	2.0×10^{-27}
34	*SLC16A8/PICK1*	rs8135665	22:38,476,276	intronic	0.205	0.607	2.9×10^{-12}

Variants with fGWAS PPs >0.5 or the highest fGWAS PPs in the loci or are listed (horizontal lines separate loci). Shown are reside/nearby genes, dbSNPIDs, positions, annotations, MAFs (unfolded, corresponding to the direction of effect-sizes), fGWAS PPs, and P-values.

Table S9: Candidate AMD loci identified by bfGWAS, accounting for summarized regulatory annotations.

Locus	Reside gene	*dbSNPID*	Chr:Position	Anno	MAF	P-value	Regional-PP	bfGWAS PP	Effect-size
1	*PPIL3*	*rs7562391*	2:201,736,166	coding	0.127	$\mathbf{4.8\times10^{-7}}$	0.967	0.475	-0.061
2	*ZNRD1-AS1*	*rs114357644*	6:29,924,728	intergenic	0.669	$\mathbf{2.3\times10^{-7}}$	0.999	0.609	0.051
3	*CPN1*	*rs61733667*	10:101,829,514	coding	0.036	$\mathbf{1.0\times10^{-7}}$	0.994	0.463	-0.118

Variants with the highest bfGWAS PP in the candidate loci are listed in this table. Shown are reside genes, dbSNPIDs, positions, functional annotations, MAFs, P-values, Bayesian regional-PPs, and Bayesian PPs/effect-sizes.

Table S10: Candidate AMD loci identified by fGWAS, accounting for summarized regulatory annotations.

Locus	Reside gene	dbSNPID	Chr:Position	Anno	MAF	P-value	Regional-PP	fGWAS PP	Effect-size
1	*PPIL3*	*rs7562391*	2:201,736,166	coding	0.127	$\mathbf{4.8\times10^{-7}}$	0.976	0.322	-0.061
2	*Intergenic*	*rs115754868*	6:29,884,646	intergenic	0.653	$\mathbf{9.6\times10^{-10}}$	0.998	0.101	0.053
3	*CPN1*	*rs61733667*	10:101,802,262	coding	0.036	$\mathbf{1.0\times10^{-7}}$	0.994	0.253	-0.118
4	*ABHD2*	*rs8042649*	15:89,740,469	UTR	0.417	$\mathbf{1.2\times10^{-7}}$	0.973	0.093	0.049

Variants with the highest fGWAS PP in the candidate loci are listed in this table. Shown are reside genes, dbSNPIDs, positions, functional annotations, MAFs, P-values, fGWAS regional-PPs, fGWAS PPs, and Bayesian effect-sizes.

Table S11: AMD risk variants by bfGWAS in the 34 known loci, accounting for chromatin states profiled with the epigenome of fetal thymus.

Signal number	Reside/nearby gene	dbSNPID	Chr:Position	Anno	MAF	bfGWAS PP	Effect-size	P-value
1.1	*KCNT2*	*rs144520124*	1:196,371,908	Quies	0.005	1.000	-0.389	$\mathbf{1.9\times10^{-23}}$
1.2	*KCNT2*	*rs10754198*	1:196,573,505	Quies	0.258	1.000	-0.078	$\mathbf{1.4\times10^{-228}}$
1.3	*Intergenic*	*rs74979069*	1:196,588,463	Quies	0.049	1.000	0.160	$\mathbf{8.1\times10^{-92}}$
1.4	*CFH*	*rs72734340*	1:196,681,376	Quies	0.037	1.000	-0.189	$\mathbf{1.1\times10^{-1}}$
1.5	*Intergenic*	*rs200467660*	1:196,721,770	Quies	0.161	1.000	-0.405	$\mathbf{1.1\times10^{-249}}$
1.6	*Intergenic*	*rs113632891*	1:196,731,186	Quies	0.155	1.000	-0.173	$\mathbf{2.8\times10^{-296}}$
1.7	*ZNF675*	*rs146093952*	1:196,811,860	Quies	0.277	1.000	-0.207	$\mathbf{2.2\times10^{-310}}$
1.8	*CFHR4*	*rs76258418*	1:196,815,863	Quies	0.130	1.000	-0.199	$\mathbf{2.7\times10^{-293}}$
2	*COL4A3*	*rs112103000*	2:228,072,336	Quies	0.064	0.072	0.064	$\mathbf{2.0\times10^{-8}}$
3.1	*ADAMTS9-AS2*	*rs57305229*	3:64,720,574	Quies	0.304	0.572	-0.057	$\mathbf{2.3\times10^{-5}}$
3.2	*ADAMTS9-AS2*	*rs11914351*	3:64,723,441	Quies	0.240	0.968	-0.064	$\mathbf{8.7\times10^{-7}}$
4	*Intergenic*	*rs140647181*	3:99,180,668	Quies	0.019	0.703	0.222	$\mathbf{5.3\times10^{-13}}$
5	*CFI*	*rs10033900*	4:110,659,067	Quies	0.506	0.999	-0.067	$\mathbf{7.2\times10^{-19}}$
6	*C9*	*rs62358361*	5:39,327,888	Quies	0.012	0.551	0.271	$\mathbf{3.1\times10^{-16}}$
7	*Intergenic*	*rs114092250*	5:35,494,448	Quies	0.019	0.213	-0.171	$\mathbf{2.5\times10^{-9}}$
8.1	*SKIV2L*	*rs116503776*	6:31,930,462	Tx	0.120	1.000	-0.307	$\mathbf{2.1\times10^{-114}}$
8.2	*STK19/C4A*	*rs144629244*	6:31,946,792	Enh	0.014	0.536	0.435	$\mathbf{4.4\times10^{-7}}$
8.3	*PBX2/AGER/GPSM3*	*rs114254831*	6:32,155,581	EnhG	0.271	0.693	0.080	$\mathbf{8.1\times10^{-13}}$
9	*Intergenic*	*rs943080*	6:43,826,627	Quies	0.518	0.422	0.063	$\mathbf{2.0\times10^{-16}}$
10	*KMT2E/SRPK2*	*rs1142*	7:104,756,326	Tx	0.357	0.197	0.051	$\mathbf{1.5\times10^{-10}}$
11	*NYAP1*	*rs67040465*	7:100,083,078	ReprPCWk	0.200	0.040	0.059	$\mathbf{5.7\times10^{-10}}$
12	*TNFRSF10A*	*rs79037040*	8:23,082,971	BivFlnk	0.534	0.967	0.053	$\mathbf{2.9\times10^{-12}}$
13	*Intergenic*	*rs10781180*	9:76,615,662	Quies	0.684	0.090	-0.052	$\mathbf{3.0\times10^{-10}}$
14	*TRPM3*	*rs71507014*	9:73,438,605	Quies	0.585	0.819	-0.046	$\mathbf{3.2\times10^{-9}}$
15	*TGFBR1*	*rs10819635*	9:10,819,635	TxWk	0.186	0.084	-0.066	$\mathbf{2.5\times10^{-11}}$
16	*ABCA1*	*rs2740488*	9:107,661,742	TxWk	0.266	0.759	-0.053	$\mathbf{1.7\times10^{-9}}$
17	*ARHGAP21*	*rs12357257*	10:24,999,593	Quies	0.232	0.308	0.053	$\mathbf{4.3\times10^{-9}}$
18.1	*Intergenic*	*rs7068411*	10:124,202,878	Quies	0.621	1.000	0.198	$\mathbf{2.4\times10^{-212}}$
18.2	*HTRA1*	*rs2672595*	10:124,227,288	ReprePCWk	0.213	0.844	-0.466	$\mathbf{8.7\times10^{-111}}$
18.3	*HTRA1*	*rs74895474*	10:124,230,397	ReprePCWk	0.094	0.578	-0.181	$\mathbf{1.3\times10^{-42}}$
18.4	*HTRA1*	*rs4752699*	10:124,234,320	ReprePCWk	0.128	1.000	-0.292	$\mathbf{2.1\times10^{-51}}$
18.5	*HTRA1*	*rs2672589*	10:124,234,988	ReprePCWk	0.653	1.000	0.274	$\mathbf{8.9\times10^{-180}}$
19	*CDK2/PMEL*	*rs2069389*	12:56,359,642	Enh	0.044	0.001	0.042	$\mathbf{5.3\times10^{-2}}$
20	*CUX2*	*rs142641895*	12:111,786,202	Het	0.019	0.635	0.249	$\mathbf{1.6\times10^{-9}}$
21	*B3GALTL*	*rs9564692*	13:31,821,240	Quies	0.288	0.411	-0.056	$\mathbf{3.2\times10^{-11}}$
22	*RAD51B*	*rs2842339*	14:68,986,999	TxWk	0.899	0.206	-0.082	$\mathbf{3.1\times10^{-7}}$
23	*ALDH1A2*	*rs2414577*	15:58,680,638	Quies	0.366	0.525	-0.067	$\mathbf{4.8\times10^{-17}}$
24	*CETP*	*rs11076175*	16:57,006,378	TxWk	0.67	0.203	-0.072	$\mathbf{5.0\times10^{-21}}$

Signal number	Reside/nearby gene	dbSNPID	Chr:Position	Anno	MAF	bfGWAS PP	Effect-size	P-value
25	*CTRB2*	*rs72802342*	16:75,234,872	Enh	0.074	0.478	-0.114	$\mathbf{2.8\times10^{-13}}$
26	*SARM1/SLC46A1*	*rs4795433*	17:26,716,821	ReprPCWk	0.524	0.138	0.045	$\mathbf{1.6\times10^{-9}}$
27	*NPLOC4*	*rs8070929*	17:79,530,993	Tx	0.378	0.226	0.058	$\mathbf{1.1\times10^{-12}}$
28.1	*FUT6*	*rs12019136*	19:5,835,677	Quies	0.042	0.639	-0.160	$\mathbf{3.7\times10^{-17}}$
28.2	*C3*	*rs147859257*	19:6,718,146	Het	0.008	1.000	0.504	$\mathbf{4.3\times10^{-31}}$
28.3	*C3*	*rs2230199*	19:6,718,387	Het	0.764	0.996	-0.172	$\mathbf{1.7\times10^{-77}}$
29	*CNN2/ABCA7*	*rs3087680*	19:1,038,289	TxFlnk	0.109	0.208	0.072	$\mathbf{8.6\times10^{-9}}$
30	*APOE/TOMM40*	*rs429358*	19:45,411,941	ReprPCWk	0.118	1.000	-0.186	$\mathbf{3.3\times10^{-46}}$
31	*MMP9*	*rs142450006*	20:44,614,991	ReprPCWk	0.132	0.251	-0.079	$\mathbf{1.4\times10^{-11}}$
32	*Intergenic*	*rs140611615*	20:56,653,111	Quies	0.062	0.080	-0.135	$\mathbf{8.2\times10^{-18}}$
33	*SYN3*	*rs5754227*	22:33,105,817	Quies	0.124	0.896	-0.128	2.0×10^{-27}
34	*SLC16A8/PICK1/ BAIAP2L2*	*rs8135665*	22:38,476,276	ReprPC	0.206	0.624	0.066	2.9×10^{-12}

Variants with Bayesian PPs >0.5 or the highest bfGWAS PPs in the loci are listed in this table. Shown are reside/nearby genes, dbSNPIDs, positions, annotations, MAFs (unfolded, corresponding to the direction of effect-sizes), P-values, and Bayesian PPs/effect-sizes.

Table S12: AMD risk variants by fGWAS in the 34 known loci, accounting for chromatin states profiled with the epigenome of fetal thymus.

Signal number	Reside/Nearby Gene	dbSNPID	Chr:Position	Anno	MAF	fGWAS PP	P-value
1	*CFH*	rs1089033	1:196,666,793	Quies	0.412	1.000	$< 9.0\times10^{-321}$
2	*COL4A3*	rs11884770	2:228,086,920	Quies	0.731	0.731	5.7×10^{-9}
3	*ADAMTS9-AS2*	rs66793786	3:64,707,880	Quies	0.243	0.050	2.0×10^{-7}
4	*COL8A1*	rs140647181	3:99,180,668	Quies	0.019	0.307	5.4×10^{-13}
5	*CFI*	rs10033900	4:110,659,067	Quies	0.506	0.994	7.2×10^{-19}
6	*C9*	rs62358361	5:39,327,888	Quies	0.012	0.559	3.1×10^{-16}
7	*PRLR/SPEF2*	rs114092250	5:35,494,448	Quies	0.019	0.468	2.5×10^{-9}
8	*NELFE/SKIV2L*	rs116503776	6:31,930,462	Tx	0.120	0.967	2.1×10^{-114}
9	*VEGFA*	rs943080	6:43,826,627	Quies	0.518	0.437	2.0×10^{-16}
10	*KMT2E/SRPK2*	rs1142	7:104,756,326	Tx	0.357	0.141	1.5×10^{-10}
11	*ZKSCAN1*	rs2406255	7:100,053,690	EnhG	0.200	0.026	5.9×10^{-10}
12	*TNFRSF10A*	rs79037040	8:23,082,971	BivFlnk	0.534	0.998	2.9×10^{-12}
13	*Intergenic*	rs10781180	9:76,615,662	Quies	0.684	0.068	3.0×10^{-10}
14	*TRPM3*	rs71507014	9:73,438,605	Quies	0.584	0.776	3.2×10^{-9}
15	*TGFBR1*	rs6478972	9:101,869,278	Enh	0.200	0.103	3.5×10^{-11}
16	*ABCA1*	rs2740488	9:107,661,742	TxWk	0.266	0.746	1.7×10^{-9}
17	*ARHGAP21*	rs12357257	10:24,999,593	Quies	0.232	0.269	4.3×10^{-9}
18	*ARMS2*	rs2672599	10:124,211,875	Quies	0.641	1.000	2.7×10^{-263}
19	*RDH5/CD63*	rs3138136	12:56,117,570	EnhG	0.099	0.001	3.9×10^{-4}
20	*MAPKAPK5*	rs61941287	12:112,330,305	Tx	0.019	0.205	1.2×10^{-10}
21	*B3GALTL*	rs9564692	13:31,821,240	Quies	0.288	0.388	3.2×10^{-11}
22	*RAD51B*	rs11158728	14:68,762,205	Enh	0.640	0.066	1.0×10^{-11}
23	*ALDH1A2*	rs2414577	15:58,680,638	Quies	0.366	0.495	4.8×10^{-17}
24	*CETP*	rs5817082	16:56,997,349	TxWk	0.248	0.254	1.7×10^{-21}
25	*CTRB2*	rs72802342	16:75,234,872	Enh	0.073	0.656	2.8×10^{-13}
26	*TNFAIP1/POLDIP2*	rs733914	17:26,671,196	EnhG	0.526	0.156	3.5×10^{-9}
27	*NPLOC4*	rs8070929	17:79,530,993	Tx	0.378	0.221	1.1×10^{-12}
28	*C3*	rs2230199	19:6,718,387	Het	0.764	0.992	1.7×10^{-77}
29	*CNN2/ABCA7*	rs58369307	19:1,038,290	TxFlnk	0.109	0.369	8.5×10^{-9}
30	*APOE/TOMM40*	rs429358	19:45,411,941	ReprPCWk	0.118	1.000	3.3×10^{-46}
31	*MMP9*	rs1888235	20:44,623,967	Enh	0.133	0.281	1.4×10^{-11}
32	*C20orf85*	rs117739907	20:56,652,781	Quies	0.062	0.079	7.8×10^{-18}
33	*SYN3*	rs5754227	22:33,105,817	Quies	0.124	0.791	2.0×10^{-27}
34	*SLC16A8/PICK1*	rs8135665	22:38,476,276	ReprPC	0.205	0.773	2.9×10^{-12}

Variants with either the highest fGWAS PP per locus or fGWAS PP > 0.5 are listed (horizontal lines separate loci). Shown are reside/nearby genes, dbSNPIDs, positions, functional annotations, MAFs (unfolded, corresponding to the direction of effect-sizes), fGWAS PPs, and P-values.

Table S13: Candidate AMD loci identified by bfGWAS, accounting for chromatin states profiled with the epigenome of fetal thymus.

Locus	Reside gene	*dbSNPID*	Chr:Position	Anno	MAF	P-value	Regional-PP	bfGWAS PP	Effect-size
1	*HLA-W*	*rs114357644*	6:29,924,728	TxWk	0.669	2.3×10^{-7}	0.988	0.877	0.051
2	*CPN1*	*rs111563092*	10:101,808,993	ReprPCWk	0.045	7.2×10^{-8}	0.998	0.171	-0.106

Variants with the highest bfGWAS PPs in the candidate loci are listed in this table. Shown are reside genes, dbSNPIDs, positions, functional annotations, MAFs, P-values, Bayesian regional-PPs, and Bayesian PPs/effect-sizes.

Table S14: Candidate AMD loci identified by fGWAS, accounting for chromatin states profiled with the epigenome of fetal thymus.

Locus	Reside gene	dbSNPID	Chr:Position	Anno	MAF	P-value	Regional-PP	fGWAS PP	Effect-size
1	*PPIL3*	*rs7562391*	2:201,736,166	Tx	0.127	6.5×10^{-8}	0.969	0.088	-0.061
2	*Intergenic*	*rs140766203*	6:29,883,869	Quies	0.652	8.5×10^{-10}	0.998	0.044	0.053
3	*CPN1*	*rs113582392*	10:101,804,258	Enh	0.045	1.4×10^{-8}	0.993	0.154	-0.106
4	*ABHD2*	*rs4932480*	15:89,723,858	EnhG	0.501	7.2×10^{-8}	0.971	0.138	-0.043

Variants with the highest fGWAS PPs in the candidate loci are listed in this table. Shown are reside genes, dbSNPIDs, positions, functional annotations, MAFs, P-values, fGWAS regional-PPs, fGWAS PPs, and Bayesian effect-sizes.

Table S15: Haplotype analysis in locus C2/CFB/SKIV2L.

Region	Haplotype			Haplotype Frequency (%)		P-value	OR (95% CI)
	SKIV2L intronic (*rs116503776*)	*CFB* missense (*rs4151667*)	*CFB* missense (*rs115270436*)	Cases	Controls		
C2/CFB/SKIV2L	1	1	1	1.5×10^{-3}	4.2×10^{-3}	8.9×10^{-11}	0.364 (0.265, 0.501)
	1	0	1	0.046	0.085	1.5×10^{-86}	0.522 (0.490, 0.557)
	1	1	0	0.023	0.041	5.0×10^{-36}	0.561 (0.513, 0.613)
	0	0	1	8.9×10^{-4}	1.5×10^{-3}	0.024	0.586 (0.375, 0.917)
	1	0	0	0.018	0.017	0.092	1.102 (0.983, 1.236)
	0	0	0	0.909	0.850	-	Reference Haplotype
	0	1	0	6.1×10^{-5}	2.8×10^{-5}	0.306	1.840 (0.243, 13.938)

Considered the haplotype consisting with the top significant intronic variant found by single variant test P-values (*rs116503776* with p-value=2.1×10^{-114}), the top two significant missense variants (in the $\pm$20KB region around *rs116503776*) found by bfGWAS (*rs4151667* with Bayesian PP=0.903, *rs115270436* with Bayesian PP= 0.638).

Table S16: Model comparison.

Region (*C2/CFB/SKIV2L)*	*SKIV2L* intronic (*rs116503776*) & *PBX2* intronic (*rs114254831*)	*CFB* missense (*rs4151667*) & *SKIV2L* missense (*rs115270436*)	Differences (col2-col3)
Akaike information criterion (AIC)	95857.36	95752.63	104.73
Bayesian information criterion (BIC)	95891.1	95786.36	104.74
Log Likelihood	-47924.68	-47872.31	-52.37

Compared the linear regression model with the top two independent significant variants (*rs116503776*, *rs114254831*) found by conditional analysis, versus the linear regression model with the top two significant variants (*rs4151667*, *rs115270436*) found by bfGWAS accounting for gene-based annotations.

Supplemental References

[1] Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447(7145):661–678, 06 2007.

[2] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):pp. 1–38, 1977.

[3] Edgar C Fieller. Some problems in interval estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 175–185, 1954.

[4] Andrew Gelman and Donald B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):pp. 457–472, 1992.

[5] Edward I. George and Robert E. McCulloch. Variable selection via gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889, 1993.

[6] Yongtao Guan and Matthew Stephens. Bayesian variable selection regression for genome-wide association studies and other large-scale problems. *Ann. Appl. Stat.*, 5(3):1780–1815, 09 2011.

[7] Xiang Zhou, Peter Carbonetto, and Matthew Stephens. Polygenic modeling with bayesian sparse linear mixed models. *PLoS Genet*, 9(2):e1003264, 02 2013.