Supplementary information for "Modeling regulatory network topology improves genome-wide analyses of complex human traits' '

Xiang Zhu

The Pennsylvania State University, University Park, PA 16802, USA

E-mail: xiangzhu@psu.edu

Summary. Below are Supplementary Notes 1-8, Figures 1-19 and Tables 1-17.

Supplementary Notes

Supplementary Note 1

RSS-NET is a Bayesian hierarchical model given by

$$\begin{aligned}
\widehat{\boldsymbol{\beta}} &\sim \mathcal{N}(\widehat{\mathbf{S}}\widehat{\mathbf{R}}\widehat{\mathbf{S}}^{-1}\boldsymbol{\beta},\ \widehat{\mathbf{S}}\widehat{\mathbf{R}}\widehat{\mathbf{S}}), && (1)\\
\beta_j &\sim \pi_j\cdot\mathcal{N}(\mu_j,\ \sigma_0^2)+(1-\pi_j)\cdot\delta_0, && (2)\\
\pi_j &= 1/[1+10^{-(\theta_0+a_j\cdot\theta)}], && (3)\\
\mu_j &= \textstyle\sum_{g\in\mathbf{O}_j} w_{jg}\cdot\gamma_{jg}, && (4)\\
\gamma_{jg} &\sim \mathcal{N}(0,\ \sigma^2), && (5)
\end{aligned}$$

where $\mathcal{N}(\boldsymbol{\mu},\ \boldsymbol{\Sigma})$ denotes a normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, $\widehat{\boldsymbol{\beta}} := (\hat{\beta}_1,\ldots,\hat{\beta}_p)'$ is a $p\times 1$ vector, $\widehat{\mathbf{S}} := \mathrm{diag}\{(\hat{s}_1,\ldots,\hat{s}_p)'\}$ is a $p\times p$ diagonal matrix, $\{\hat{\beta}_j,\hat{s}_j\}$ are single-SNP effect estimate and its standard error for each SNP j in a GWAS, $\widehat{\mathbf{R}} := [\hat{r}_{ij}]$ is a $p\times p$ LD matrix estimated from a reference panel with ancestry matching the GWAS, $\boldsymbol{\beta} := (\beta_1,\ldots,\beta_p)'$ is a $p\times 1$ vector, β_j is the true effect of each SNP j, π_j denotes the prior probability that $\beta_j\neq 0$, μ_j denotes the prior expected value of β_j conditioning on $\beta_j\neq 0$, δ_0 denotes point mass at zero, $\{a_j,\mathbf{O}_j,w_{jg}\}$ are known annotations derived from a given regulatory network (see main text Equations (3)-(5)), γ_{jg} denotes the random effect of SNP j due to gene g, and $\{\theta_0,\theta,\sigma_0,\sigma\}$ are hyper-parameters.

To fit the RSS-NET model, we first integrate out γ_{jg}:

$$\begin{aligned}
\widehat{\boldsymbol{\beta}} &\sim \mathcal{N}(\widehat{\mathbf{S}}\widehat{\mathbf{R}}\widehat{\mathbf{S}}^{-1}\boldsymbol{\beta},\ \widehat{\mathbf{S}}\widehat{\mathbf{R}}\widehat{\mathbf{S}}), && (6)\\
\beta_j &\sim \pi_j\cdot\mathcal{N}(0,\ \sigma_j^2)+(1-\pi_j)\cdot\delta_0, && (7)\\
\pi_j &= 1/[1+10^{-(\theta_0+a_j\theta)}], && (8)\\
\sigma_j^2 &= \sigma_0^2+\sigma^2\cdot\textstyle\sum_{g\in\mathbf{O}_j} w_{jg}^2. && (9)
\end{aligned}$$

We then write the posterior distribution of $\boldsymbol{\beta}$ as

$$f(\boldsymbol{\beta}\mid\mathbf{D}) = \int f(\boldsymbol{\beta}\mid\mathbf{D},\theta_0,\theta,\sigma_0,\sigma)f(\theta_0,\theta,\sigma_0,\sigma\mid\mathbf{D})\mathrm{d}\theta_0\mathrm{d}\theta\mathrm{d}\sigma_0\mathrm{d}\sigma, \quad (10)$$

where $\mathbf{D}$ is a shorthand for the input data of RSS-NET including GWAS summary statistics $\{\widehat{\boldsymbol{\beta}},\widehat{\mathbf{S}}\}$, LD estimates $\widehat{\mathbf{R}}$ and network annotations $\{\mathbf{a},\mathbf{O},\mathbf{W}\}$ (see main text Equations

(3)-(5)). Given $\{\theta_0, \theta, \sigma_0, \sigma\}$, we use the mean-field variational approximation algorithm from RSS-E [1] to estimate $f(\boldsymbol{\beta} \mid \mathbf{D}, \theta_0, \theta, \sigma_0, \sigma)$:

$$f(\boldsymbol{\beta} \mid \mathbf{D}, \theta_0, \theta, \sigma_0, \sigma) \approx \prod_{j=1}^{p} q_j^\star(\beta_j \mid \mathbf{D}, \theta_0, \theta, \sigma_0, \sigma), \tag{11}$$

$$q_j^\star(\beta_j \mid \mathbf{D}, \theta_0, \theta, \sigma_0, \sigma) = \alpha_j^\star \cdot \mathcal{N}(\beta_j;\, \nu_j^\star, (\tau_j^\star)^2) + (1-\alpha_j^\star) \cdot \delta_0(\beta_j), \tag{12}$$

where the optimal variational parameters $\{\alpha_j^\star, \nu_j^\star, \tau_j^\star\}$ are given by:

$$(\tau_j^\star)^2 = \frac{1}{\hat{s}_j^{-2} + \sigma_j^{-2}}, \tag{13}$$

$$\nu_j^\star = (\tau_j^\star)^2 \cdot \left(\frac{\hat{\beta}_j}{\hat{s}_j^2} - \sum_{i \neq j} \frac{\hat{r}_{ij} \alpha_i^\star \nu_i^\star}{\hat{s}_i \hat{s}_j} \right), \tag{14}$$

$$\frac{\alpha_j^\star}{1-\alpha_j^\star} = \frac{\pi_j}{1-\pi_j} \cdot \frac{\tau_j^\star}{\sigma_j} \cdot \exp\left\{ \frac{(\nu_j^\star)^2}{2(\tau_j^\star)^2} \right\}. \tag{15}$$

The per-iteration complexity of the coordinate descent algorithm specified by Equations (13)-(15) is linear with p, the number of genome-wide SNPs analyzed.

We use $F^\star(\mathbf{D}, \theta_0, \theta, \sigma_0, \sigma)$, the variational lower bound corresponding to the optimal variational parameters $\{\alpha_j^\star, \nu_j^\star, \tau_j^\star\}$ in Equations (13)-(15) to approximate the log marginal likelihood $\log f(\widehat{\boldsymbol{\beta}} \mid \widehat{\mathbf{S}}, \widehat{\mathbf{R}}, \mathbf{a}, \mathbf{O}, \mathbf{W}, \theta_0, \theta, \sigma_0, \sigma)$:

$$\begin{aligned} F^\star(\mathbf{D}, \theta_0, \theta, \sigma_0, \sigma) = {} & F_0(\mathbf{D}) + \widehat{\boldsymbol{\beta}}' \widehat{\mathbf{S}}^{-2} \mathrm{E}_{q^\star}(\boldsymbol{\beta}) - \frac{1}{2} \mathrm{E}'_{q^\star}(\boldsymbol{\beta}) \widehat{\mathbf{S}}^{-1} \widehat{\mathbf{R}} \widehat{\mathbf{S}}^{-1} \mathrm{E}_{q^\star}(\boldsymbol{\beta}) - \frac{1}{2} \sum_{j=1}^{p} \frac{\mathrm{Var}_{q^\star}(\beta_j)}{\hat{s}_j^2} \\ & - \sum_{j=1}^{p} \alpha_j^\star \log\left(\frac{\alpha_j^\star}{\pi_j} \right) - \sum_{j=1}^{p} (1-\alpha_j^\star) \log\left(\frac{1-\alpha_j^\star}{1-\pi_j} \right) \\ & + \sum_{j=1}^{p} \frac{\alpha_j^\star}{2} \left\{ 1 + \log\left[\frac{(\tau_j^\star)^2}{\sigma_j^2} \right] - \frac{(\tau_j^\star)^2 + (\nu_j^\star)^2}{\sigma_j^2} \right\}, \end{aligned} \tag{16}$$

where $F_0(\mathbf{D}) = -[\log|2\pi \cdot \widehat{\mathbf{S}}\widehat{\mathbf{R}}\widehat{\mathbf{S}}| + \widehat{\boldsymbol{\beta}}'(\widehat{\mathbf{S}}\widehat{\mathbf{R}}\widehat{\mathbf{S}})^{-1}\widehat{\boldsymbol{\beta}}]$ / 2 is a constant of $\{\theta_0, \theta, \sigma_0, \sigma\}$, $\mathrm{E}_{q^\star}(\boldsymbol{\beta}) = (\mathrm{E}_{q^\star}(\beta_1), \ldots, \mathrm{E}_{q^\star}(\beta_p))'$, $\mathrm{E}_{q^\star}(\beta_j) = \alpha_j^\star \nu_j^\star$ and $\mathrm{Var}_{q^\star}(\beta_j) = \alpha_j^\star[(\tau_j^\star)^2 + (\nu_j^\star)^2] - (\alpha_j^\star \nu_j^\star)^2$. Of note, RSS-NET and RSS-E have the same computation scheme if we set $\sigma_j^2 = \sigma_\beta^2$ for all $j = 1, \ldots, p$ in Equations (13)-(16).

Finally, we approximate the joint posterior of hyper-parameters $f(\theta_0, \theta, \sigma_0, \sigma \mid \mathbf{D})$ as:

$$f(\theta_0, \theta, \sigma_0, \sigma \mid \mathbf{D}) \approx \omega^\star(\theta_0, \theta, \sigma_0, \sigma) \propto \exp\{F^\star(\mathbf{D}, \theta_0, \theta, \sigma_0, \sigma)\}, \tag{17}$$

since we use independent uniform grid hyper-priors (Supplementary Table 14).

Supplementary Note 2

To assess whether a regulatory network is enriched for genetic associations with a trait, we evaluate the following Bayes factor (BF):

$$\mathrm{BF} = \frac{f(\widehat{\boldsymbol{\beta}} \mid \widehat{\mathbf{S}}, \widehat{\mathbf{R}}, \mathbf{a}, \mathbf{O}, \mathbf{W}, M_1)}{f(\widehat{\boldsymbol{\beta}} \mid \widehat{\mathbf{S}}, \widehat{\mathbf{R}}, \mathbf{a}, \mathbf{O}, \mathbf{W}, M_0)}, \tag{18}$$

where M_1 denotes the enrichment model with $\theta > 0$ or $\sigma^2 > 0$, and M_0 denotes the baseline model with $\theta = 0$ and $\sigma^2 = 0$. To compute BF we approximate intractable marginal likelihoods in Equation (18) by corresponding variational lower bounds $F^\star$:

$$\begin{aligned} \text{BF} &= \frac{\int f(\widehat{\boldsymbol{\beta}} \mid \widehat{\mathbf{S}}, \widehat{\mathbf{R}}, \mathbf{a}, \mathbf{O}, \mathbf{W}, \theta_0, \theta, \sigma_0, \sigma) f(\theta_0) f(\theta) f(\sigma_0) f(\sigma) \mathrm{d}\theta \mathrm{d}\theta_0 \mathrm{d}\sigma_0 \mathrm{d}\sigma}{\int f(\widehat{\boldsymbol{\beta}} \mid \widehat{\mathbf{S}}, \widehat{\mathbf{R}}, \mathbf{a}, \mathbf{O}, \mathbf{W}, \theta_0, \theta = 0, \sigma_0, \sigma = 0) f(\theta_0) f(\sigma_0) \mathrm{d}\theta_0 \mathrm{d}\sigma_0} \\ &\approx \frac{H^{-1} \sum_{h=1}^{H} \exp\{F^\star(\mathbf{D}, \theta_0^{(h)}, \theta^{(h)}, \sigma_0^{(h)}, \sigma^{(h)})\}}{L^{-1} \sum_{\ell=1}^{L} \exp\{F^\star(\mathbf{D}, \theta_0^{(\ell)}, \theta = 0, \sigma_0^{(\ell)}, \sigma = 0)\}}, \end{aligned} \tag{19}$$

where $\{\theta_0^{(h)}, \theta^{(h)}, \sigma_0^{(h)}, \sigma^{(h)}\}$ and $\{\theta_0^{(\ell)}, \sigma_0^{(\ell)}\}$ are pre-defined (Supplementary Table 14).

Supplementary Note 3

To identify association between a locus and a trait, we compute P_1, the posterior probability that at least one SNP in the locus is associated with the trait:

$$P_1 = 1 - \Pr(\beta_j = 0, \forall j \in \text{locus} \mid \mathbf{D}, \text{model}), \tag{20}$$

where the "model" here can be the baseline model M_0 (yielding P_1^{base}), the enrichment model M_1 for the near-gene control network (yielding P_1^{near}) or M_1 for a given regulatory network (yielding P_1^{net}). Given a grid $\{\theta_0^{(h)}, \theta^{(h)}, \sigma_0^{(h)}, \sigma^{(h)}\}$, P_1 is estimated as

$$\begin{aligned} P_1 &= 1 - \int \Pr(\beta_j = 0,\ \forall j \in \text{locus} \mid \mathbf{D}, \theta_0, \theta, \sigma_0, \sigma) f(\theta_0, \theta, \sigma_0, \sigma \mid \mathbf{D}) \mathrm{d}\theta \mathrm{d}\theta_0 \mathrm{d}\sigma_0 \mathrm{d}\sigma \\ &\approx 1 - \sum_{h=1}^{H} \prod_{j \in \text{locus}} \left[1 - \alpha_j^\star(\theta_0^{(h)}, \theta^{(h)}, \sigma_0^{(h)}, \sigma^{(h)})\right] \cdot \omega^\star(\theta_0^{(h)}, \theta^{(h)}, \sigma_0^{(h)}, \sigma^{(h)}). \end{aligned} \tag{21}$$

Supplementary Note 4

In this study a regulatory network is a directed bipartite graph $\{\mathbf{V}_{\text{TF}}, \mathbf{V}_{\text{TG}}, \mathbf{E}_{\text{TF}\to\text{TG}}\}$, where $\mathbf{V}_{\text{TF}}$ denotes the node set of transcription factors (TFs), $\mathbf{V}_{\text{TG}}$ denotes the node set of target genes (TGs), and $\mathbf{E}_{\text{TF}\to\text{TG}}$ denotes the set of directed TF-to-TG edges, summarizing how TFs regulate TGs through regulatory elements (REs). Each edge has a weight between 0 and 1, measuring the relative regulation strength of a TF on a TG.

Below is a small subset of B cell regulatory network (`https://github.com/suwonglab/rss-net/tree/master/data`). In this example, $\mathbf{V}_{\text{TF}}$ corresponds to the column `TF`, $\mathbf{V}_{\text{TG}}$ corresponds to the column `TG`, $\mathbf{E}_{\text{TF}\to\text{TG}}$ corresponds to all rows of `TF` and `TG`, and the edge weights are specified by the column `Score`.

```
       TF     TG    Score
1:   RARG EEF1A1 0.723037
2:   SOX9   CD74 0.690921
3:   EBF1  FXYD5 0.659009
4:   PAX5   CD44 0.704366
5: TRIM28   GSAP 0.629798
6: POU6F1  RPS3A 0.632690
```

Given a TF, here we only consider the downstream effects of all genes that are **directly** regulated by this TF; see main text Figure 1c and Equation (5). For example, it is biologically possible that TF 1 regulates a target gene which is also a TF, say TF 2; when

specifying the RSS-NET SNP-level effect size distribution, we only incorporate the direct downstream effect of TF 1 on TF 2, and ignore all other indirect downstream effect of TF 1 on genes that are only directly regulated by TF 2.

Supplementary Note 5

To specify the relative impact of a cis SNP j on a gene g c_{jg} in main text Equation (5), we use a simple approach based on cis expression quantitative trait loci (eQTL). Specifically, if a pair of SNP j and gene g is available in an eQTL database, we define

$$c_{jg} = 1 + \sqrt{z_{jg}^2 / (z_{jg}^2 + n_{jg})}, \tag{22}$$

where n_{jg} is the number of individuals with both genotype of SNP j and expression of gene g available and z_{jg} is the single-SNP z-score measuring the marginal association between genotype of SNP j and expression of gene g across n_{jg} individuals; if the pair does not have eQTL association data available, we set $c_{jg} = 1$. When deriving c_{jg} from cis-eQTL data, we consider all available SNP-gene pairs, not just the significant ones.

In this study we specify c_{jg} in a context-matched manner. For example, if we analyze the regulatory network derived from liver samples, we also use cis-eQTL data from liver samples to define c_{jg}. For networks of 5 brain regions and 27 non-brain tissues, we use the tissue-matched cis-eQTL data from GTEx (`https://storage.googleapis.com/gtex_analysis_v7/single_tissue_eqtl_data/GTEx_Analysis_v7_eQTL_all_associations.tar.gz`, accessed April 7, 2019). For networks of 5 immune cells, we use the cell-type-matched cis-eQTL data from DICE (`http://downloads.dice-database.org/downloads/DICE_DB_1/eqtl/unfiltered/`, accessed July 3, 2019). For the "omnibus" network, we use the cis-eQTL data of blood samples from the eQTLGen (`http://www.eqtlgen.org/cis-eqtls.html`, accessed May 17, 2019). Summaries of c_{jg} are provided in Supplementary Tables 12-13.

The small sample sizes of eQTL studies make it hard to produce reliable model-based estimates of c_{jg}. Hence, we keep our specification of c_{jg} deliberately simple. We acknowledge that Equation (22) could be potentially improved, although we have not fully investigated it here. Due to the modular design of RSS-NET, once improved estimates of c_{jg} become available, they can be incorporated into RSS-NET in a straightforward manner.

Supplementary Note 6

In this work we model the random effect γ_{jg} of SNP j due to gene g as

$$\gamma_{jg} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\, \sigma^2).$$

We emphasize that the SNP-level subscript j in γ_{jg} ensures the exchangeability of β_j required by Equation (2), as illustrated below. Suppose there are two nearby SNPs 1 and 2, and their effects are sums of random effects of genes 1 and 2. If we ignore the SNP-level subscript j and use the γ_g notation, then

$$\beta_1 = 0.18 \cdot \gamma_1 + 0.36 \cdot \gamma_2,\;\; \beta_2 = 0.36 \cdot \gamma_1 + 0.72 \cdot \gamma_2,\;\; \gamma_g \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\, \sigma^2),\; g = 1, 2.$$

This is inconsistent with the independent prior placed on β_j as shown in Equation (2), because under the γ_g notation, $\beta_2 = 2\beta_1$ and the correlation of β_1 and β_2 is exactly 1. In

contrast, this inconsistency disappears if we use the γ_{jg} notation:

$$\beta_1 = 0.18 \cdot \gamma_{11} + 0.36 \cdot \gamma_{12}, \;\; \beta_2 = 0.36 \cdot \gamma_{21} + 0.72 \cdot \gamma_{22}, \;\; \gamma_{jg} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\, \sigma^2),\; j = 1,2,\; g = 1,2.$$

Now β_1 and β_2 are independent, because γ_{jg}'s are all i.i.d for $j = 1,2$ and $g = 1,2$.

Supplementary Note 7

Here we assume a multiple regression model for the phenotype-genotype relationship:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{23}$$

where $\mathbf{y}$ is an $n \times 1$ centered vector of phenotype, $\mathbf{X}$ is an $n \times p$ column-centered matrix of genotype, $\boldsymbol{\beta}$ is a $p \times 1$ vector of multiple regression coefficients (i.e., SNP-level effect sizes), and $\boldsymbol{\epsilon}$ is an independent error term. Let $\sigma^2_{x,j}$ denote the population variance of genotype for SNP j, $j = 1, \ldots, p$, and let σ^2_y denote the population variance of phenotype.

PROPOSITION 1. If SNP-level effect sizes $\boldsymbol{\beta}$ follow the prior distribution specified in Equations (2), (4) and (5), then, for all $j = 1, \ldots, p$,

$$\mathrm{E}(\beta_j) = 0, \;\; \mathrm{Var}(\beta_j) = \pi_j \left(\sigma_0^2 + \sigma^2 \cdot \textstyle\sum_{g \in \mathbf{O}_j} w_{jg}^2\right), \;\; \mathrm{Cov}(\beta_i,\, \beta_j) = 0, \;\; i \neq j. \tag{24}$$

PROOF. This is a direct application of the law of total expectation. ■

PROPOSITION 2. $\mathrm{E}[V(\mathbf{X}\boldsymbol{\beta})] = \mathrm{E}^{\mathsf{base}}[V(\mathbf{X}\boldsymbol{\beta})] + \mathrm{E}^{\mathsf{net}}[V(\mathbf{X}\boldsymbol{\beta})]$, where

$$\mathrm{E}^{\mathsf{base}}[V(\mathbf{X}\boldsymbol{\beta})] \;=\; \sigma_0^2 \cdot \sum_{j=1}^{p} \pi_j \cdot \mathrm{E}[V(\mathbf{X}_j)], \tag{25}$$

$$\mathrm{E}^{\mathsf{net}}[V(\mathbf{X}\boldsymbol{\beta})] \;=\; \sigma^2 \cdot \sum_{j=1}^{p} \pi_j \cdot \left(\textstyle\sum_{g \in \mathbf{O}_j} w_{jg}^2\right) \cdot \mathrm{E}[V(\mathbf{X}_j)], \tag{26}$$

$\mathbf{X}_j$ is the jth column of $\mathbf{X}$, $j = 1, \ldots, p$, and $V(\mathbf{d})$ denotes the sample variance of a vector $\mathbf{d}$.

PROOF. Define a $p \times p$ diagonal matrix $\boldsymbol{\Sigma}_\beta$ where the jth diagonal entry is $\mathrm{Var}(\beta_j)$ in PROPOSITION 1. It suffices to show that $V(\mathbf{X}\boldsymbol{\beta}) = n^{-1}\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$ and

$$\begin{aligned} \mathrm{E}[V(\mathbf{X}\boldsymbol{\beta})] \;&=\; \mathrm{E}\{\mathrm{E}[V(\mathbf{X}\boldsymbol{\beta}) \mid \mathbf{X}]\} = \mathrm{E}[\mathrm{trace}(n^{-1}\mathbf{X}'\mathbf{X} \cdot \boldsymbol{\Sigma}_\beta)], \\ &=\; \mathrm{E}\left[\sum_{j=1}^{p} n^{-1}\mathbf{X}_j'\mathbf{X}_j \cdot \mathrm{Var}(\beta_j)\right] = \sum_{j=1}^{p} \mathrm{Var}(\beta_j) \cdot \mathrm{E}[V(\mathbf{X}_j)]. \end{aligned} \tag{27}$$

Use PROPOSITION 1 to complete the proof. ■

REMARK 1. PROPOSITION 2 shows that the total genetic variation $\mathrm{E}[V(\mathbf{X}\boldsymbol{\beta})]$ can be decomposed into a component that is shared by all p SNPs (Equation (25)) and a component that is due to network annotations $\{\mathbf{O}_j, w_{jg}\}$ (Equation (26)).

PROPOSITION 3. Let $ns_j^2 = \sigma_y^2 \,/\, \sigma_{x,j}^2$, for $j = 1, \ldots, p$. If $\{\sigma_0^2, \sigma^2\}$ are re-parameterized as

$$\sigma_0^2 = \eta \cdot (1 - \rho) \cdot \left(\sum_{j=1}^{p} \frac{\pi_j}{ns_j^2}\right)^{-1}, \;\; \sigma^2 = \eta \cdot \rho \cdot \left(\sum_{j=1}^{p} \frac{\pi_j \cdot \sum_{g \in \mathbf{O}_j} w_{jg}^2}{ns_j^2}\right)^{-1}, \tag{28}$$

then $\eta = \mathrm{E}[V(\mathbf{X}\boldsymbol{\beta})] \,/\, \mathrm{E}[V(\mathbf{y})]$ and $\rho = \mathrm{E}^{\mathsf{net}}[V(\mathbf{X}\boldsymbol{\beta})] \,/\, \mathrm{E}[V(\mathbf{X}\boldsymbol{\beta})]$.

PROOF. It suffices to show that, for all $j = 1, \ldots, p$,

$$ns_j^2 = \sigma_y^2 \,/\, \sigma_{x,j}^2 = \mathrm{E}[V(\mathbf{y})] \,/\, \mathrm{E}[V(\mathbf{X}_j)]. \tag{29}$$

Use PROPOSITION 2 to complete the proof. ■

REMARK 2. PROPOSITION 3 provides the mathematical basis to interpret $\{\eta, \rho\}$. Specifically, η represents the proportion of the total phenotypic variation explained by p SNPs, and ρ represents the proportion of total genetic variation explained by network annotations $\{\mathbf{O}_j, w_{jg}\}$.

REMARK 3. Since $ns_j^2 = \sigma_y^2 \,/\, \sigma_{x,j}^2$, where σ_y^2 is the population variance of phenotype and $\sigma_{x,j}^2$ is the population variance of genotype at SNP j, this re-parameterization based on $\{\eta, \rho\}$ ensures that the induced prior of $\{\sigma_0^2, \sigma^2\}$ in Equation (28) does not depend on GWAS sample size n, and the resulting prior genetic effects ($\boldsymbol{\beta}$) have the same measurement unit as the phenotype.

REMARK 4. The induced prior of $\{\sigma_0^2, \sigma^2\}$ in Equation (28) contains a quantity $ns_j^2 = \sigma_y^2 \,/\, \sigma_{x,j}^2$ that depends on unknown population parameters $\{\sigma_y^2, \sigma_{x,j}^2\}$. As shown in [2], ns_j^2 can be reliably estimated from GWAS summary data: $ns_j^2 \approx n\hat{s}_j^2$. In practice, we replace the unknown ns_j^2 in Equation (28) with observed $n\hat{s}_j^2$; see main text Equation (9).

Supplementary Note 8

This study uses data generated by the WTCCC, 1000 Genomes Project, ENCODE Consortium, GTEx Project, DICE Project, eQTLGen Consortium, and multiple GWAS consortia. We thank them for making their data publicly available. Detailed acknowledgments and data sources are listed below.

- Wellcome Trust Case Control Consortium (WTCCC). This study uses individual-level genotype data generated by WTCCC to design simulations. A full list of the investigators who contributed to the generation of the data is available from `https://www.wtccc.org.uk/`. The WTCCC data are available at the European Genome-phenome Archive (`https://www.ebi.ac.uk/ega/`).
- 1000 Genomes Project. This study uses haplotypes of individuals with European ancestry from the 1000 Genomes Project Phase 3 to estimate LD. The 1000 Genomes Phase 3 haplotype data have been contributed by 1000 Genomes investigators and have been downloaded from `ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/`.
- Encyclopedia of DNA Elements (ENCODE) Consortium. This study uses gene expression and chromatin accessibility data generated by the ENCODE Consortium and the ENCODE production laboratories. The ENCODE data are available at the ENCODE portal (`https://www.encodeproject.org/`).
- Genotype-Tissue Expression (GTEx) Project. This study uses cis-eQTL data generated by the GTEx Project. The GTEx Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The GTEx data are available at the GTEx Portal (`https://gtexportal.org/home/`).

- Database of Immune Cell Expression, Expression quantitative trait loci and Epigenomics (DICE) Project. This study uses cis-eQTL data generated by the DICE Project. The DICE Project was supported by NIH R24AI108564. The DICE data are available at https://dice-database.org/.

- eQTLGen Consortium. This study uses cis-eQTL data generated by the eQTLGen Consortium. The eQTLGen data are available at http://www.eqtlgen.org/.

- Genetic Investigation of ANthropometric Traits (GIANT) Consortium. GWAS summary statistics on adult human height [3], body mass index [4] and body fat distribution [5] have been contributed by GIANT investigators and have been downloaded from http://portals.broadinstitute.org/collaboration/giant.

- Psychiatric Genomics Consortium (PGC). GWAS summary statistics on schizophrenia [6] have been contributed by PGC investigators and have been downloaded from http://www.med.unc.edu/pgc.

- International Inflammatory Bowel Disease Genetics Consortium (IIBDGC). GWAS summary statistics on inflammatory bowel disease, Crohn's disease and ulcerative colitis [7] have been contributed by IIBDGC investigators and have been downloaded from https://www.ibdgenetics.org.

- Coronary ARtery DIsease Genome wide Replication and Meta-analysis (CARDIoGRAM) plus The Coronary Artery Disease (C4D) Genetics (CARDIoGRAMplusC4D) Consortium. GWAS summary statistics on coronary artery disease and myocardial infarction [8] have been contributed by CARDIoGRAMplusC4D investigators and downloaded from http://www.cardiogramplusc4d.org/.

- GWAS summary statistics of heart rate. GWAS summary statistics on heart rate have been contributed by authors of [9] and have been downloaded from https://walker05.u.hpc.mssm.edu/.

- International Genomics of Alzheimer's Project (IGAP). GWAS summary statistics on Alzheimer's disease [10] have been contributed by IGAP investigators and have been downloaded from http://web.pasteur-lille.fr/en/recherche/u744/igap/igap_download.php. We thank the International Genomics of Alzheimer's Project (IGAP) for providing summary results data for these analyses. The investigators within IGAP contributed to the design and implementation of IGAP and/or provided data but did not participate in analysis or writing of this report. IGAP was made possible by the generous participation of the control subjects, the patients, and their families. The i-Select chips was funded by the French National Foundation on Alzheimer's disease and related disorders. EADI was supported by the LABEX (laboratory of excellence program investment for the future) DISTALZ grant, Inserm, Institut Pasteur de Lille, Universite de Lille 2 and the Lille University Hospital. GERAD was supported by the Medical Research Council (Grant no. 503480), Alzheimer's Research UK (Grant no. 503176), the Wellcome Trust (Grant no. 082604/2/07/Z) and German Federal Ministry of Education and Research (BMBF): Competence Network Dementia (CND) grant no. 01GI0102, 01GI0711, 01GI0420. CHARGE was partly supported by the NIH/NIA grant R01 AG033193 and the NIA AG081220 and AGES contract N01-AG-12100, the NHLBI grant R01 HL105756, the Icelandic Heart Association, and the Erasmus Medical Center and Erasmus University. ADGC was supported by the NIH/NIA grants: U01 AG032984, U24 AG021886, U01 AG016976,

and the Alzheimer's Association grant ADGC-10-196728.

- Social Science Genetic Association Consortium (SSGAC). GWAS summary statistics on neuroticism [11] have been contributed by SSGAC investigators and have been downloaded from https://www.thessgac.org. For financial support, the SSGAC thanks the U.S. National Science Foundation, the U.S. National Institutes of Health (National Institute on Aging, and the Office for Behavioral and Social Science Research), the Ragnar Soderberg Foundation, the Swedish Research Council, The Jan Wallander and Tom Hedelius Foundation, the European Research Council, and the Pershing Square Fund of the Foundations of Human Behavior.
- GWAS summary statistics of rheumatoid arthritis. GWAS summary statistics on rheumatoid arthritis have been contributed by authors of [12] and have been downloaded from http://plaza.umin.ac.jp/yokada/datasource/software.htm.
- DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium. GWAS summary statistics on type 2 diabetes [13] have been contributed by DIAGRAM investigators and have been downloaded from http://www.diagram-consortium.org.
- Global Lipids Genetics Consortium (GLGC). GWAS summary statistics on low-density and high-density lipoprotein cholesterol [14] have been contributed by GLGC investigators and have been downloaded from http://csg.sph.umich.edu//abecasis/public/lipids2010.
- GWAS summary statistics of atrial fibrillation. GWAS summary statistics on atrial fibrillation have been contributed by authors of [15] and are available from the corresponding author on reasonable request; see "Data availability" section of [15].
- Genetic Associations and Mechanisms in Oncology (GAME-ON), Discovery, Biology, and Risk of Inherited Variants in Breast Cancer (DRIVE) Project. GWAS summary statistics on breast cancer have been contributed by the GAME-ON/DRIVE Project and have been downloaded from http://gameon.dfci.harvard.edu. The DRIVE project acknowledges the following GWASs and investigators that shared genome-wide summary data as part of the breast-cancer GWAS meta-analysis: the Australian Breast Cancer Family Study (ABCFS) (John L. Hopper, Melissa C. Southey, Enes Makalic, Daniel F. Schmidt), the British Breast Cancer Study (BBCS) (Olivia Fletcher, Julian Peto, Lorna Gibson, Isabel dos Santos Silva), the Breast and Prostate Cancer Cohort Consortium (BPC3) (David J. Hunter, Sara Lindstrom, Peter Kraft), the Breast Cancer Family Registries (BCFR) (Habib Ahsan, Alice Whittemore), the Dutch Familial Bilateral Breast Cancer Study (DFBBCS) (Quinten Waisfisz, Hanne Meijers-Heijboer, Muriel Adank, Rob B. van der Luijt, Andre G. Uitterlinden, Albert Hofman), German Consortium for Hereditary Breast and Ovarian Cancer (GC-HBOC) (Alfons Meindl, Rita K. Schmutzler, Bertram Muller-Myhsok, Peter Lichtner), the Helsinki Breast Cancer Study (HEBCS) (Heli Nevanlinna, Taru A. Muranen, Kristiina Aittomaki, Carl Blomqvist), the Mammary Carcinoma Risk Factor Investigation (MARIE) (Jenny Chang-Claude, Rebecca Hein, Norbert Dahmen, Lars Beckman), SardiNIA (Laura Crisponi), the Singapore and Sweden Breast Cancer Study (SASBAC) (Per Hall, Kamila Czene, Astrid Irwanto, Jianjun Liu), and the UK2 (Douglas F. Easton, Clare Turnbull, Nazneen Rahman).

Supplementary Figures

Supplementary Figure 1

Simulation details and additional results of Figure 2. We use real genotypes of 348,965 genome-wide SNPs on chromosomes 1-22 from 1,458 individuals in the UK Blood Service Control Group [16] to simulate phenotype data, and then compute single-SNP association summary statistics. On these summary data, we compare RSS-NET with existing enrichment methods.

We infer a B cell regulatory network from the paired gene expression and chromatin accessibility data of primary B cells from peripheral blood [17, 18]. We use the B cell network to create SNP-level annotation for 348,965 SNPs. Specifically, we let $a_j = 1$ if SNP j is within 100 kb of either a regulatory element or the transcribed region of a member gene in the B cell network, and let $a_j = 0$ otherwise. There are 121,308 SNPs mapped to the B cell network with annotation $a_j = 1$.

To ensure that simulated datasets have roughly the same proportion of true genetic signals, we simulate causal indicators for negative and positive datasets in a paired way. We first simulate the causal indicator ζ_j of each SNP j in a positive dataset as:

$$\zeta_j \sim \text{Bernoulli}(\pi_j), \quad \pi_j = 1 \,/\, [1 + 10^{-(\theta_0 + a_j\theta)}],$$

where θ_0 and θ are the baseline and enrichment proportion parameters respectively. We then count the number of causal SNPs in this positive dataset as $n_c := \sum_j \zeta_j$, and randomly choose n_c SNPs as causal SNPs for the corresponding negative dataset.

For positive datasets, we simulate the true effect β_j of each SNP j as

$$\begin{aligned} \beta_j \mid \zeta_j = 0 &\sim \delta_0, \\ \beta_j \mid \zeta_j = 1 &\sim \mathcal{N}(0,\, \sigma_0^2 + \sigma^2 \cdot \textstyle\sum_{g \in \mathbf{O}_j} w_{jg}^2), \end{aligned}$$

where δ_0 denotes point mass at zero, σ_0^2 and σ^2 are the baseline and enrichment magnitude parameter respectively, $\{\mathbf{O}_j, w_{jg}\}$ are network annotations defined in main text Equations (4)-(5). For negative datasets, we simulate β_j from the same model above with $\sigma^2 = 0$.

To ensure that negative and positive datasets have roughly the same magnitude of total genetic signals, we simulate phenotypes by matching signal-to-noise ratios. Specifically, we simulate phenotype y_i of individual i as follows:

$$y_i = \textstyle\sum_{j=1}^{p} x_{ij} \cdot \beta_j + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \tau^{-1}),$$

where p is 348,965, x_{ij} is the genotype of SNP j for individual i. The true value of residual variance τ^{-1} is determined by the total proportion of variance in phenotype $\mathbf{y}$ explained by effects of all available SNPs in $\mathbf{X}$:

$$\text{PVE} = V(\mathbf{X}\boldsymbol{\beta}) \,/\, [\tau^{-1} + V(\mathbf{X}\boldsymbol{\beta})],$$

where $V(\mathbf{X}\boldsymbol{\beta})$ is the sample variance of $\mathbf{X}\boldsymbol{\beta}$.

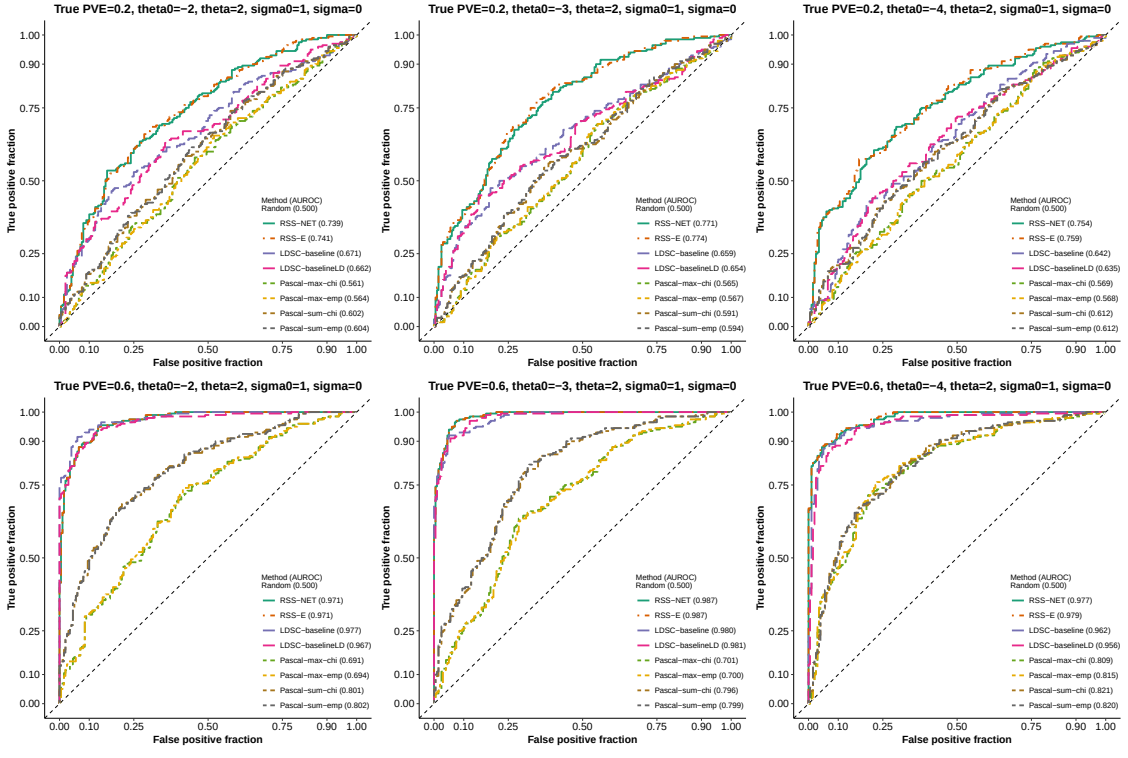
We set the true values of hyper-parameter as follows: baseline proportion parameter $\theta_0 \in \{-2, -3, -4\}$, enrichment proportion parameter $\theta \in \{0, 2\}$, baseline magnitude parameter

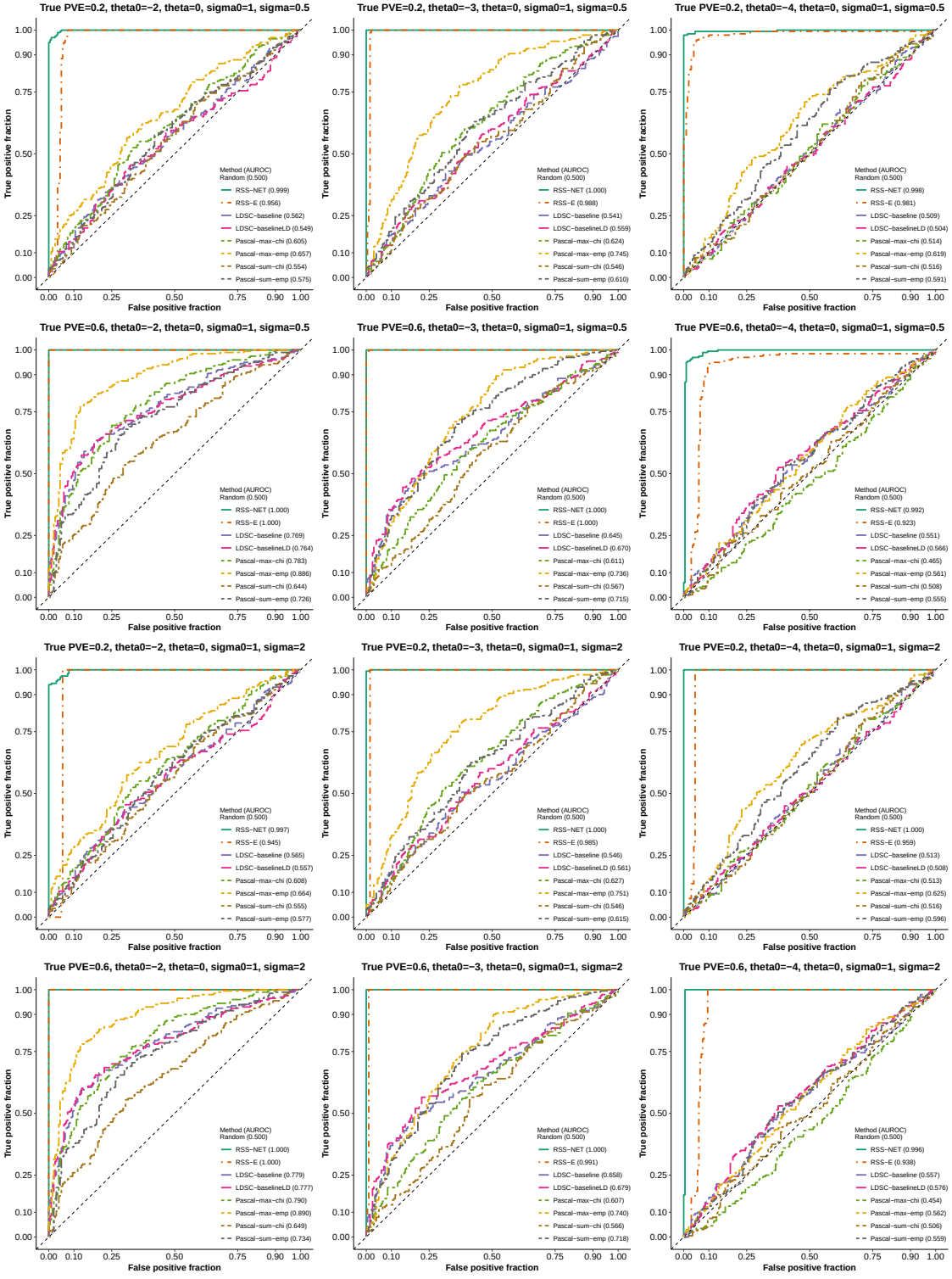
$\sigma_0 = 1$, enrichment magnitude parameter $\sigma \in \{0, 0.5, 2\}$, and PVE $\in \{0.2, 0.6\}$. For each combination of $\{\theta_0, \theta, \sigma_0, \sigma, \text{PVE}\}$ (excluding negative cases with $\theta = \sigma = 0$), we simulate 200 negative and 200 positive datasets. The caption of each panel in this supplementary figure shows the true hyper-parameter values of positive datasets. The table below lists true values of positive datasets for each panel in **Figure 2** ($\sigma_0 = 1$).

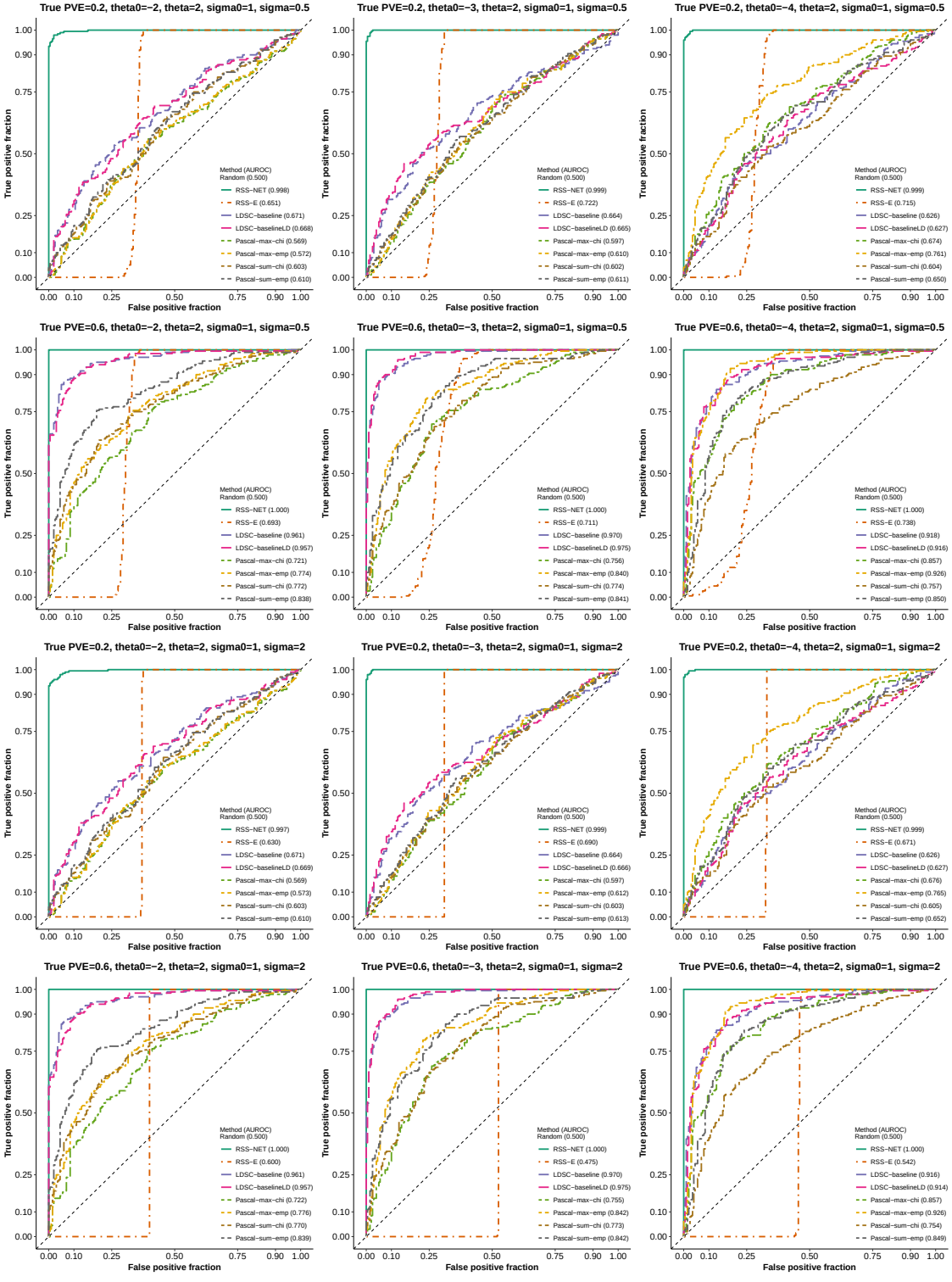
Figure 2 panel	a	b	c	d
sparse: $\theta_0 = -4$	$\theta = 2, \sigma = 0, \text{PVE} = 0.6$	$\theta = 0, \sigma = 0.5, \text{PVE} = 0.6$	$\theta = 2, \sigma = 0.5, \text{PVE} = 0.6$	$\theta = 2, \sigma = 0, \text{PVE} = 0.2$
polygenic: $\theta_0 = -2$	$\theta = 2, \sigma = 0, \text{PVE} = 0.6$	$\theta = 0, \sigma = 0.5, \text{PVE} = 0.6$	$\theta = 2, \sigma = 0.5, \text{PVE} = 0.6$	$\theta = 2, \sigma = 0, \text{PVE} = 0.2$

We apply RSS-NET to simulated datasets with the following hyper-parameter grids. The baseline parameters θ_0 and σ_0 are fixed as their true values. The grids on the enrichment parameters θ and σ are `(0:0.25:1)` if their true values are zero; otherwise they are `[0 ((truth-0.5):0.25:(truth+0.5))]`. For the same simulation scenario, we use the same hyper-parameter grid in RSS-NET analyses of all negative and positive datasets.

For all datasets the target of enrichment testing is the B cell network. A false positive occurs if a method identifies enrichment in a negative dataset. A true positive occurs if a method identifies enrichment in a positive dataset. We evaluate the performance of enrichment methods by plotting the receiver operating characteristic (ROC) curve and computing the area under the ROC curve (AUROC) for each method. Both metrics are implemented in the R package `plotROC` [19].

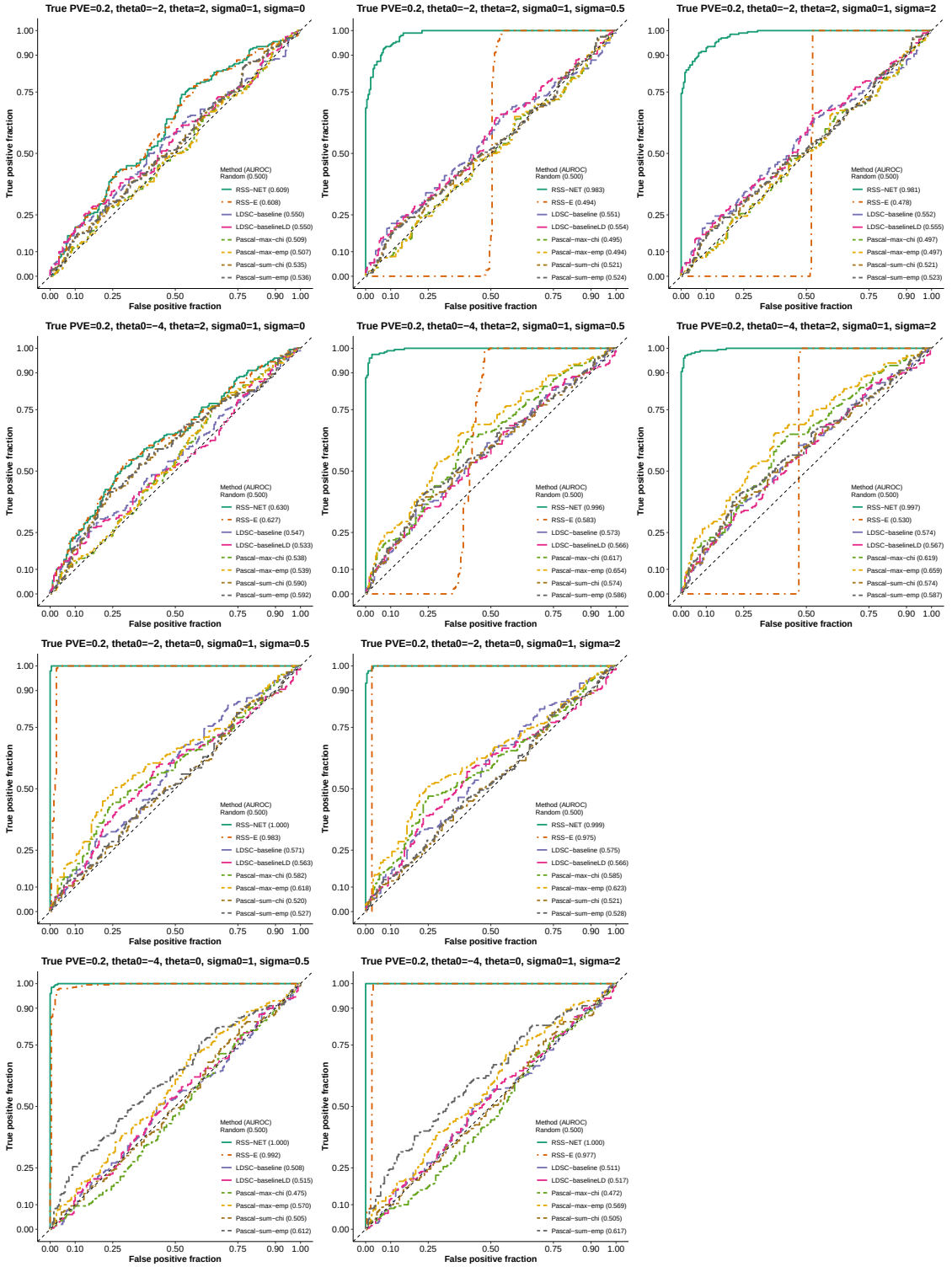



True PVE=0.2, theta0=−2, theta=0, sigma0=1, sigma=0.5
Method (AUROC)
Random (0.500)
RSS−NET (0.999)
RSS−E (0.956)
LDSC−baseline (0.562)
LDSC−baselineLD (0.549)
Pascal−max−chi (0.605)
Pascal−max−emp (0.657)
Pascal−sum−chi (0.554)
Pascal−sum−emp (0.575)
True PVE=0.2, theta0=−3, theta=0, sigma0=1, sigma=0.5
Method (AUROC)
Random (0.500)
RSS−NET (1.000)
RSS−E (0.988)
LDSC−baseline (0.541)
LDSC−baselineLD (0.559)
Pascal−max−chi (0.624)
Pascal−max−emp (0.745)
Pascal−sum−chi (0.546)
Pascal−sum−emp (0.610)
True PVE=0.2, theta0=−4, theta=0, sigma0=1, sigma=0.5
Method (AUROC)
Random (0.500)
RSS−NET (0.998)
RSS−E (0.981)
LDSC−baseline (0.509)
LDSC−baselineLD (0.504)
Pascal−max−chi (0.514)
Pascal−max−emp (0.619)
Pascal−sum−chi (0.516)
Pascal−sum−emp (0.591)
True PVE=0.6, theta0=−2, theta=0, sigma0=1, sigma=0.5
Method (AUROC)
Random (0.500)
RSS−NET (1.000)
RSS−E (1.000)
LDSC−baseline (0.769)
LDSC−baselineLD (0.764)
Pascal−max−chi (0.783)
Pascal−max−emp (0.886)
Pascal−sum−chi (0.644)
Pascal−sum−emp (0.726)
True PVE=0.6, theta0=−3, theta=0, sigma0=1, sigma=0.5
Method (AUROC)
Random (0.500)
RSS−NET (1.000)
RSS−E (1.000)
LDSC−baseline (0.645)
LDSC−baselineLD (0.670)
Pascal−max−chi (0.611)
Pascal−max−emp (0.736)
Pascal−sum−chi (0.567)
Pascal−sum−emp (0.715)
True PVE=0.6, theta0=−4, theta=0, sigma0=1, sigma=0.5
Method (AUROC)
Random (0.500)
RSS−NET (0.992)
RSS−E (0.923)
LDSC−baseline (0.551)
LDSC−baselineLD (0.566)
Pascal−max−chi (0.465)
Pascal−max−emp (0.561)
Pascal−sum−chi (0.508)
Pascal−sum−emp (0.555)
True PVE=0.2, theta0=−2, theta=0, sigma0=1, sigma=2
Method (AUROC)
Random (0.500)
RSS−NET (0.997)
RSS−E (0.945)
LDSC−baseline (0.565)
LDSC−baselineLD (0.557)
Pascal−max−chi (0.608)
Pascal−max−emp (0.664)
Pascal−sum−chi (0.555)
Pascal−sum−emp (0.577)
True PVE=0.2, theta0=−3, theta=0, sigma0=1, sigma=2
Method (AUROC)
Random (0.500)
RSS−NET (1.000)
RSS−E (0.985)
LDSC−baseline (0.546)
LDSC−baselineLD (0.561)
Pascal−max−chi (0.627)
Pascal−max−emp (0.751)
Pascal−sum−chi (0.546)
Pascal−sum−emp (0.615)
True PVE=0.2, theta0=−4, theta=0, sigma0=1, sigma=2
Method (AUROC)
Random (0.500)
RSS−NET (1.000)
RSS−E (0.959)
LDSC−baseline (0.513)
LDSC−baselineLD (0.508)
Pascal−max−chi (0.513)
Pascal−max−emp (0.625)
Pascal−sum−chi (0.516)
Pascal−sum−emp (0.596)
True PVE=0.6, theta0=−2, theta=0, sigma0=1, sigma=2
Method (AUROC)
Random (0.500)
RSS−NET (1.000)
RSS−E (1.000)
LDSC−baseline (0.779)
LDSC−baselineLD (0.777)
Pascal−max−chi (0.790)
Pascal−max−emp (0.890)
Pascal−sum−chi (0.649)
Pascal−sum−emp (0.734)
True PVE=0.6, theta0=−3, theta=0, sigma0=1, sigma=2
Method (AUROC)
Random (0.500)
RSS−NET (1.000)
RSS−E (0.991)
LDSC−baseline (0.658)
LDSC−baselineLD (0.679)
Pascal−max−chi (0.607)
Pascal−max−emp (0.740)
Pascal−sum−chi (0.566)
Pascal−sum−emp (0.718)
True PVE=0.6, theta0=−4, theta=0, sigma0=1, sigma=2
Method (AUROC)
Random (0.500)
RSS−NET (0.996)
RSS−E (0.938)
LDSC−baseline (0.557)
LDSC−baselineLD (0.576)
Pascal−max−chi (0.454)
Pascal−max−emp (0.562)
Pascal−sum−chi (0.506)
Pascal−sum−emp (0.559)
True positive fraction
False positive fraction
0.00
0.10
0.25
0.50
0.75
0.90
1.00


True PVE=0.2, theta0=−2, theta=2, sigma0=1, sigma=0.5
Method (AUROC)
Random (0.500)
RSS−NET (0.998)
RSS−E (0.651)
LDSC−baseline (0.671)
LDSC−baselineLD (0.668)
Pascal−max−chi (0.569)
Pascal−max−emp (0.572)
Pascal−sum−chi (0.603)
Pascal−sum−emp (0.610)
True PVE=0.2, theta0=−3, theta=2, sigma0=1, sigma=0.5
Method (AUROC)
Random (0.500)
RSS−NET (0.999)
RSS−E (0.722)
LDSC−baseline (0.664)
LDSC−baselineLD (0.665)
Pascal−max−chi (0.597)
Pascal−max−emp (0.610)
Pascal−sum−chi (0.602)
Pascal−sum−emp (0.611)
True PVE=0.2, theta0=−4, theta=2, sigma0=1, sigma=0.5
Method (AUROC)
Random (0.500)
RSS−NET (0.999)
RSS−E (0.715)
LDSC−baseline (0.626)
LDSC−baselineLD (0.627)
Pascal−max−chi (0.674)
Pascal−max−emp (0.761)
Pascal−sum−chi (0.604)
Pascal−sum−emp (0.650)
True PVE=0.6, theta0=−2, theta=2, sigma0=1, sigma=0.5
Method (AUROC)
Random (0.500)
RSS−NET (1.000)
RSS−E (0.693)
LDSC−baseline (0.961)
LDSC−baselineLD (0.957)
Pascal−max−chi (0.721)
Pascal−max−emp (0.774)
Pascal−sum−chi (0.772)
Pascal−sum−emp (0.838)
True PVE=0.6, theta0=−3, theta=2, sigma0=1, sigma=0.5
Method (AUROC)
Random (0.500)
RSS−NET (1.000)
RSS−E (0.711)
LDSC−baseline (0.970)
LDSC−baselineLD (0.975)
Pascal−max−chi (0.756)
Pascal−max−emp (0.840)
Pascal−sum−chi (0.774)
Pascal−sum−emp (0.841)
True PVE=0.6, theta0=−4, theta=2, sigma0=1, sigma=0.5
Method (AUROC)
Random (0.500)
RSS−NET (1.000)
RSS−E (0.738)
LDSC−baseline (0.918)
LDSC−baselineLD (0.916)
Pascal−max−chi (0.857)
Pascal−max−emp (0.926)
Pascal−sum−chi (0.757)
Pascal−sum−emp (0.850)
True PVE=0.2, theta0=−2, theta=2, sigma0=1, sigma=2
Method (AUROC)
Random (0.500)
RSS−NET (0.997)
RSS−E (0.630)
LDSC−baseline (0.671)
LDSC−baselineLD (0.669)
Pascal−max−chi (0.569)
Pascal−max−emp (0.573)
Pascal−sum−chi (0.603)
Pascal−sum−emp (0.610)
True PVE=0.2, theta0=−3, theta=2, sigma0=1, sigma=2
Method (AUROC)
Random (0.500)
RSS−NET (0.999)
RSS−E (0.690)
LDSC−baseline (0.664)
LDSC−baselineLD (0.666)
Pascal−max−chi (0.597)
Pascal−max−emp (0.612)
Pascal−sum−chi (0.603)
Pascal−sum−emp (0.613)
True PVE=0.2, theta0=−4, theta=2, sigma0=1, sigma=2
Method (AUROC)
Random (0.500)
RSS−NET (0.999)
RSS−E (0.671)
LDSC−baseline (0.626)
LDSC−baselineLD (0.627)
Pascal−max−chi (0.676)
Pascal−max−emp (0.765)
Pascal−sum−chi (0.605)
Pascal−sum−emp (0.652)
True PVE=0.6, theta0=−2, theta=2, sigma0=1, sigma=2
Method (AUROC)
Random (0.500)
RSS−NET (1.000)
RSS−E (0.600)
LDSC−baseline (0.961)
LDSC−baselineLD (0.957)
Pascal−max−chi (0.722)
Pascal−max−emp (0.776)
Pascal−sum−chi (0.770)
Pascal−sum−emp (0.839)
True PVE=0.6, theta0=−3, theta=2, sigma0=1, sigma=2
Method (AUROC)
Random (0.500)
RSS−NET (1.000)
RSS−E (0.475)
LDSC−baseline (0.970)
LDSC−baselineLD (0.975)
Pascal−max−chi (0.755)
Pascal−max−emp (0.842)
Pascal−sum−chi (0.773)
Pascal−sum−emp (0.842)
True PVE=0.6, theta0=−4, theta=2, sigma0=1, sigma=2
Method (AUROC)
Random (0.500)
RSS−NET (1.000)
RSS−E (0.542)
LDSC−baseline (0.916)
LDSC−baselineLD (0.914)
Pascal−max−chi (0.857)
Pascal−max−emp (0.926)
Pascal−sum−chi (0.754)
Pascal−sum−emp (0.849)
True positive fraction
False positive fraction
0.00
0.10
0.25
0.50
0.75
0.90
1.00

Supplementary Figure 2

We repeat the simulations in Supplementary Figure 1 on a different network based on data from vagina (`https://github.com/suwonglab/rss-net/tree/master/data/`).



Supplementary Figure 3

Simulation details and additional results of Figure 3a. This part aims to assess the robustness of RSS-NET to model mis-specification where a random set of "near-gene" SNPs is enriched for genetic association in negative datasets. Details of this part are almost identical to those in Supplementary Figure 1. Here we only highlight the differences.

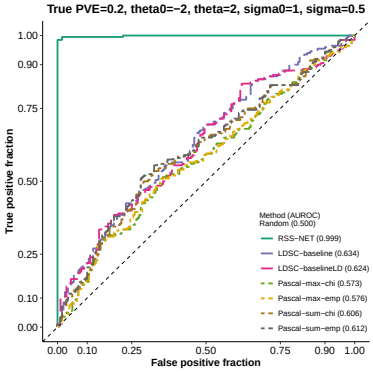
We define a SNP as "near-gene" if this SNP is within 100 kb of the transcribed region of any autosomal protein-coding gene. In total, there are 254,296 near-gene SNPs among 348,965 genome-wide SNPs from [16].

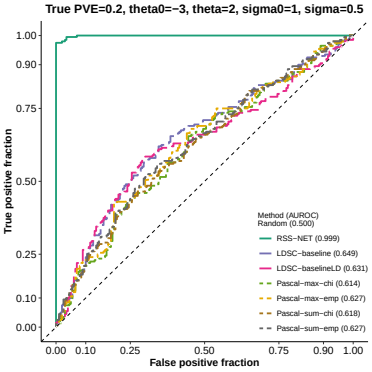
Here we simulate negative and positive datasets in a paired way. We first simulate a positive dataset as in Supplementary Figure 1. For this positive dataset, we count the total number of causal SNPs as $n_c = \sum_j \zeta_j$, and count the number of causal SNPs in the target network as $n_p = \sum_j \zeta_j a_j$ (recall that $a_j = 1$ if SNP j is in the network). We then randomly choose n_p SNPs from the 254,296 near-gene SNPs and $(n_c - n_p)$ SNPs from the remaining SNPs, and use them as causal SNPs for the corresponding negative dataset. With causal indicators $\{\zeta_j\}$ in place, we simulate the true genetic effects $\boldsymbol{\beta}$ in a negative dataset as follows

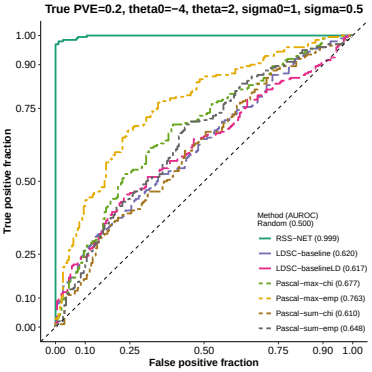
$$\begin{aligned} \beta_j \mid \zeta_j = 0 &\sim \delta_0, \\ \beta_j \mid \zeta_j = 1,\ \alpha_j = 0 &\sim \mathcal{N}(0,\ \sigma_0^2), \\ \beta_j \mid \zeta_j = 1,\ \alpha_j = 1 &\sim \mathcal{N}(0,\ (2 \cdot \sigma_0)^2), \end{aligned}$$

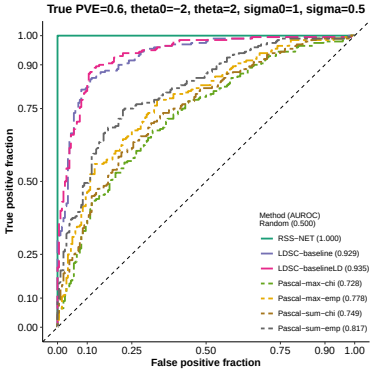
where $\zeta_j = 1$ if SNP j is causal and 0 otherwise, $\alpha_j = 1$ if SNP j is near-gene and 0 otherwise. The rest of simulations is the same as Supplementary Figure 1.

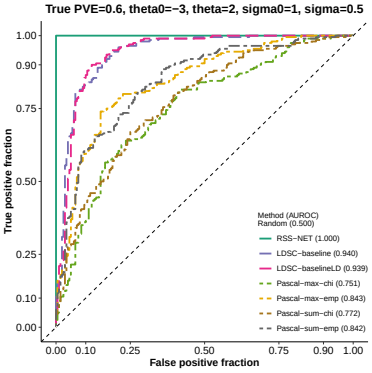
The caption of each panel in this supplementary figure shows the true hyper-parameter values of positive datasets. The sparse scenario in Figure 3a corresponds to simulations with true hyper-parameter values of positive datasets being $\theta_0 = -4$, $\theta = 2$, $\sigma_0 = 1$, $\sigma = 0.5$ and PVE = 0.6. The polygenic scenario in Figure 3a corresponds to simulations with true hyper-parameter values of positive datasets being $\theta_0 = -2$, $\theta = 2$, $\sigma_0 = 1$, $\sigma = 0.5$ and PVE = 0.6.

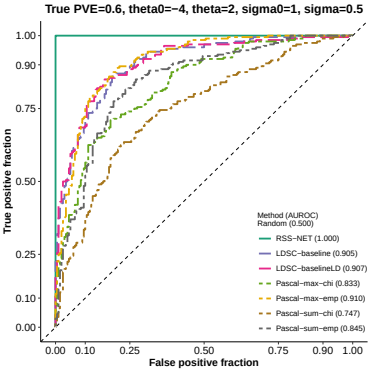

True PVE=0.2, theta0=−2, theta=2, sigma0=1, sigma=0.5
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.999)
LDSC−baseline (0.634)
LDSC−baselineLD (0.624)
Pascal−max−chi (0.573)
Pascal−max−emp (0.576)
Pascal−sum−chi (0.606)
Pascal−sum−emp (0.612)

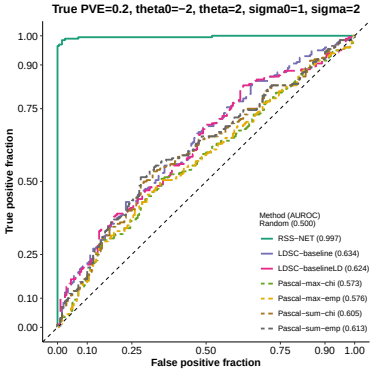

True PVE=0.2, theta0=−3, theta=2, sigma0=1, sigma=0.5
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.999)
LDSC−baseline (0.649)
LDSC−baselineLD (0.631)
Pascal−max−chi (0.614)
Pascal−max−emp (0.627)
Pascal−sum−chi (0.618)
Pascal−sum−emp (0.627)

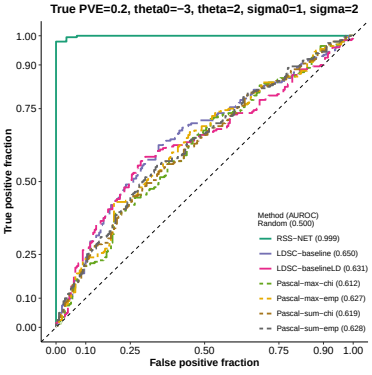

True PVE=0.2, theta0=−4, theta=2, sigma0=1, sigma=0.5
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.999)
LDSC−baseline (0.620)
LDSC−baselineLD (0.617)
Pascal−max−chi (0.677)
Pascal−max−emp (0.763)
Pascal−sum−chi (0.610)
Pascal−sum−emp (0.648)

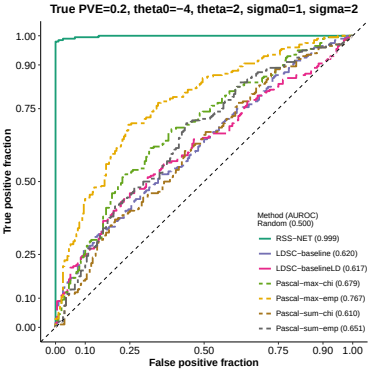

True PVE=0.6, theta0=−2, theta=2, sigma0=1, sigma=0.5
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (1.000)
LDSC−baseline (0.929)
LDSC−baselineLD (0.935)
Pascal−max−chi (0.728)
Pascal−max−emp (0.778)
Pascal−sum−chi (0.749)
Pascal−sum−emp (0.817)

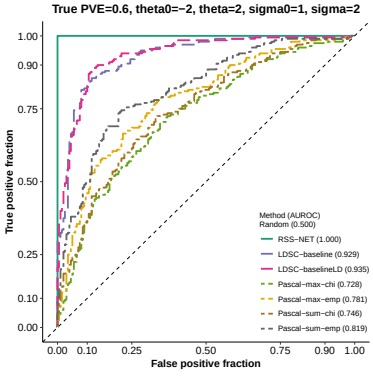

True PVE=0.6, theta0=−3, theta=2, sigma0=1, sigma=0.5
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (1.000)
LDSC−baseline (0.940)
LDSC−baselineLD (0.939)
Pascal−max−chi (0.751)
Pascal−max−emp (0.843)
Pascal−sum−chi (0.772)
Pascal−sum−emp (0.842)

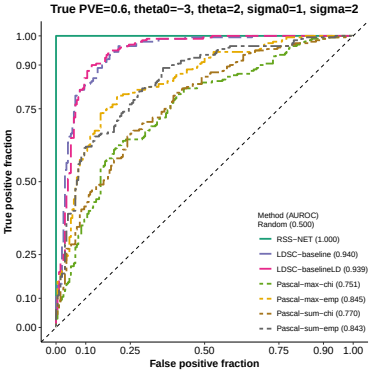

True PVE=0.6, theta0=−4, theta=2, sigma0=1, sigma=0.5
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (1.000)
LDSC−baseline (0.905)
LDSC−baselineLD (0.907)
Pascal−max−chi (0.833)
Pascal−max−emp (0.910)
Pascal−sum−chi (0.747)
Pascal−sum−emp (0.845)

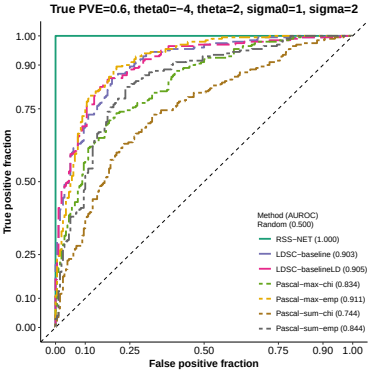

True PVE=0.2, theta0=−2, theta=2, sigma0=1, sigma=2
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.997)
LDSC−baseline (0.634)
LDSC−baselineLD (0.624)
Pascal−max−chi (0.573)
Pascal−max−emp (0.576)
Pascal−sum−chi (0.605)
Pascal−sum−emp (0.613)


True PVE=0.2, theta0=−3, theta=2, sigma0=1, sigma=2
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.999)
LDSC−baseline (0.650)
LDSC−baselineLD (0.631)
Pascal−max−chi (0.612)
Pascal−max−emp (0.627)
Pascal−sum−chi (0.619)
Pascal−sum−emp (0.628)


True PVE=0.2, theta0=−4, theta=2, sigma0=1, sigma=2
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.999)
LDSC−baseline (0.620)
LDSC−baselineLD (0.617)
Pascal−max−chi (0.679)
Pascal−max−emp (0.767)
Pascal−sum−chi (0.610)
Pascal−sum−emp (0.651)


True PVE=0.6, theta0=−2, theta=2, sigma0=1, sigma=2
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (1.000)
LDSC−baseline (0.929)
LDSC−baselineLD (0.935)
Pascal−max−chi (0.728)
Pascal−max−emp (0.781)
Pascal−sum−chi (0.746)
Pascal−sum−emp (0.819)


True PVE=0.6, theta0=−3, theta=2, sigma0=1, sigma=2
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (1.000)
LDSC−baseline (0.940)
LDSC−baselineLD (0.939)
Pascal−max−chi (0.751)
Pascal−max−emp (0.845)
Pascal−sum−chi (0.770)
Pascal−sum−emp (0.843)


True PVE=0.6, theta0=−4, theta=2, sigma0=1, sigma=2
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (1.000)
LDSC−baseline (0.903)
LDSC−baselineLD (0.905)
Pascal−max−chi (0.834)
Pascal−max−emp (0.911)
Pascal−sum−chi (0.744)
Pascal−sum−emp (0.844)

Supplementary Figure 4

Simulation details and additional results of Figure 3b. This part aims to assess the robustness of RSS-NET to model mis-specification where a random set of "near-RE" SNPs is enriched for genetic association in negative datasets. Details of this part are almost identical to those in Supplementary Figure 1. Here we only highlight the differences.

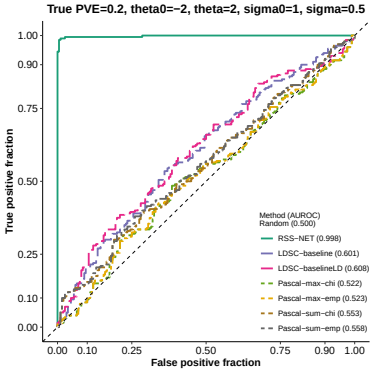
We define a SNP as "near-RE" if this SNP is within 10 kb of any RE. There are 3,895,021 unique autosomal REs associated with 38 regulatory networks analyzed in this study. In total, there are 169,100 near-RE SNPs among 348,965 genome-wide SNPs from [16].

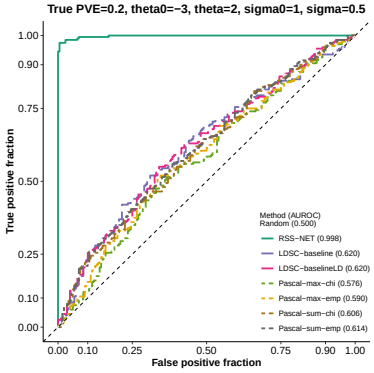
Here we simulate negative and positive datasets in a paired way. We first simulate a positive dataset as in Supplementary Figure 1. For this positive dataset, we count the total number of causal SNPs as $n_c = \sum_j \zeta_j$, and count the number of causal SNPs in the target network as $n_p = \sum_j \zeta_j a_j$ (recall that $a_j = 1$ if SNP j is in the network). We then randomly choose n_p SNPs from the 169,100 near-RE SNPs and $(n_c - n_p)$ SNPs from the remaining SNPs, and use them as causal SNPs for the corresponding negative dataset. With causal indicators $\{\zeta_j\}$ in place, we simulate the true genetic effects $\boldsymbol{\beta}$ in a negative dataset as follows

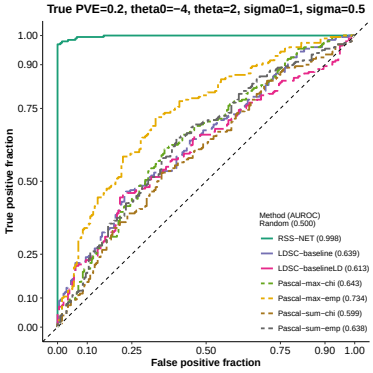
$$\begin{aligned} \beta_j \mid \zeta_j = 0 &\sim \delta_0, \\ \beta_j \mid \zeta_j = 1,\ \alpha_j = 0 &\sim \mathcal{N}(0,\ \sigma_0^2), \\ \beta_j \mid \zeta_j = 1,\ \alpha_j = 1 &\sim \mathcal{N}(0,\ (2 \cdot \sigma_0)^2), \end{aligned}$$

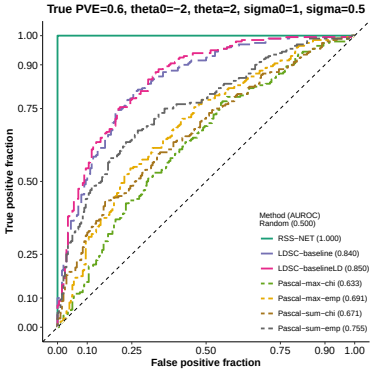
where $\zeta_j = 1$ if SNP j is causal and 0 otherwise, $\alpha_j = 1$ if SNP j is near-RE and 0 otherwise. The rest of simulations is the same as Supplementary Figure 1.

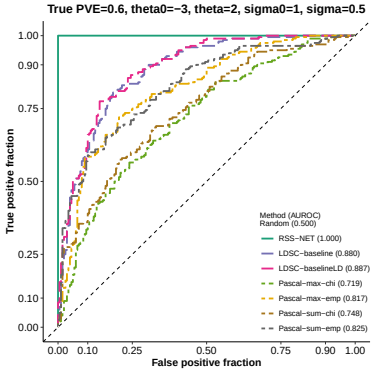
The caption of each panel in this supplementary figure shows the true hyper-parameter values of positive datasets. The sparse scenario in Figure 3b corresponds to simulations with true hyper-parameter values of positive datasets being $\theta_0 = -4$, $\theta = 2$, $\sigma_0 = 1$, $\sigma = 0.5$ and PVE = 0.6. The polygenic scenario in Figure 3b corresponds to simulations with true hyper-parameter values of positive datasets being $\theta_0 = -2$, $\theta = 2$, $\sigma_0 = 1$, $\sigma = 0.5$ and PVE = 0.6.

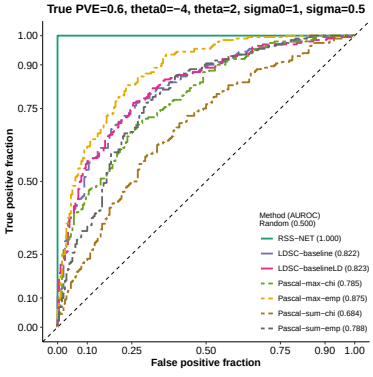

True PVE=0.2, theta0=−2, theta=2, sigma0=1, sigma=0.5
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.998)
LDSC−baseline (0.601)
LDSC−baselineLD (0.608)
Pascal−max−chi (0.522)
Pascal−max−emp (0.523)
Pascal−sum−chi (0.553)
Pascal−sum−emp (0.558)

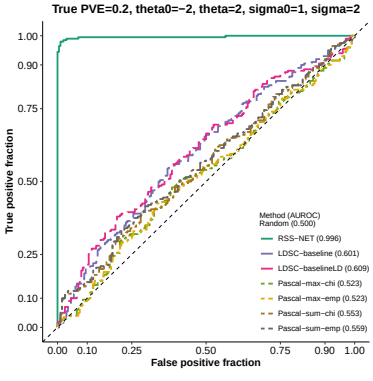

True PVE=0.2, theta0=−3, theta=2, sigma0=1, sigma=0.5
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.998)
LDSC−baseline (0.620)
LDSC−baselineLD (0.620)
Pascal−max−chi (0.576)
Pascal−max−emp (0.590)
Pascal−sum−chi (0.606)
Pascal−sum−emp (0.614)

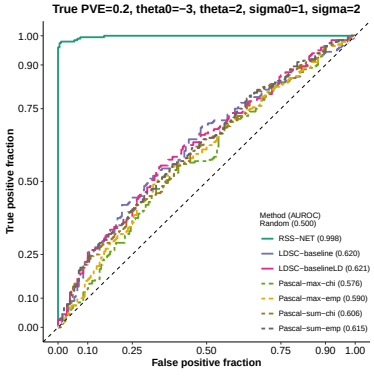

True PVE=0.2, theta0=−4, theta=2, sigma0=1, sigma=0.5
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.998)
LDSC−baseline (0.639)
LDSC−baselineLD (0.613)
Pascal−max−chi (0.643)
Pascal−max−emp (0.734)
Pascal−sum−chi (0.599)
Pascal−sum−emp (0.638)

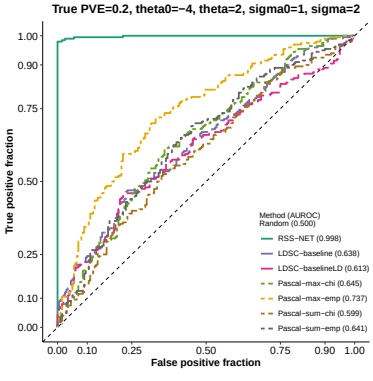

True PVE=0.6, theta0=−2, theta=2, sigma0=1, sigma=0.5
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (1.000)
LDSC−baseline (0.840)
LDSC−baselineLD (0.850)
Pascal−max−chi (0.633)
Pascal−max−emp (0.691)
Pascal−sum−chi (0.671)
Pascal−sum−emp (0.755)

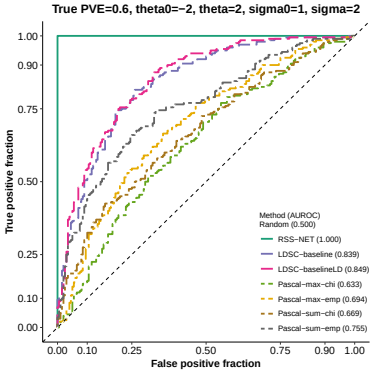

True PVE=0.6, theta0=−3, theta=2, sigma0=1, sigma=0.5
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (1.000)
LDSC−baseline (0.880)
LDSC−baselineLD (0.887)
Pascal−max−chi (0.719)
Pascal−max−emp (0.817)
Pascal−sum−chi (0.748)
Pascal−sum−emp (0.825)

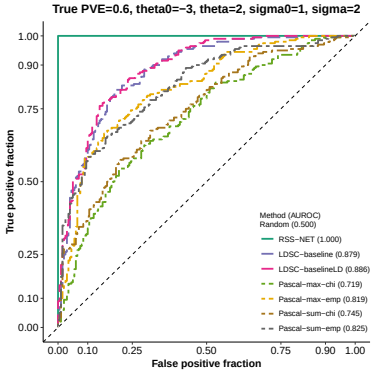

True PVE=0.6, theta0=−4, theta=2, sigma0=1, sigma=0.5
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (1.000)
LDSC−baseline (0.822)
LDSC−baselineLD (0.823)
Pascal−max−chi (0.785)
Pascal−max−emp (0.875)
Pascal−sum−chi (0.684)
Pascal−sum−emp (0.788)

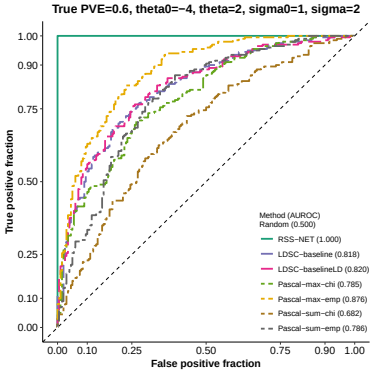

True PVE=0.2, theta0=−2, theta=2, sigma0=1, sigma=2
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.996)
LDSC−baseline (0.601)
LDSC−baselineLD (0.609)
Pascal−max−chi (0.523)
Pascal−max−emp (0.523)
Pascal−sum−chi (0.553)
Pascal−sum−emp (0.559)


True PVE=0.2, theta0=−3, theta=2, sigma0=1, sigma=2
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.998)
LDSC−baseline (0.620)
LDSC−baselineLD (0.621)
Pascal−max−chi (0.576)
Pascal−max−emp (0.590)
Pascal−sum−chi (0.606)
Pascal−sum−emp (0.615)


True PVE=0.2, theta0=−4, theta=2, sigma0=1, sigma=2
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.998)
LDSC−baseline (0.638)
LDSC−baselineLD (0.613)
Pascal−max−chi (0.645)
Pascal−max−emp (0.737)
Pascal−sum−chi (0.599)
Pascal−sum−emp (0.641)


True PVE=0.6, theta0=−2, theta=2, sigma0=1, sigma=2
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (1.000)
LDSC−baseline (0.839)
LDSC−baselineLD (0.849)
Pascal−max−chi (0.633)
Pascal−max−emp (0.694)
Pascal−sum−chi (0.669)
Pascal−sum−emp (0.755)


True PVE=0.6, theta0=−3, theta=2, sigma0=1, sigma=2
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (1.000)
LDSC−baseline (0.879)
LDSC−baselineLD (0.886)
Pascal−max−chi (0.719)
Pascal−max−emp (0.819)
Pascal−sum−chi (0.745)
Pascal−sum−emp (0.825)


True PVE=0.6, theta0=−4, theta=2, sigma0=1, sigma=2
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (1.000)
LDSC−baseline (0.818)
LDSC−baselineLD (0.820)
Pascal−max−chi (0.785)
Pascal−max−emp (0.876)
Pascal−sum−chi (0.682)
Pascal−sum−emp (0.786)

Supplementary Figure 5

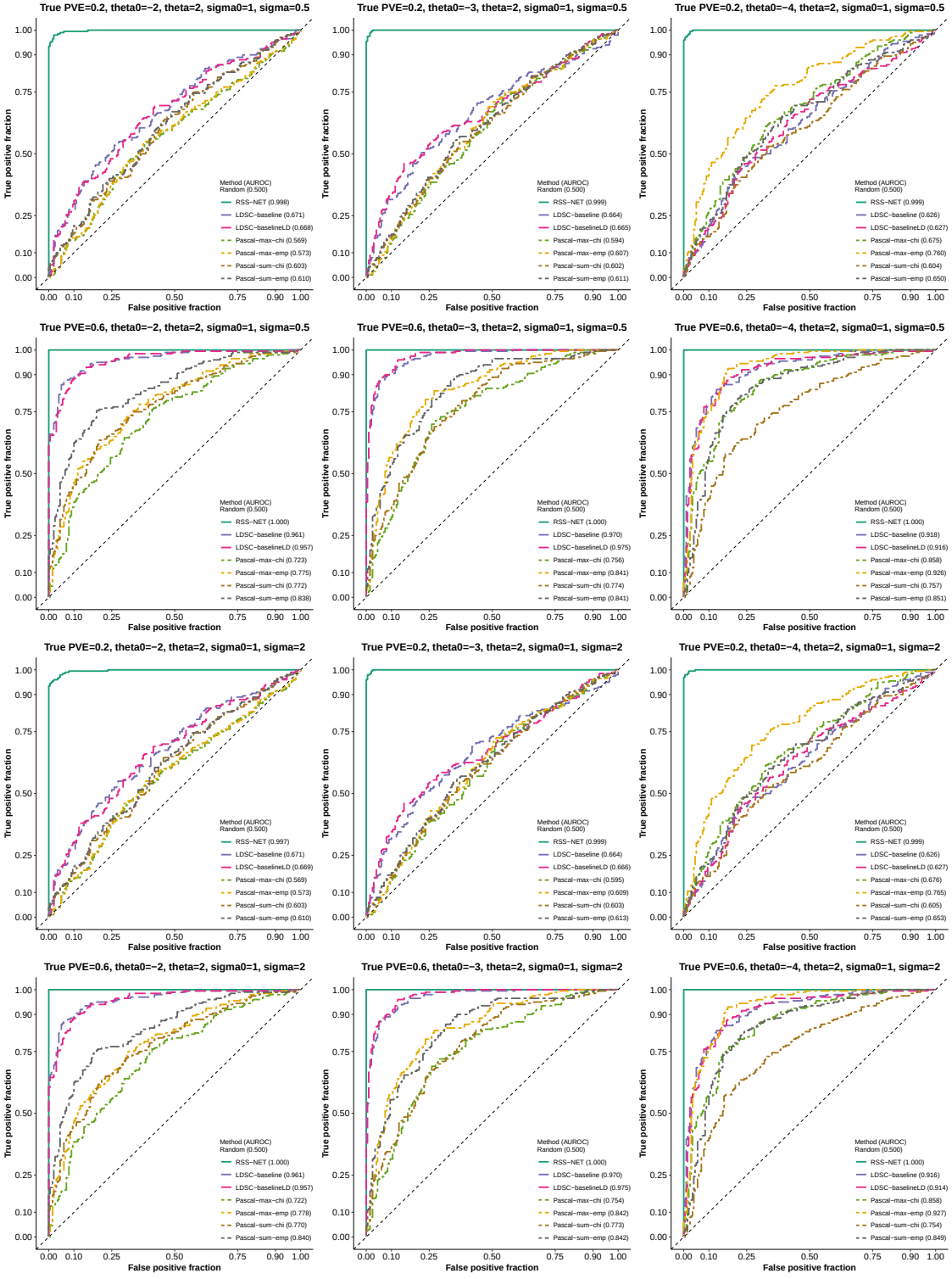
Simulation details and additional results of Figure 3c. This part aims to assess the robustness of RSS-NET to model mis-specification where true genetic effects are MAF- and LD-dependent in negative datasets. Details of this part are almost identical to those in Supplementary Figure 1. Here we only highlight the differences.

Here we simulate negative and positive datasets in a paired way. We first simulate a positive dataset as in **Supplementary Figure 1**. For this positive dataset, we count the total number of causal SNPs as $n_c = \sum_j \zeta_j$. We then randomly choose n_c genome-wide SNPs as causal SNPs for the corresponding negative dataset. With causal indicators $\{\zeta_j\}$ in place, we simulate the true genetic effects $\boldsymbol{\beta}$ in a negative dataset as follows

$$\begin{aligned} \beta_j \mid \zeta_j = 0 &\sim \delta_0, \\ \beta_j \mid \zeta_j = 1 &\sim \mathcal{N}(0,\, \sigma_0^2 + \textstyle\sum_{k=1}^{16} a_{jk} \cdot \tau_k), \\ \tau_k &\sim \mathcal{N}(\hat{\mu}_k,\, \hat{\sigma}_k^2), \end{aligned}$$

where $\zeta_j = 1$ if SNP j is causal and 0 otherwise, a_{jk} is the value of annotation k at SNP j, the 16 annotations consist of 10 MAF bins and 6 LD-related annotations defined in [20], $\{\hat{\mu}_k, \hat{\sigma}_k\}$ are meta-analyzed mean and standard error estimates for τ_k across 31 independent human traits, which are provided in the Supplementary Table 9 of [20]. The rest of simulations is the same as Supplementary Figure 1.

The caption of each panel in this supplementary figure shows the true hyper-parameter values of positive datasets. The sparse scenario in **Figure 3c** corresponds to simulations with true hyper-parameter values of positive datasets being $\theta_0 = -4$, $\theta = 2$, $\sigma_0 = 1$, $\sigma = 0.5$ and PVE = 0.6. The polygenic scenario in **Figure 3c** corresponds to simulations with true hyper-parameter values of positive datasets being $\theta_0 = -2$, $\theta = 2$, $\sigma_0 = 1$, $\sigma = 0.5$ and PVE = 0.6.


True PVE=0.2, theta0=−2, theta=2, sigma0=1, sigma=0.5
Method (AUROC)
Random (0.500)
RSS−NET (0.998)
LDSC−baseline (0.671)
LDSC−baselineLD (0.668)
Pascal−max−chi (0.569)
Pascal−max−emp (0.573)
Pascal−sum−chi (0.603)
Pascal−sum−emp (0.610)
True PVE=0.2, theta0=−3, theta=2, sigma0=1, sigma=0.5
Method (AUROC)
Random (0.500)
RSS−NET (0.999)
LDSC−baseline (0.664)
LDSC−baselineLD (0.665)
Pascal−max−chi (0.594)
Pascal−max−emp (0.607)
Pascal−sum−chi (0.602)
Pascal−sum−emp (0.611)
True PVE=0.2, theta0=−4, theta=2, sigma0=1, sigma=0.5
Method (AUROC)
Random (0.500)
RSS−NET (0.999)
LDSC−baseline (0.626)
LDSC−baselineLD (0.627)
Pascal−max−chi (0.675)
Pascal−max−emp (0.760)
Pascal−sum−chi (0.604)
Pascal−sum−emp (0.650)
True PVE=0.6, theta0=−2, theta=2, sigma0=1, sigma=0.5
Method (AUROC)
Random (0.500)
RSS−NET (1.000)
LDSC−baseline (0.961)
LDSC−baselineLD (0.957)
Pascal−max−chi (0.723)
Pascal−max−emp (0.775)
Pascal−sum−chi (0.772)
Pascal−sum−emp (0.838)
True PVE=0.6, theta0=−3, theta=2, sigma0=1, sigma=0.5
Method (AUROC)
Random (0.500)
RSS−NET (1.000)
LDSC−baseline (0.970)
LDSC−baselineLD (0.975)
Pascal−max−chi (0.756)
Pascal−max−emp (0.841)
Pascal−sum−chi (0.774)
Pascal−sum−emp (0.841)
True PVE=0.6, theta0=−4, theta=2, sigma0=1, sigma=0.5
Method (AUROC)
Random (0.500)
RSS−NET (1.000)
LDSC−baseline (0.918)
LDSC−baselineLD (0.916)
Pascal−max−chi (0.858)
Pascal−max−emp (0.926)
Pascal−sum−chi (0.757)
Pascal−sum−emp (0.851)
True PVE=0.2, theta0=−2, theta=2, sigma0=1, sigma=2
Method (AUROC)
Random (0.500)
RSS−NET (0.997)
LDSC−baseline (0.671)
LDSC−baselineLD (0.669)
Pascal−max−chi (0.569)
Pascal−max−emp (0.573)
Pascal−sum−chi (0.603)
Pascal−sum−emp (0.610)
True PVE=0.2, theta0=−3, theta=2, sigma0=1, sigma=2
Method (AUROC)
Random (0.500)
RSS−NET (0.999)
LDSC−baseline (0.664)
LDSC−baselineLD (0.666)
Pascal−max−chi (0.595)
Pascal−max−emp (0.609)
Pascal−sum−chi (0.603)
Pascal−sum−emp (0.613)
True PVE=0.2, theta0=−4, theta=2, sigma0=1, sigma=2
Method (AUROC)
Random (0.500)
RSS−NET (0.999)
LDSC−baseline (0.626)
LDSC−baselineLD (0.627)
Pascal−max−chi (0.676)
Pascal−max−emp (0.765)
Pascal−sum−chi (0.605)
Pascal−sum−emp (0.653)
True PVE=0.6, theta0=−2, theta=2, sigma0=1, sigma=2
Method (AUROC)
Random (0.500)
RSS−NET (1.000)
LDSC−baseline (0.961)
LDSC−baselineLD (0.957)
Pascal−max−chi (0.722)
Pascal−max−emp (0.778)
Pascal−sum−chi (0.770)
Pascal−sum−emp (0.840)
True PVE=0.6, theta0=−3, theta=2, sigma0=1, sigma=2
Method (AUROC)
Random (0.500)
RSS−NET (1.000)
LDSC−baseline (0.970)
LDSC−baselineLD (0.975)
Pascal−max−chi (0.754)
Pascal−max−emp (0.842)
Pascal−sum−chi (0.773)
Pascal−sum−emp (0.842)
True PVE=0.6, theta0=−4, theta=2, sigma0=1, sigma=2
Method (AUROC)
Random (0.500)
RSS−NET (1.000)
LDSC−baseline (0.916)
LDSC−baselineLD (0.914)
Pascal−max−chi (0.858)
Pascal−max−emp (0.927)
Pascal−sum−chi (0.754)
Pascal−sum−emp (0.849)
True positive fraction
False positive fraction
0.00
0.10
0.25
0.50
0.75
0.90
1.00

Supplementary Figure 6

Simulation details and additional results of Figure 3d. This part aims to assess the robustness of RSS-NET to model mis-specification where an edge-altered version of the true network is enriched for genetic association in negative datasets. Details of this part are almost identical to those in Supplementary Figure 1. Here we only highlight the differences.

We simulate random edge-altered networks on the basis of the B cell regulatory network introduced in Supplementary Figure 1. Specifically, we keep all associated REs and member genes of the B cell network, remove the actual connections between TFs and TGs, and then create fake edges by randomly connecting TFs to TGs with random edge weights. We ensure that the edge-altered network has the same number of edges as the actual B cell network and their edge weights have the same distribution. For each negative dataset, we simulate a new edge-altered network independently.

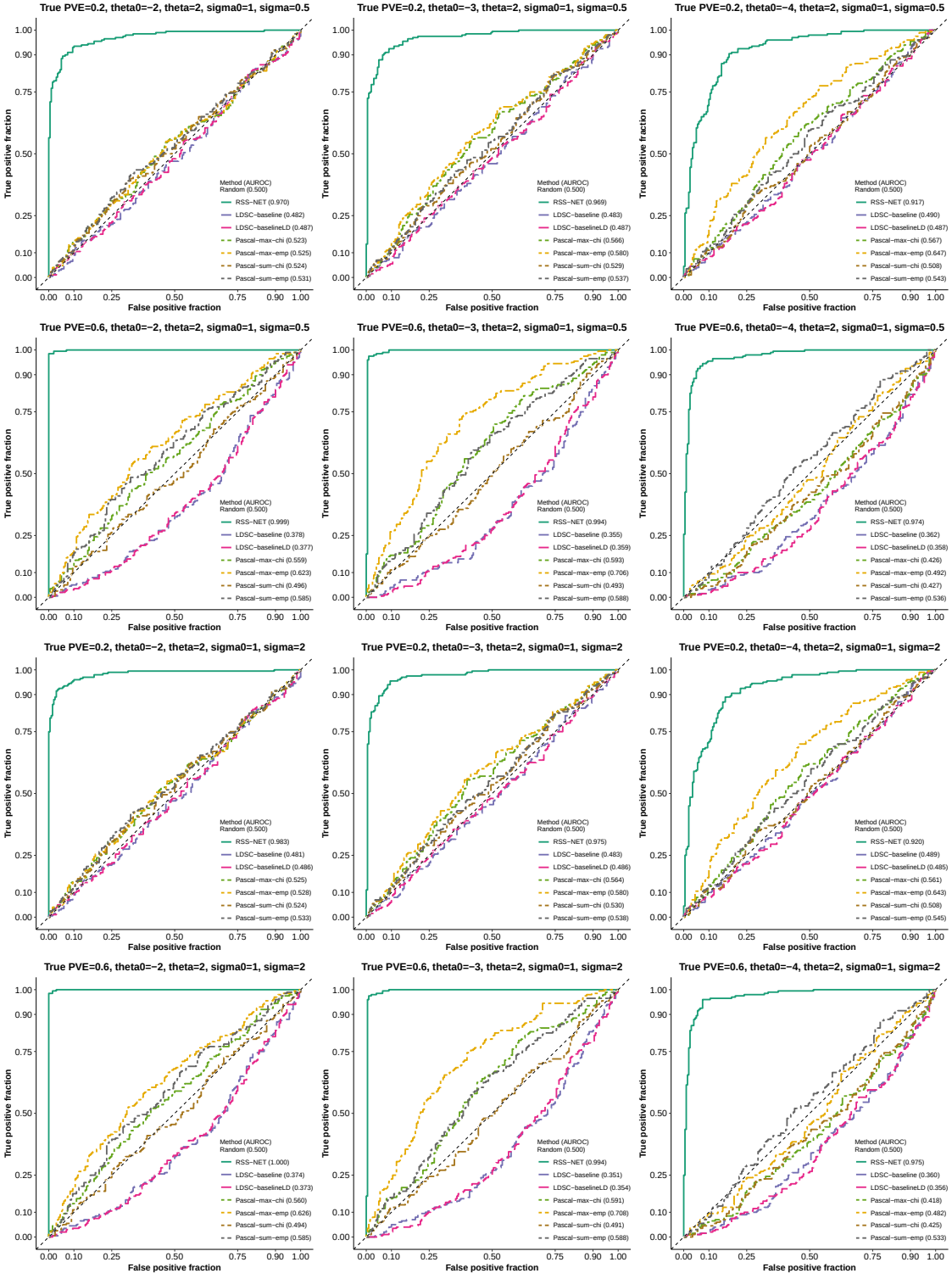
We first simulate positive datasets as in Supplementary Figure 1. Since the true and edge-altered networks have the same associated REs and member genes, their SNP-level annotations $\{a_j\}$ are the same, and thus we let a negative dataset and its corresponding positive dataset have the same causal indicators $\{\zeta_j\}$. We then simulate true genetic effects $\boldsymbol{\beta}$ in a negative dataset as:

$$\begin{aligned} \beta_j \mid \zeta_j = 0 &\sim \delta_0, \\ \beta_j \mid \zeta_j = 1 &\sim \mathcal{N}(0,\ \sigma_0^2 + \sigma^2 \cdot \textstyle\sum_{g \in \mathbf{O}_j^\dagger} (w_{jg}^\dagger)^2), \end{aligned}$$

where $\{\mathbf{O}_j^\dagger, w_{jg}^\dagger\}$ are annotations of the random edge-altered network. For all datasets the target of enrichment testing is the actual B cell network.

The caption of each panel in this supplementary figure shows the true hyper-parameter values of positive datasets. The sparse scenario in Figure 3d corresponds to simulations with true hyper-parameter values of positive datasets being $\theta_0 = -4$, $\theta = 2$, $\sigma_0 = 1$, $\sigma = 0.5$ and PVE = 0.6. The polygenic scenario in Figure 3d corresponds to simulations with true hyper-parameter values of positive datasets being $\theta_0 = -2$, $\theta = 2$, $\sigma_0 = 1$, $\sigma = 0.5$ and PVE = 0.6.

The true and random edge-altered networks have the same associated REs and member genes but totally different topology (TF-TG edges and edge weights). Existing methods like LDSC and Pascal only exploits proximity between SNPs and network genes and/or REs, and thus they cannot distinguish the true and random edge-altered networks. In contrast, because of the enrichment parameter σ^2, RSS-NET is able to capture the topological differences among networks, and thus it can reliably distinguish true enrichments of the B cell network from enrichments of its edge-altered counterparts.


True PVE=0.2, theta0=−2, theta=2, sigma0=1, sigma=0.5
Method (AUROC)
Random (0.500)
RSS−NET (0.970)
LDSC−baseline (0.482)
LDSC−baselineLD (0.487)
Pascal−max−chi (0.523)
Pascal−max−emp (0.525)
Pascal−sum−chi (0.524)
Pascal−sum−emp (0.531)
True PVE=0.2, theta0=−3, theta=2, sigma0=1, sigma=0.5
Method (AUROC)
Random (0.500)
RSS−NET (0.969)
LDSC−baseline (0.483)
LDSC−baselineLD (0.487)
Pascal−max−chi (0.566)
Pascal−max−emp (0.580)
Pascal−sum−chi (0.529)
Pascal−sum−emp (0.537)
True PVE=0.2, theta0=−4, theta=2, sigma0=1, sigma=0.5
Method (AUROC)
Random (0.500)
RSS−NET (0.917)
LDSC−baseline (0.490)
LDSC−baselineLD (0.487)
Pascal−max−chi (0.567)
Pascal−max−emp (0.647)
Pascal−sum−chi (0.508)
Pascal−sum−emp (0.543)
True PVE=0.6, theta0=−2, theta=2, sigma0=1, sigma=0.5
Method (AUROC)
Random (0.500)
RSS−NET (0.999)
LDSC−baseline (0.378)
LDSC−baselineLD (0.377)
Pascal−max−chi (0.559)
Pascal−max−emp (0.623)
Pascal−sum−chi (0.496)
Pascal−sum−emp (0.585)
True PVE=0.6, theta0=−3, theta=2, sigma0=1, sigma=0.5
Method (AUROC)
Random (0.500)
RSS−NET (0.994)
LDSC−baseline (0.355)
LDSC−baselineLD (0.359)
Pascal−max−chi (0.593)
Pascal−max−emp (0.706)
Pascal−sum−chi (0.493)
Pascal−sum−emp (0.588)
True PVE=0.6, theta0=−4, theta=2, sigma0=1, sigma=0.5
Method (AUROC)
Random (0.500)
RSS−NET (0.974)
LDSC−baseline (0.362)
LDSC−baselineLD (0.358)
Pascal−max−chi (0.426)
Pascal−max−emp (0.492)
Pascal−sum−chi (0.427)
Pascal−sum−emp (0.536)
True PVE=0.2, theta0=−2, theta=2, sigma0=1, sigma=2
Method (AUROC)
Random (0.500)
RSS−NET (0.983)
LDSC−baseline (0.481)
LDSC−baselineLD (0.486)
Pascal−max−chi (0.525)
Pascal−max−emp (0.528)
Pascal−sum−chi (0.524)
Pascal−sum−emp (0.533)
True PVE=0.2, theta0=−3, theta=2, sigma0=1, sigma=2
Method (AUROC)
Random (0.500)
RSS−NET (0.975)
LDSC−baseline (0.483)
LDSC−baselineLD (0.486)
Pascal−max−chi (0.564)
Pascal−max−emp (0.580)
Pascal−sum−chi (0.530)
Pascal−sum−emp (0.538)
True PVE=0.2, theta0=−4, theta=2, sigma0=1, sigma=2
Method (AUROC)
Random (0.500)
RSS−NET (0.920)
LDSC−baseline (0.489)
LDSC−baselineLD (0.485)
Pascal−max−chi (0.561)
Pascal−max−emp (0.643)
Pascal−sum−chi (0.508)
Pascal−sum−emp (0.545)
True PVE=0.6, theta0=−2, theta=2, sigma0=1, sigma=2
Method (AUROC)
Random (0.500)
RSS−NET (1.000)
LDSC−baseline (0.374)
LDSC−baselineLD (0.373)
Pascal−max−chi (0.560)
Pascal−max−emp (0.626)
Pascal−sum−chi (0.494)
Pascal−sum−emp (0.585)
True PVE=0.6, theta0=−3, theta=2, sigma0=1, sigma=2
Method (AUROC)
Random (0.500)
RSS−NET (0.994)
LDSC−baseline (0.351)
LDSC−baselineLD (0.354)
Pascal−max−chi (0.591)
Pascal−max−emp (0.708)
Pascal−sum−chi (0.491)
Pascal−sum−emp (0.588)
True PVE=0.6, theta0=−4, theta=2, sigma0=1, sigma=2
Method (AUROC)
Random (0.500)
RSS−NET (0.975)
LDSC−baseline (0.360)
LDSC−baselineLD (0.356)
Pascal−max−chi (0.418)
Pascal−max−emp (0.482)
Pascal−sum−chi (0.425)
Pascal−sum−emp (0.533)
True positive fraction
False positive fraction
0.00
0.10
0.25
0.50
0.75
0.90
1.00

Supplementary Figure 7

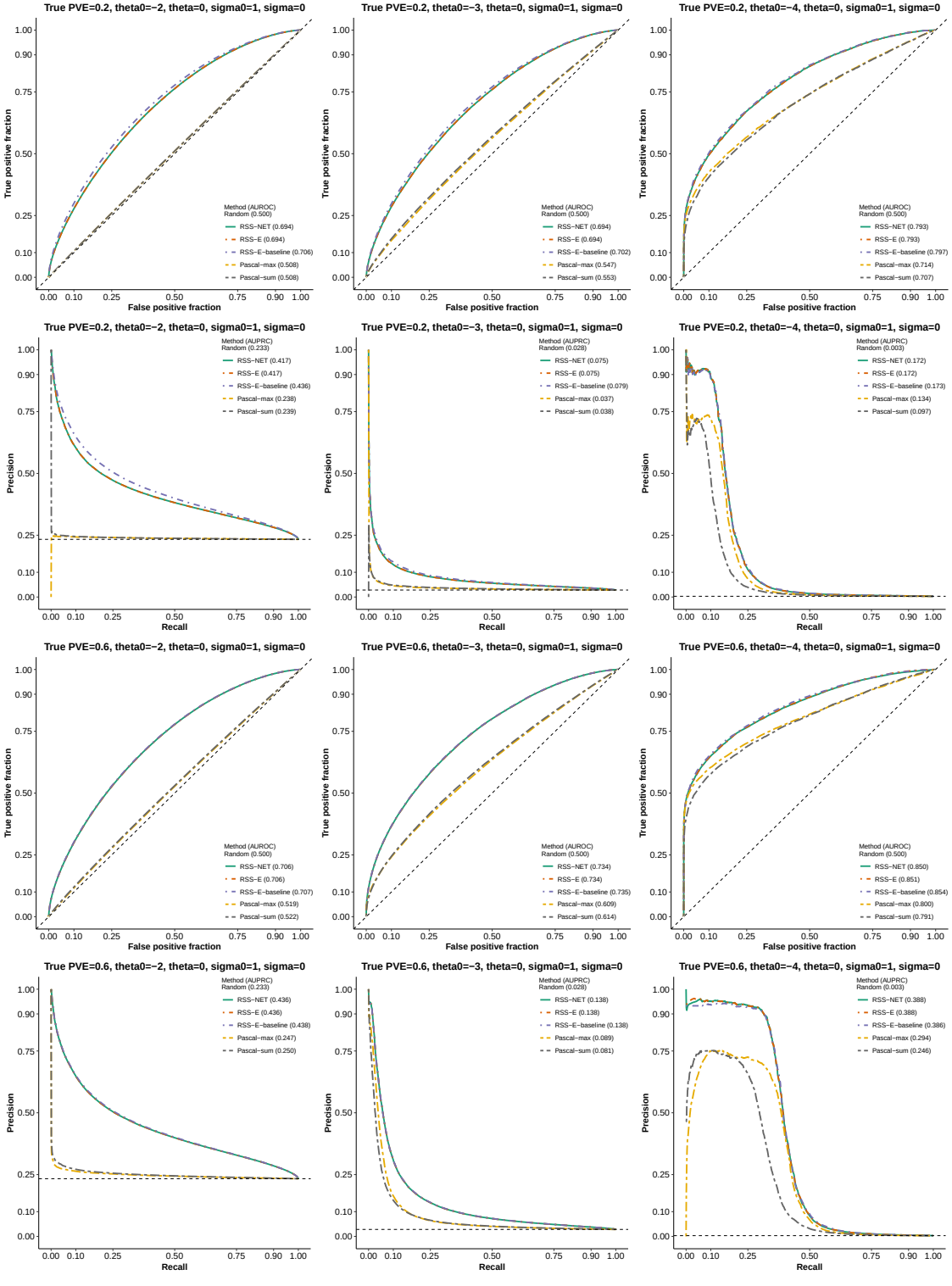
Simulation details and additional results of Figure 4. Here we compare RSS-NET with existing gene-level association testing methods on the same simulated summary data used in Supplementary Figure 1. Details of this part are almost identical to those in Supplementary Figure 1. Here we only highlight the differences.

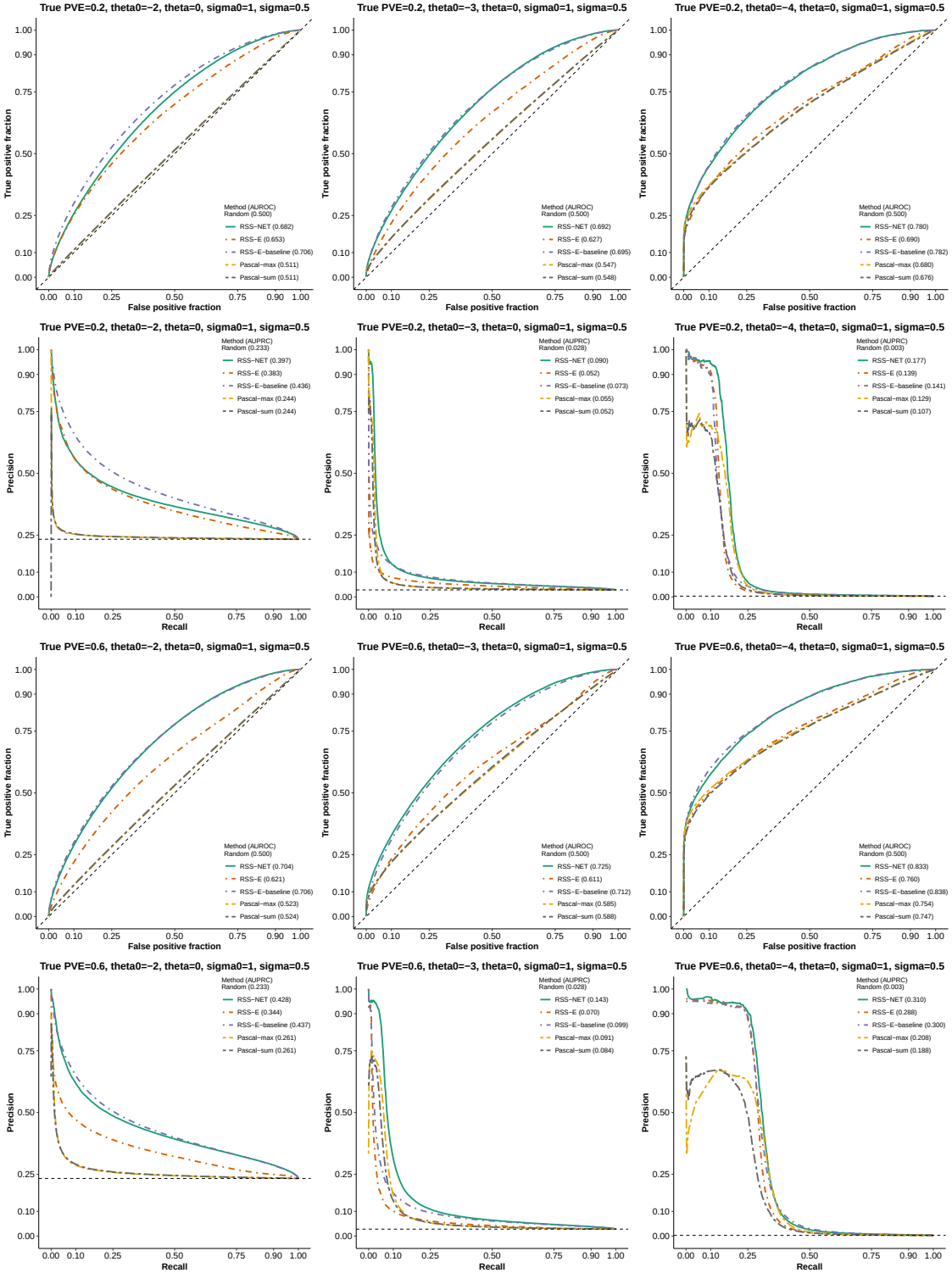
For each simulated dataset, we define a gene as "trait-associated" if at lease one SNP j within 100 kb of the transcribed region of this gene has non-zero effect ($\beta_j \neq 0$). For each gene in each simulated dataset, RSS-NET and RSS-E methods [1] produce P_1, the posterior probability that the gene is trait-associated, whereas Pascal methods [21] produce P-value for the null hypothesis that the gene is not trait-associated; these statistics are used to rank the significance of gene-level associations. If a method identifies association between a non-trait-associated gene and the trait, then it is a "false positive". If a method identifies association between a trait-associated gene and the trait, then it is a "true positive".

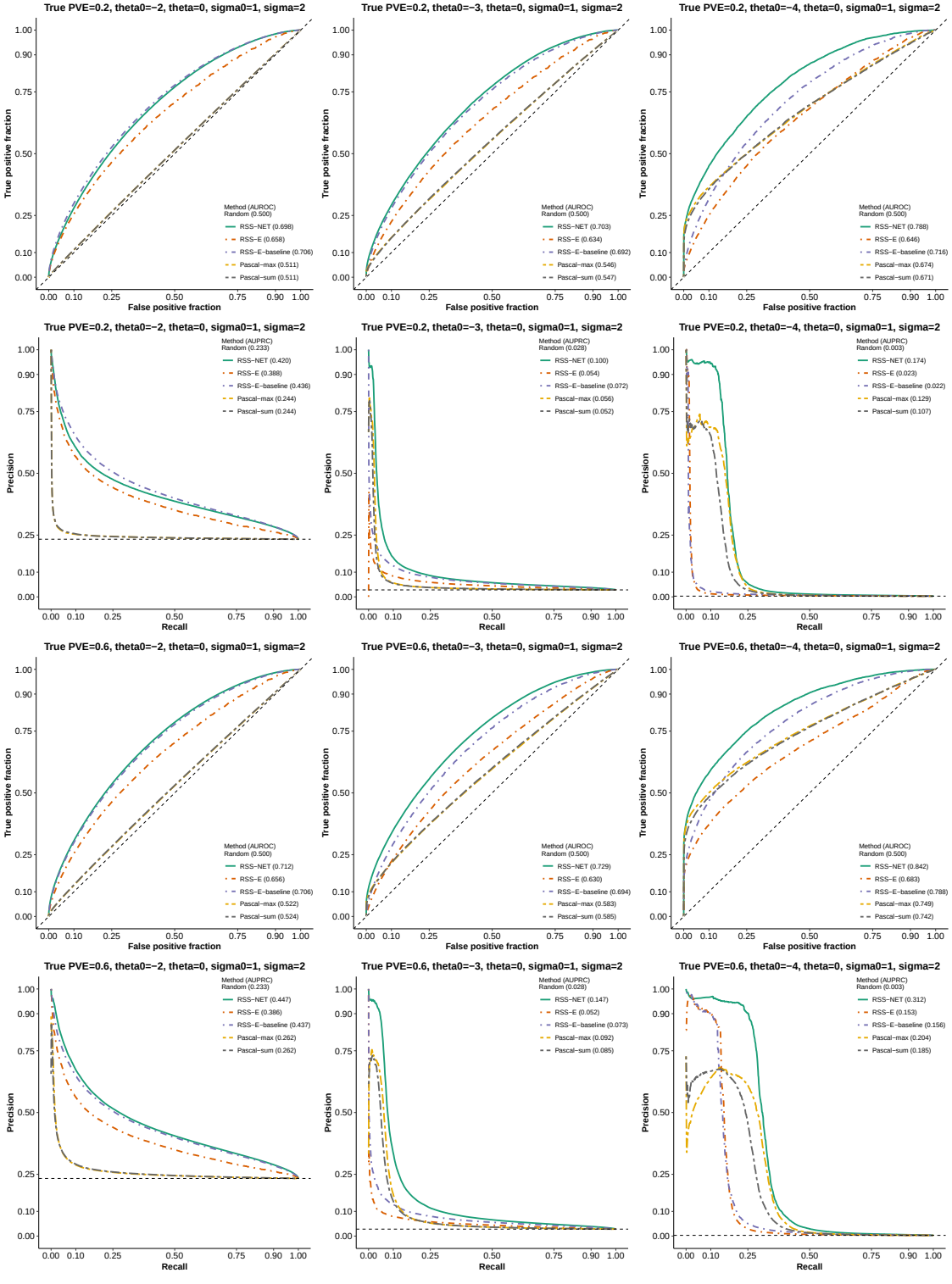
Here the true values of baseline proportion parameter θ_0 are $\{-2,-3,-4\}$, the true values of enrichment proportion parameter θ are $\{0,2\}$, the true value of baseline magnitude parameter σ_0 is 1, the true values of enrichment magnitude parameter σ are $\{0,0.5,2\}$, and the true values of PVE are $\{0.2,0.6\}$. In total there are 36 ($= 3 \times 2 \times 3 \times 2$) simulation scenarios. Each scenario contains 200 independent datasets. For all datasets we test associations of 16,954 autosomal protein-coding genes that are available in Pascal-required database.

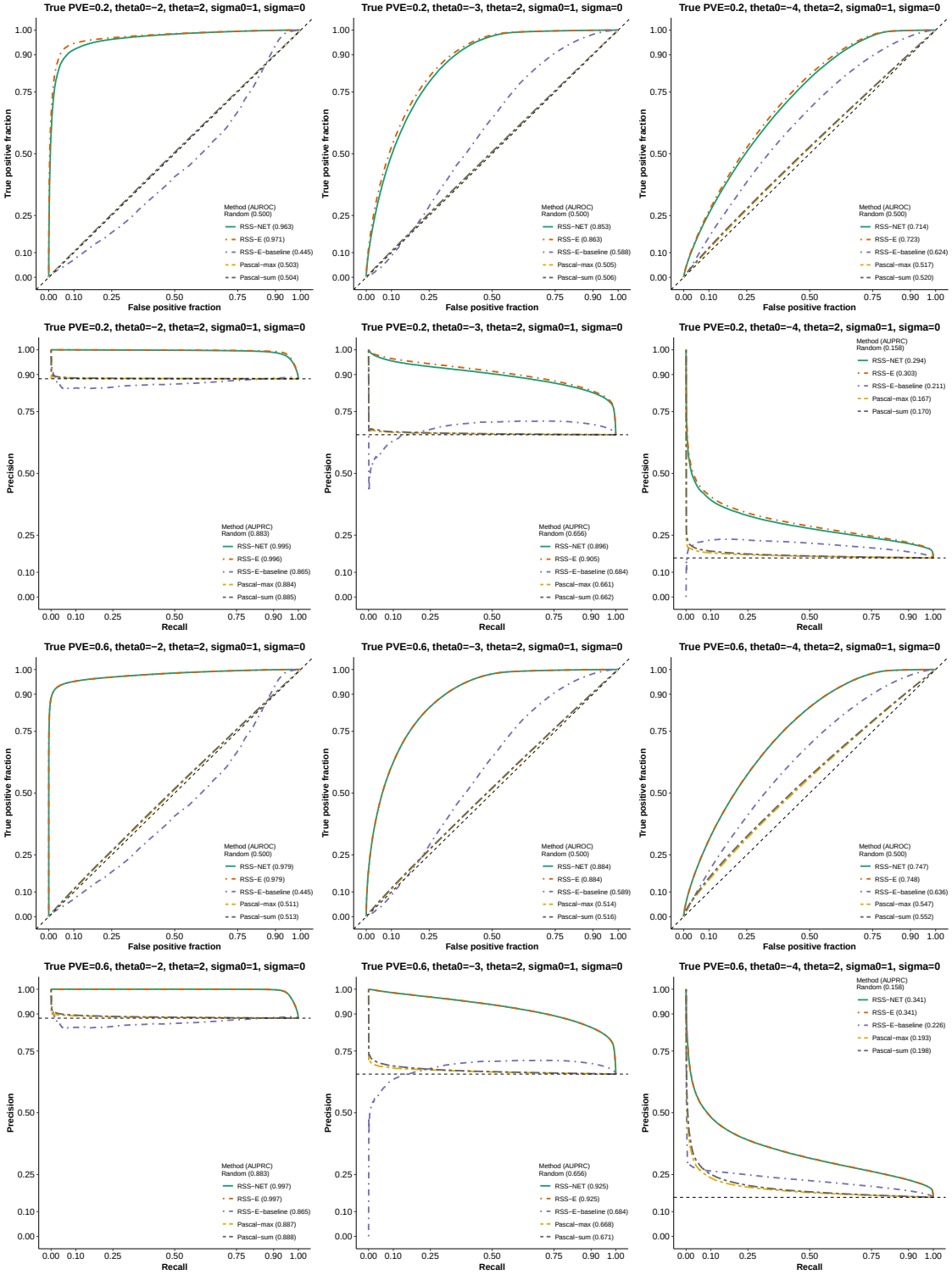
The caption of each panel in this supplementary figure shows the true hyper-parameter values of simulated datasets. All panels of Figure 4 correspond to simulations with true $\theta_0 = -4$, $\sigma_0 = 1$ and PVE = 0.6. Figure 4a corresponds to simulations with true $\theta = 0$ and $\sigma = 0$. Figure 4b corresponds to simulations with true $\theta = 2$ and $\sigma = 0$. Figure 4c corresponds to simulations with true $\theta = 0$ and $\sigma = 2$. Figure 4d corresponds to simulations with true $\theta = 2$ and $\sigma = 2$.

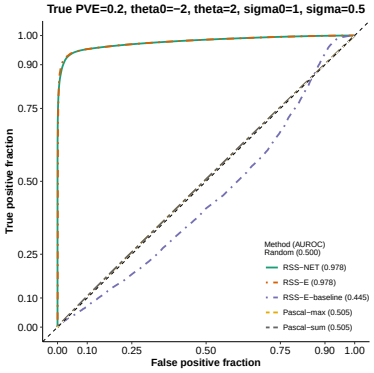
We evaluate the performance of these gene-level association methods by plotting the receiver operating characteristic (ROC) curve and computing the area under the ROC curve (AUROC) for each method. Both metrics are implemented in the R package `plotROC` [19]. Unlike the balanced enrichment simulations (200 positive and 200 negative datasets in each scenario), the numbers of trait-associated genes (true labels) and non-trait-associated genes (false labels) are often very different. Due to this imbalance nature, we also plot the precision-recall (PRC) curve and compute the area under the PRC curve (AUPRC) for each method. Both metrics are implemented in the R package `precrec` [22].

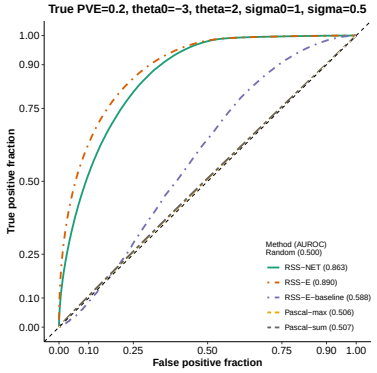

True PVE=0.2, theta0=−2, theta=0, sigma0=1, sigma=0
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.694)
RSS−E (0.694)
RSS−E−baseline (0.706)
Pascal−max (0.508)
Pascal−sum (0.508)
True PVE=0.2, theta0=−3, theta=0, sigma0=1, sigma=0
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.694)
RSS−E (0.694)
RSS−E−baseline (0.702)
Pascal−max (0.547)
Pascal−sum (0.553)
True PVE=0.2, theta0=−4, theta=0, sigma0=1, sigma=0
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.793)
RSS−E (0.793)
RSS−E−baseline (0.797)
Pascal−max (0.714)
Pascal−sum (0.707)
True PVE=0.2, theta0=−2, theta=0, sigma0=1, sigma=0
Precision
Recall
Method (AUPRC)
Random (0.233)
RSS−NET (0.417)
RSS−E (0.417)
RSS−E−baseline (0.436)
Pascal−max (0.238)
Pascal−sum (0.239)
True PVE=0.2, theta0=−3, theta=0, sigma0=1, sigma=0
Precision
Recall
Method (AUPRC)
Random (0.028)
RSS−NET (0.075)
RSS−E (0.075)
RSS−E−baseline (0.079)
Pascal−max (0.037)
Pascal−sum (0.038)
True PVE=0.2, theta0=−4, theta=0, sigma0=1, sigma=0
Precision
Recall
Method (AUPRC)
Random (0.003)
RSS−NET (0.172)
RSS−E (0.172)
RSS−E−baseline (0.173)
Pascal−max (0.134)
Pascal−sum (0.097)
True PVE=0.6, theta0=−2, theta=0, sigma0=1, sigma=0
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.706)
RSS−E (0.706)
RSS−E−baseline (0.707)
Pascal−max (0.519)
Pascal−sum (0.522)
True PVE=0.6, theta0=−3, theta=0, sigma0=1, sigma=0
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.734)
RSS−E (0.734)
RSS−E−baseline (0.735)
Pascal−max (0.609)
Pascal−sum (0.614)
True PVE=0.6, theta0=−4, theta=0, sigma0=1, sigma=0
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.850)
RSS−E (0.851)
RSS−E−baseline (0.854)
Pascal−max (0.800)
Pascal−sum (0.791)
True PVE=0.6, theta0=−2, theta=0, sigma0=1, sigma=0
Precision
Recall
Method (AUPRC)
Random (0.233)
RSS−NET (0.436)
RSS−E (0.436)
RSS−E−baseline (0.438)
Pascal−max (0.247)
Pascal−sum (0.250)
True PVE=0.6, theta0=−3, theta=0, sigma0=1, sigma=0
Precision
Recall
Method (AUPRC)
Random (0.028)
RSS−NET (0.138)
RSS−E (0.138)
RSS−E−baseline (0.138)
Pascal−max (0.089)
Pascal−sum (0.081)
True PVE=0.6, theta0=−4, theta=0, sigma0=1, sigma=0
Precision
Recall
Method (AUPRC)
Random (0.003)
RSS−NET (0.388)
RSS−E (0.388)
RSS−E−baseline (0.386)
Pascal−max (0.294)
Pascal−sum (0.246)

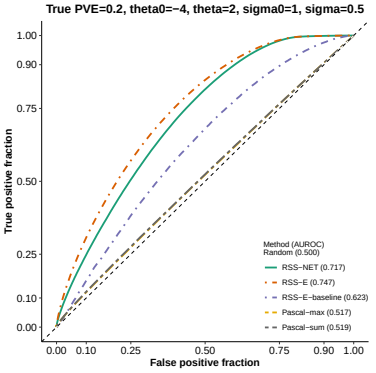

True PVE=0.2, theta0=−2, theta=0, sigma0=1, sigma=0.5
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.682)
RSS−E (0.653)
RSS−E−baseline (0.706)
Pascal−max (0.511)
Pascal−sum (0.511)
True PVE=0.2, theta0=−3, theta=0, sigma0=1, sigma=0.5
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.692)
RSS−E (0.627)
RSS−E−baseline (0.695)
Pascal−max (0.547)
Pascal−sum (0.548)
True PVE=0.2, theta0=−4, theta=0, sigma0=1, sigma=0.5
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.780)
RSS−E (0.690)
RSS−E−baseline (0.782)
Pascal−max (0.680)
Pascal−sum (0.676)
True PVE=0.2, theta0=−2, theta=0, sigma0=1, sigma=0.5
Precision
Recall
Method (AUPRC)
Random (0.233)
RSS−NET (0.397)
RSS−E (0.383)
RSS−E−baseline (0.436)
Pascal−max (0.244)
Pascal−sum (0.244)
True PVE=0.2, theta0=−3, theta=0, sigma0=1, sigma=0.5
Precision
Recall
Method (AUPRC)
Random (0.028)
RSS−NET (0.090)
RSS−E (0.052)
RSS−E−baseline (0.073)
Pascal−max (0.055)
Pascal−sum (0.052)
True PVE=0.2, theta0=−4, theta=0, sigma0=1, sigma=0.5
Precision
Recall
Method (AUPRC)
Random (0.003)
RSS−NET (0.177)
RSS−E (0.139)
RSS−E−baseline (0.141)
Pascal−max (0.129)
Pascal−sum (0.107)
True PVE=0.6, theta0=−2, theta=0, sigma0=1, sigma=0.5
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.704)
RSS−E (0.621)
RSS−E−baseline (0.706)
Pascal−max (0.523)
Pascal−sum (0.524)
True PVE=0.6, theta0=−3, theta=0, sigma0=1, sigma=0.5
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.725)
RSS−E (0.611)
RSS−E−baseline (0.712)
Pascal−max (0.585)
Pascal−sum (0.588)
True PVE=0.6, theta0=−4, theta=0, sigma0=1, sigma=0.5
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.833)
RSS−E (0.760)
RSS−E−baseline (0.838)
Pascal−max (0.754)
Pascal−sum (0.747)
True PVE=0.6, theta0=−2, theta=0, sigma0=1, sigma=0.5
Precision
Recall
Method (AUPRC)
Random (0.233)
RSS−NET (0.428)
RSS−E (0.344)
RSS−E−baseline (0.437)
Pascal−max (0.261)
Pascal−sum (0.261)
True PVE=0.6, theta0=−3, theta=0, sigma0=1, sigma=0.5
Precision
Recall
Method (AUPRC)
Random (0.028)
RSS−NET (0.143)
RSS−E (0.070)
RSS−E−baseline (0.099)
Pascal−max (0.091)
Pascal−sum (0.084)
True PVE=0.6, theta0=−4, theta=0, sigma0=1, sigma=0.5
Precision
Recall
Method (AUPRC)
Random (0.003)
RSS−NET (0.310)
RSS−E (0.288)
RSS−E−baseline (0.300)
Pascal−max (0.208)
Pascal−sum (0.188)

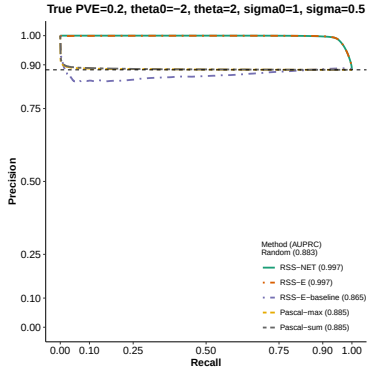

True PVE=0.2, theta0=−2, theta=0, sigma0=1, sigma=2
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.698)
RSS−E (0.658)
RSS−E−baseline (0.706)
Pascal−max (0.511)
Pascal−sum (0.511)
True PVE=0.2, theta0=−3, theta=0, sigma0=1, sigma=2
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.703)
RSS−E (0.634)
RSS−E−baseline (0.692)
Pascal−max (0.546)
Pascal−sum (0.547)
True PVE=0.2, theta0=−4, theta=0, sigma0=1, sigma=2
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.788)
RSS−E (0.646)
RSS−E−baseline (0.716)
Pascal−max (0.674)
Pascal−sum (0.671)
True PVE=0.2, theta0=−2, theta=0, sigma0=1, sigma=2
Precision
Recall
Method (AUPRC)
Random (0.233)
RSS−NET (0.420)
RSS−E (0.388)
RSS−E−baseline (0.436)
Pascal−max (0.244)
Pascal−sum (0.244)
True PVE=0.2, theta0=−3, theta=0, sigma0=1, sigma=2
Precision
Recall
Method (AUPRC)
Random (0.028)
RSS−NET (0.100)
RSS−E (0.054)
RSS−E−baseline (0.072)
Pascal−max (0.056)
Pascal−sum (0.052)
True PVE=0.2, theta0=−4, theta=0, sigma0=1, sigma=2
Precision
Recall
Method (AUPRC)
Random (0.003)
RSS−NET (0.174)
RSS−E (0.023)
RSS−E−baseline (0.022)
Pascal−max (0.129)
Pascal−sum (0.107)
True PVE=0.6, theta0=−2, theta=0, sigma0=1, sigma=2
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.712)
RSS−E (0.656)
RSS−E−baseline (0.706)
Pascal−max (0.522)
Pascal−sum (0.524)
True PVE=0.6, theta0=−3, theta=0, sigma0=1, sigma=2
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.729)
RSS−E (0.630)
RSS−E−baseline (0.694)
Pascal−max (0.583)
Pascal−sum (0.585)
True PVE=0.6, theta0=−4, theta=0, sigma0=1, sigma=2
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.842)
RSS−E (0.683)
RSS−E−baseline (0.788)
Pascal−max (0.749)
Pascal−sum (0.742)
True PVE=0.6, theta0=−2, theta=0, sigma0=1, sigma=2
Precision
Recall
Method (AUPRC)
Random (0.233)
RSS−NET (0.447)
RSS−E (0.386)
RSS−E−baseline (0.437)
Pascal−max (0.262)
Pascal−sum (0.262)
True PVE=0.6, theta0=−3, theta=0, sigma0=1, sigma=2
Precision
Recall
Method (AUPRC)
Random (0.028)
RSS−NET (0.147)
RSS−E (0.052)
RSS−E−baseline (0.073)
Pascal−max (0.092)
Pascal−sum (0.085)
True PVE=0.6, theta0=−4, theta=0, sigma0=1, sigma=2
Precision
Recall
Method (AUPRC)
Random (0.003)
RSS−NET (0.312)
RSS−E (0.153)
RSS−E−baseline (0.156)
Pascal−max (0.204)
Pascal−sum (0.185)

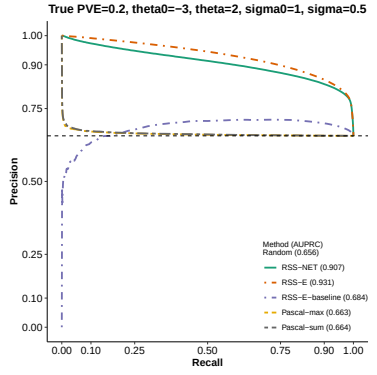

True PVE=0.2, theta0=−2, theta=2, sigma0=1, sigma=0
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.963)
RSS−E (0.971)
RSS−E−baseline (0.445)
Pascal−max (0.503)
Pascal−sum (0.504)
True PVE=0.2, theta0=−3, theta=2, sigma0=1, sigma=0
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.853)
RSS−E (0.863)
RSS−E−baseline (0.588)
Pascal−max (0.505)
Pascal−sum (0.506)
True PVE=0.2, theta0=−4, theta=2, sigma0=1, sigma=0
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.714)
RSS−E (0.723)
RSS−E−baseline (0.624)
Pascal−max (0.517)
Pascal−sum (0.520)
True PVE=0.2, theta0=−2, theta=2, sigma0=1, sigma=0
Precision
Recall
Method (AUPRC)
Random (0.883)
RSS−NET (0.995)
RSS−E (0.996)
RSS−E−baseline (0.865)
Pascal−max (0.884)
Pascal−sum (0.885)
True PVE=0.2, theta0=−3, theta=2, sigma0=1, sigma=0
Precision
Recall
Method (AUPRC)
Random (0.656)
RSS−NET (0.896)
RSS−E (0.905)
RSS−E−baseline (0.684)
Pascal−max (0.661)
Pascal−sum (0.662)
True PVE=0.2, theta0=−4, theta=2, sigma0=1, sigma=0
Precision
Recall
Method (AUPRC)
Random (0.158)
RSS−NET (0.294)
RSS−E (0.303)
RSS−E−baseline (0.211)
Pascal−max (0.167)
Pascal−sum (0.170)
True PVE=0.6, theta0=−2, theta=2, sigma0=1, sigma=0
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.979)
RSS−E (0.979)
RSS−E−baseline (0.445)
Pascal−max (0.511)
Pascal−sum (0.513)
True PVE=0.6, theta0=−3, theta=2, sigma0=1, sigma=0
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.884)
RSS−E (0.884)
RSS−E−baseline (0.589)
Pascal−max (0.514)
Pascal−sum (0.516)
True PVE=0.6, theta0=−4, theta=2, sigma0=1, sigma=0
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.747)
RSS−E (0.748)
RSS−E−baseline (0.636)
Pascal−max (0.547)
Pascal−sum (0.552)
True PVE=0.6, theta0=−2, theta=2, sigma0=1, sigma=0
Precision
Recall
Method (AUPRC)
Random (0.883)
RSS−NET (0.997)
RSS−E (0.997)
RSS−E−baseline (0.865)
Pascal−max (0.887)
Pascal−sum (0.888)
True PVE=0.6, theta0=−3, theta=2, sigma0=1, sigma=0
Precision
Recall
Method (AUPRC)
Random (0.656)
RSS−NET (0.925)
RSS−E (0.925)
RSS−E−baseline (0.684)
Pascal−max (0.668)
Pascal−sum (0.671)
True PVE=0.6, theta0=−4, theta=2, sigma0=1, sigma=0
Precision
Recall
Method (AUPRC)
Random (0.158)
RSS−NET (0.341)
RSS−E (0.341)
RSS−E−baseline (0.226)
Pascal−max (0.193)
Pascal−sum (0.198)

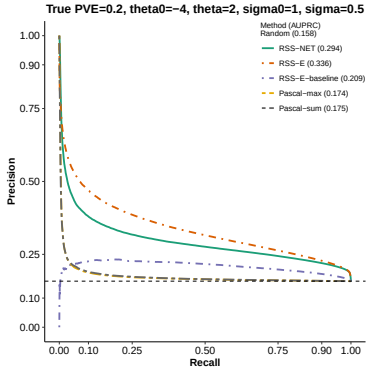

True PVE=0.2, theta0=−2, theta=2, sigma0=1, sigma=0.5
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.978)
RSS−E (0.978)
RSS−E−baseline (0.445)
Pascal−max (0.505)
Pascal−sum (0.505)

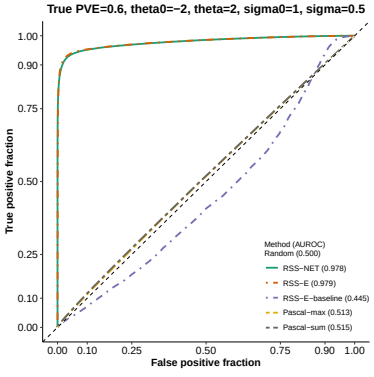

True PVE=0.2, theta0=−3, theta=2, sigma0=1, sigma=0.5
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.863)
RSS−E (0.890)
RSS−E−baseline (0.588)
Pascal−max (0.506)
Pascal−sum (0.507)

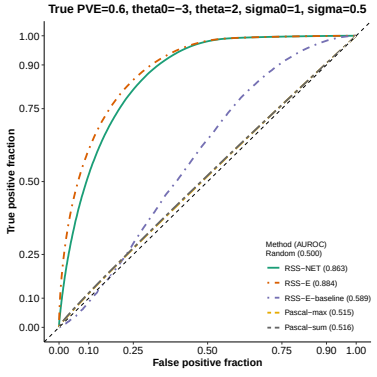

True PVE=0.2, theta0=−4, theta=2, sigma0=1, sigma=0.5
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.717)
RSS−E (0.747)
RSS−E−baseline (0.623)
Pascal−max (0.517)
Pascal−sum (0.519)

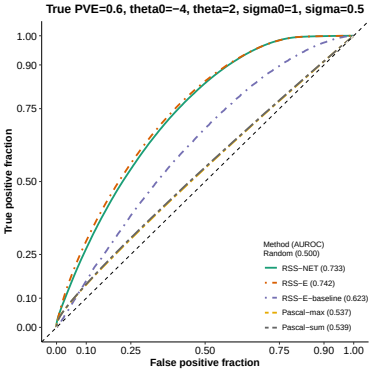

True PVE=0.2, theta0=−2, theta=2, sigma0=1, sigma=0.5
Precision
Recall
Method (AUPRC)
Random (0.883)
RSS−NET (0.997)
RSS−E (0.997)
RSS−E−baseline (0.865)
Pascal−max (0.885)
Pascal−sum (0.885)

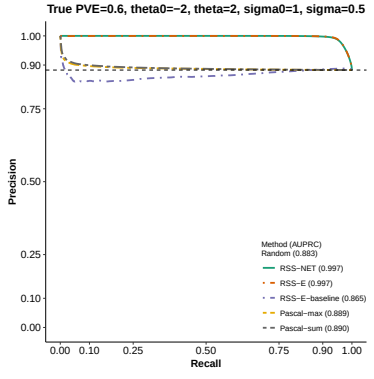

True PVE=0.2, theta0=−3, theta=2, sigma0=1, sigma=0.5
Precision
Recall
Method (AUPRC)
Random (0.656)
RSS−NET (0.907)
RSS−E (0.931)
RSS−E−baseline (0.684)
Pascal−max (0.663)
Pascal−sum (0.664)

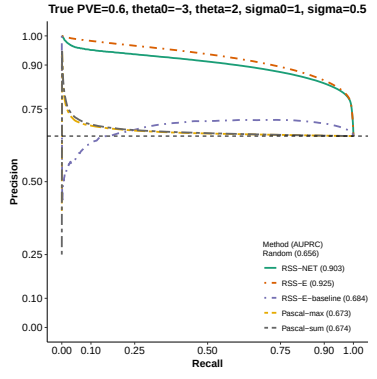

True PVE=0.2, theta0=−4, theta=2, sigma0=1, sigma=0.5
Precision
Recall
Method (AUPRC)
Random (0.158)
RSS−NET (0.294)
RSS−E (0.336)
RSS−E−baseline (0.209)
Pascal−max (0.174)
Pascal−sum (0.175)

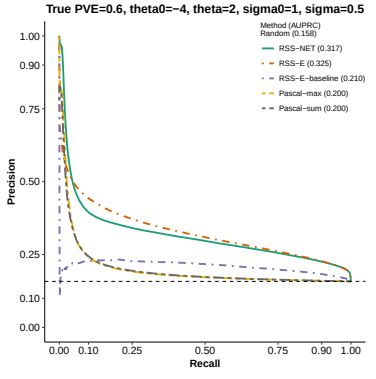

True PVE=0.6, theta0=−2, theta=2, sigma0=1, sigma=0.5
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.978)
RSS−E (0.979)
RSS−E−baseline (0.445)
Pascal−max (0.513)
Pascal−sum (0.515)

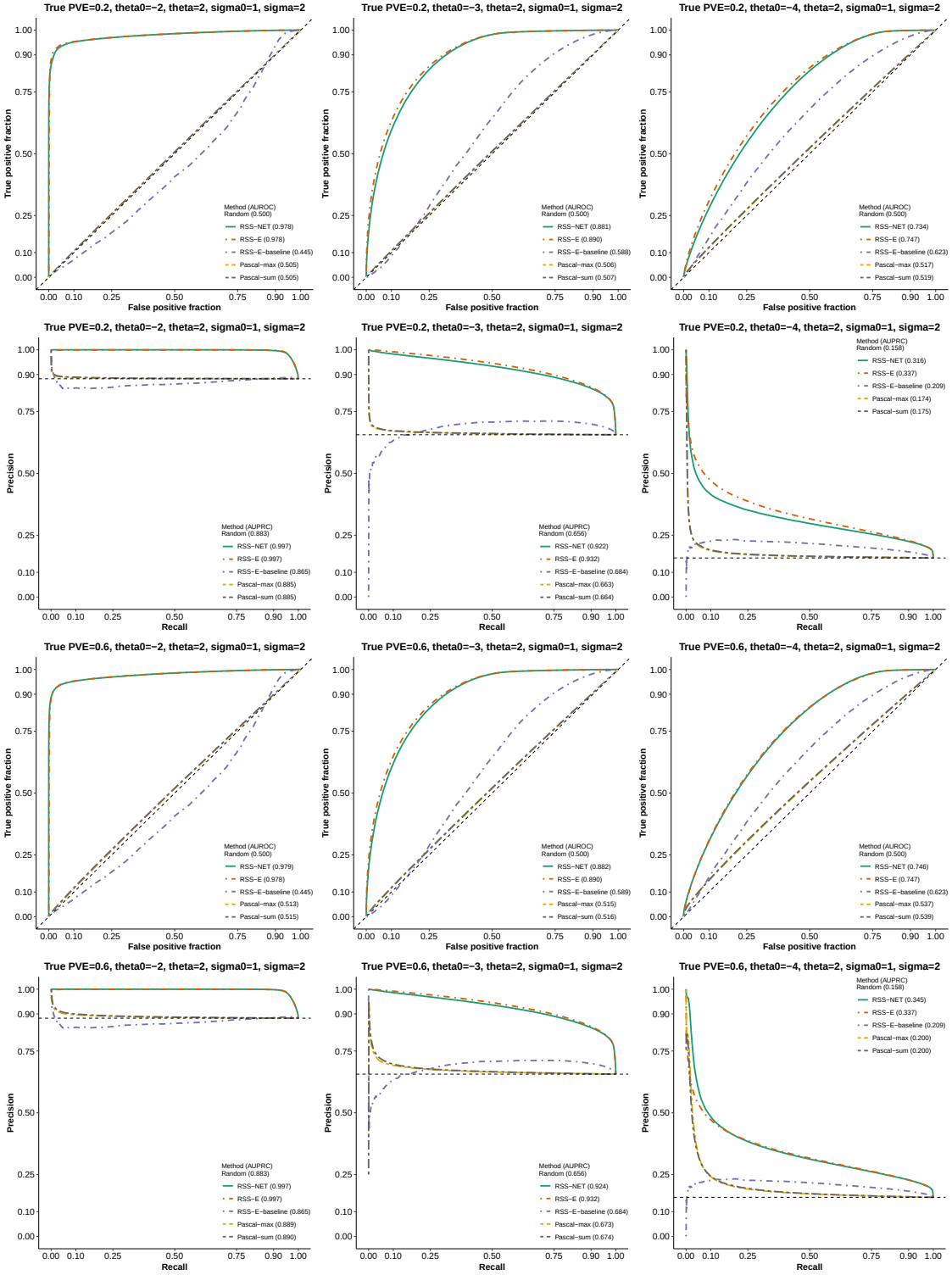

True PVE=0.6, theta0=−3, theta=2, sigma0=1, sigma=0.5
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.863)
RSS−E (0.884)
RSS−E−baseline (0.589)
Pascal−max (0.515)
Pascal−sum (0.516)


True PVE=0.6, theta0=−4, theta=2, sigma0=1, sigma=0.5
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.733)
RSS−E (0.742)
RSS−E−baseline (0.623)
Pascal−max (0.537)
Pascal−sum (0.539)


True PVE=0.6, theta0=−2, theta=2, sigma0=1, sigma=0.5
Precision
Recall
Method (AUPRC)
Random (0.883)
RSS−NET (0.997)
RSS−E (0.997)
RSS−E−baseline (0.865)
Pascal−max (0.889)
Pascal−sum (0.890)

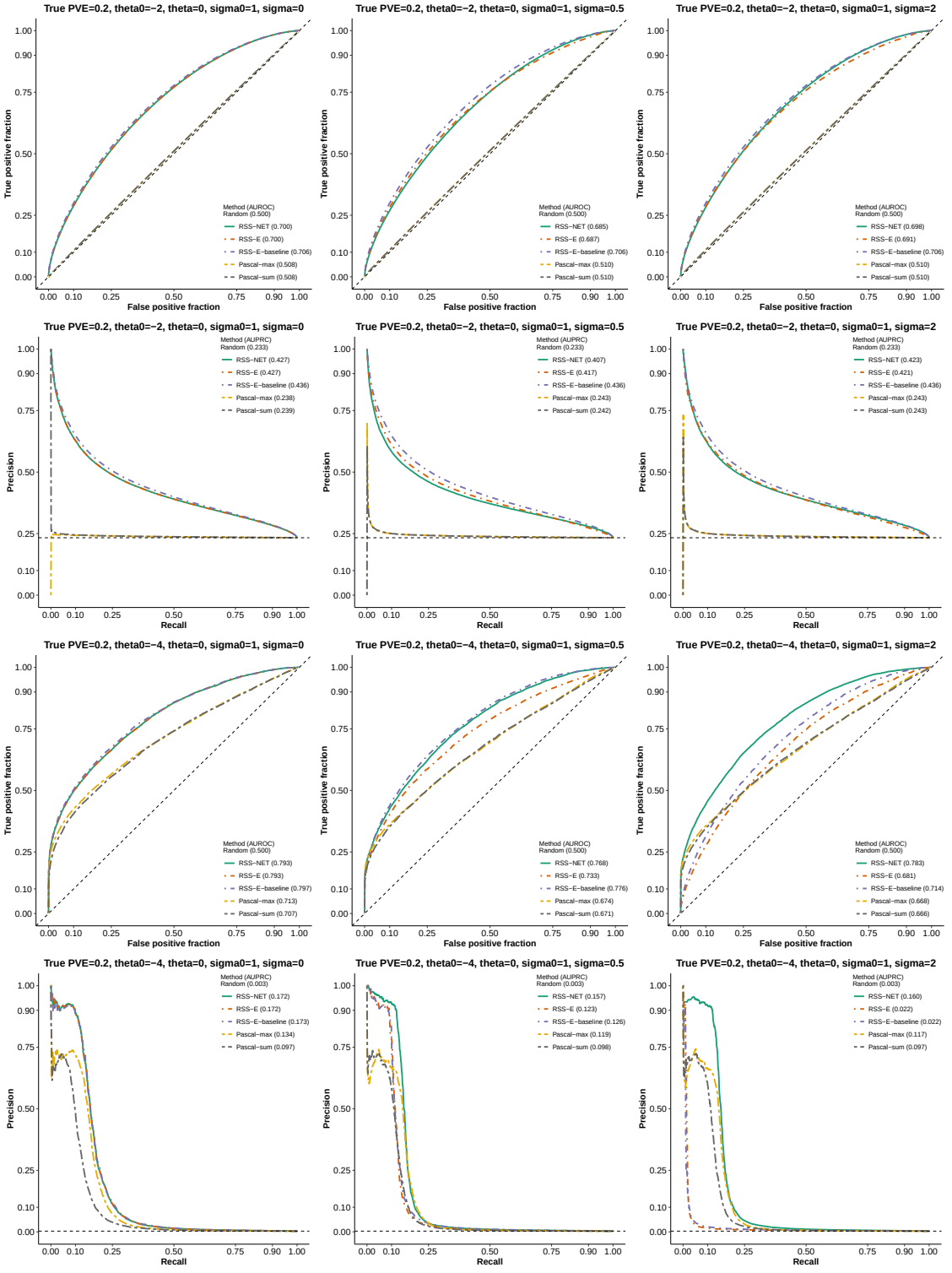

True PVE=0.6, theta0=−3, theta=2, sigma0=1, sigma=0.5
Precision
Recall
Method (AUPRC)
Random (0.656)
RSS−NET (0.903)
RSS−E (0.925)
RSS−E−baseline (0.684)
Pascal−max (0.673)
Pascal−sum (0.674)

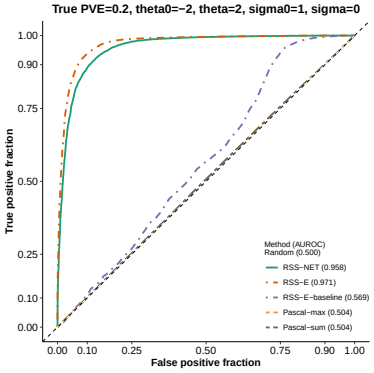

True PVE=0.6, theta0=−4, theta=2, sigma0=1, sigma=0.5
Precision
Recall
Method (AUPRC)
Random (0.158)
RSS−NET (0.317)
RSS−E (0.325)
RSS−E−baseline (0.210)
Pascal−max (0.200)
Pascal−sum (0.200)

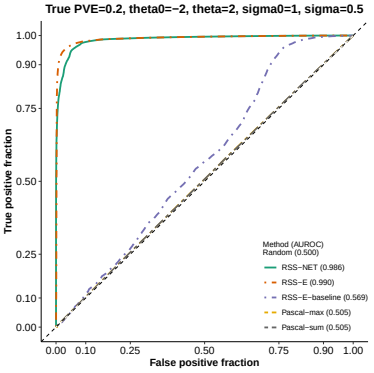

True PVE=0.2, theta0=−2, theta=2, sigma0=1, sigma=2
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.978)
RSS−E (0.978)
RSS−E−baseline (0.445)
Pascal−max (0.505)
Pascal−sum (0.505)
True PVE=0.2, theta0=−3, theta=2, sigma0=1, sigma=2
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.881)
RSS−E (0.890)
RSS−E−baseline (0.588)
Pascal−max (0.506)
Pascal−sum (0.507)
True PVE=0.2, theta0=−4, theta=2, sigma0=1, sigma=2
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.734)
RSS−E (0.747)
RSS−E−baseline (0.623)
Pascal−max (0.517)
Pascal−sum (0.519)
True PVE=0.2, theta0=−2, theta=2, sigma0=1, sigma=2
Precision
Recall
Method (AUPRC)
Random (0.883)
RSS−NET (0.997)
RSS−E (0.997)
RSS−E−baseline (0.865)
Pascal−max (0.885)
Pascal−sum (0.885)
True PVE=0.2, theta0=−3, theta=2, sigma0=1, sigma=2
Precision
Recall
Method (AUPRC)
Random (0.656)
RSS−NET (0.922)
RSS−E (0.932)
RSS−E−baseline (0.684)
Pascal−max (0.663)
Pascal−sum (0.664)
True PVE=0.2, theta0=−4, theta=2, sigma0=1, sigma=2
Precision
Recall
Method (AUPRC)
Random (0.158)
RSS−NET (0.316)
RSS−E (0.337)
RSS−E−baseline (0.209)
Pascal−max (0.174)
Pascal−sum (0.175)
True PVE=0.6, theta0=−2, theta=2, sigma0=1, sigma=2
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.979)
RSS−E (0.978)
RSS−E−baseline (0.445)
Pascal−max (0.513)
Pascal−sum (0.515)
True PVE=0.6, theta0=−3, theta=2, sigma0=1, sigma=2
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.882)
RSS−E (0.890)
RSS−E−baseline (0.589)
Pascal−max (0.515)
Pascal−sum (0.516)
True PVE=0.6, theta0=−4, theta=2, sigma0=1, sigma=2
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.746)
RSS−E (0.747)
RSS−E−baseline (0.623)
Pascal−max (0.537)
Pascal−sum (0.539)
True PVE=0.6, theta0=−2, theta=2, sigma0=1, sigma=2
Precision
Recall
Method (AUPRC)
Random (0.883)
RSS−NET (0.997)
RSS−E (0.997)
RSS−E−baseline (0.865)
Pascal−max (0.889)
Pascal−sum (0.890)
True PVE=0.6, theta0=−3, theta=2, sigma0=1, sigma=2
Precision
Recall
Method (AUPRC)
Random (0.656)
RSS−NET (0.924)
RSS−E (0.932)
RSS−E−baseline (0.684)
Pascal−max (0.673)
Pascal−sum (0.674)
True PVE=0.6, theta0=−4, theta=2, sigma0=1, sigma=2
Precision
Recall
Method (AUPRC)
Random (0.158)
RSS−NET (0.345)
RSS−E (0.337)
RSS−E−baseline (0.209)
Pascal−max (0.200)
Pascal−sum (0.200)

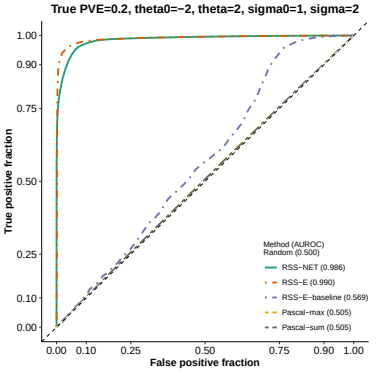
Supplementary Figure 8

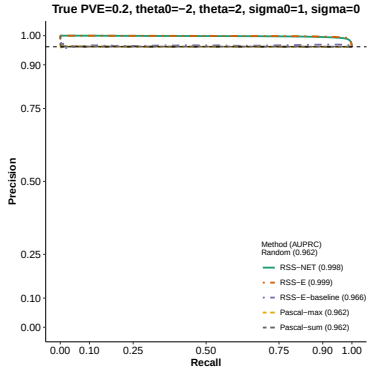
We repeat the simulations in Supplementary Figure 7 on GWAS summary statistics simulated in Supplementary Figure 2 and the vagina regulatory network.

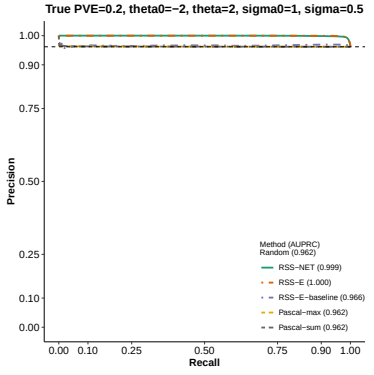


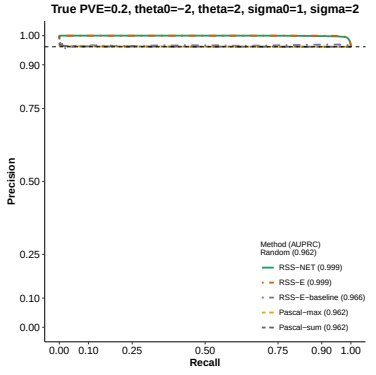

True PVE=0.2, theta0=−2, theta=2, sigma0=1, sigma=0
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.958)
RSS−E (0.971)
RSS−E−baseline (0.569)
Pascal−max (0.504)
Pascal−sum (0.504)

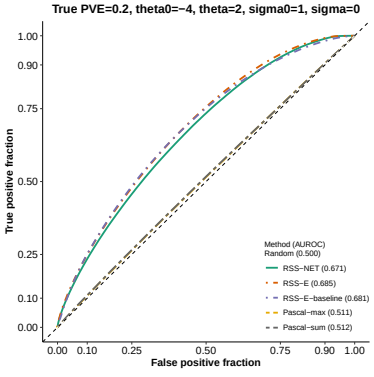

True PVE=0.2, theta0=−2, theta=2, sigma0=1, sigma=0.5
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.986)
RSS−E (0.990)
RSS−E−baseline (0.569)
Pascal−max (0.505)
Pascal−sum (0.505)

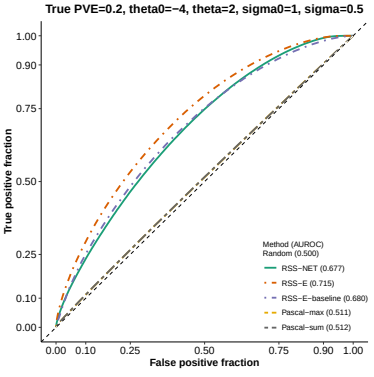

True PVE=0.2, theta0=−2, theta=2, sigma0=1, sigma=2
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.986)
RSS−E (0.990)
RSS−E−baseline (0.569)
Pascal−max (0.505)
Pascal−sum (0.505)

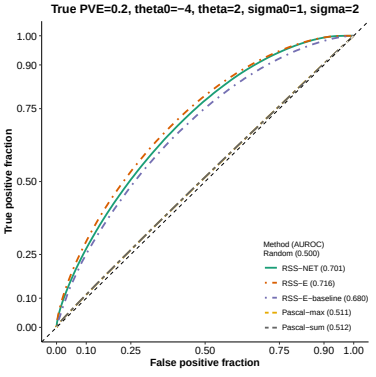

True PVE=0.2, theta0=−2, theta=2, sigma0=1, sigma=0
Precision
Recall
Method (AUPRC)
Random (0.962)
RSS−NET (0.998)
RSS−E (0.999)
RSS−E−baseline (0.966)
Pascal−max (0.962)
Pascal−sum (0.962)

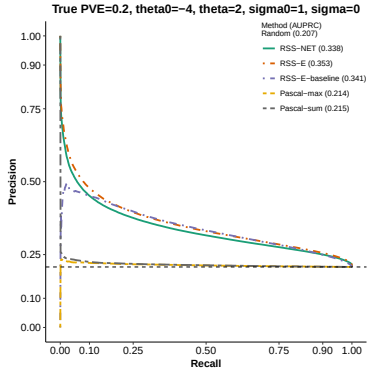

True PVE=0.2, theta0=−2, theta=2, sigma0=1, sigma=0.5
Precision
Recall
Method (AUPRC)
Random (0.962)
RSS−NET (0.999)
RSS−E (1.000)
RSS−E−baseline (0.966)
Pascal−max (0.962)
Pascal−sum (0.962)

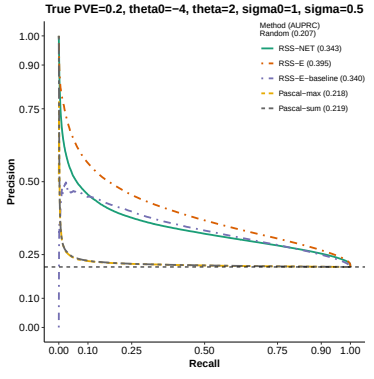

True PVE=0.2, theta0=−2, theta=2, sigma0=1, sigma=2
Precision
Recall
Method (AUPRC)
Random (0.962)
RSS−NET (0.999)
RSS−E (0.999)
RSS−E−baseline (0.966)
Pascal−max (0.962)
Pascal−sum (0.962)

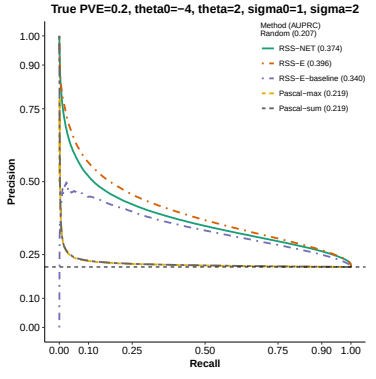

True PVE=0.2, theta0=−4, theta=2, sigma0=1, sigma=0
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.671)
RSS−E (0.685)
RSS−E−baseline (0.681)
Pascal−max (0.511)
Pascal−sum (0.512)


True PVE=0.2, theta0=−4, theta=2, sigma0=1, sigma=0.5
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.677)
RSS−E (0.715)
RSS−E−baseline (0.680)
Pascal−max (0.511)
Pascal−sum (0.512)


True PVE=0.2, theta0=−4, theta=2, sigma0=1, sigma=2
True positive fraction
False positive fraction
Method (AUROC)
Random (0.500)
RSS−NET (0.701)
RSS−E (0.716)
RSS−E−baseline (0.680)
Pascal−max (0.511)
Pascal−sum (0.512)


True PVE=0.2, theta0=−4, theta=2, sigma0=1, sigma=0
Precision
Recall
Method (AUPRC)
Random (0.207)
RSS−NET (0.338)
RSS−E (0.353)
RSS−E−baseline (0.341)
Pascal−max (0.214)
Pascal−sum (0.215)

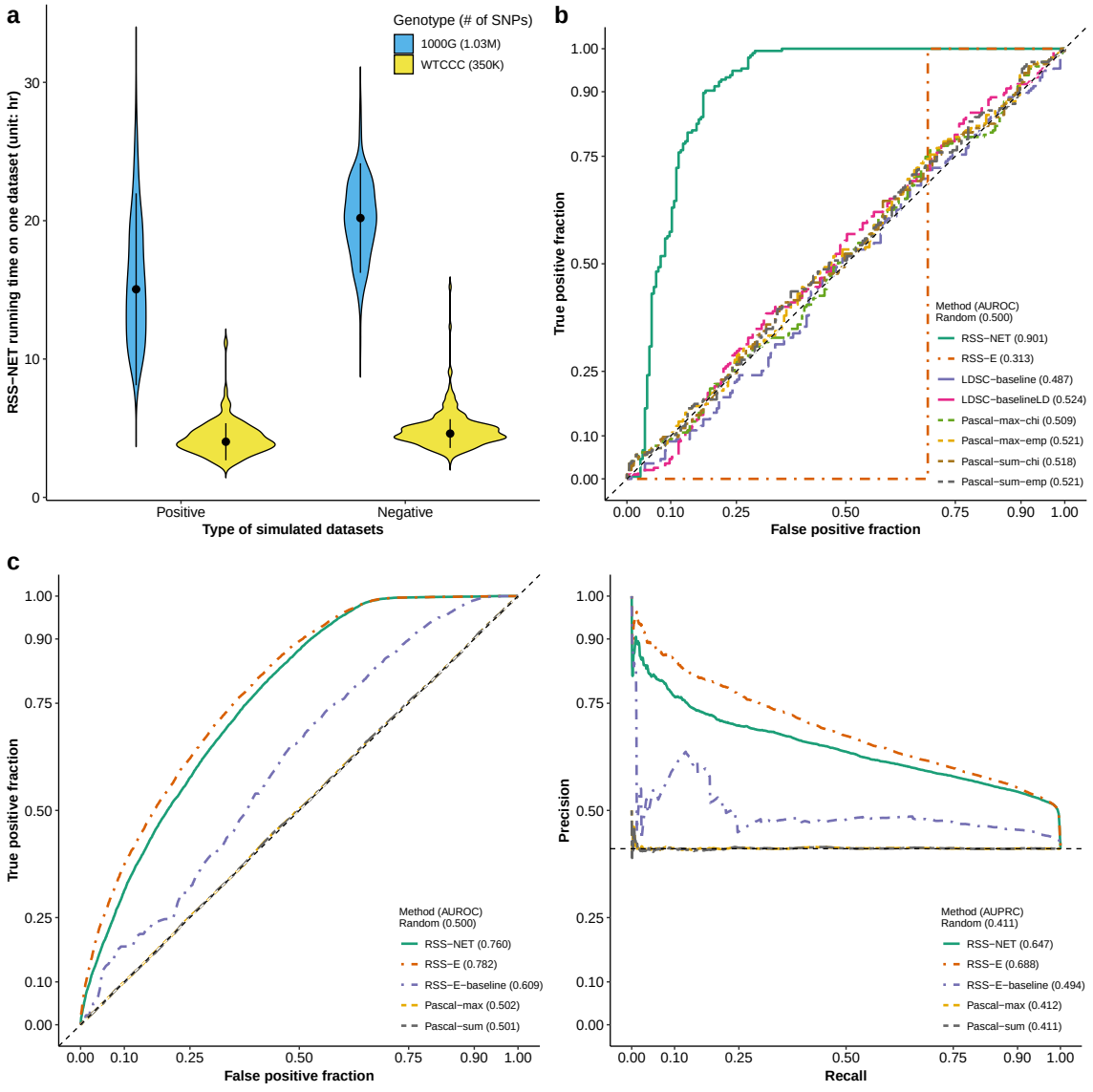

True PVE=0.2, theta0=−4, theta=2, sigma0=1, sigma=0.5
Precision
Recall
Method (AUPRC)
Random (0.207)
RSS−NET (0.343)
RSS−E (0.395)
RSS−E−baseline (0.340)
Pascal−max (0.218)
Pascal−sum (0.219)


True PVE=0.2, theta0=−4, theta=2, sigma0=1, sigma=2
Precision
Recall
Method (AUPRC)
Random (0.207)
RSS−NET (0.374)
RSS−E (0.396)
RSS−E−baseline (0.340)
Pascal−max (0.219)
Pascal−sum (0.219)

Supplementary Figure 9

RSS-NET simulations based on 1 million common SNPs. This part is the same as Supplementary Figure 1, except that here we use the genotype data of 503 European-ancestry individuals on 1,030,397 common SNPs (MAF ≥ 1%) from [23] to specify the genotype matrix **X**. Results below (**a**: timing; **b**: network enrichment; **c**: gene-level association) correspond to the simulation scenario with true hyper-parameters of positive datasets being $\theta_0 = -4$, $\theta = 2$, $\sigma_0 = 1$, $\sigma = 2$ and PVE = 0.6. Error bars in Panel **a** represent interquartile range (from 25th to 75th percentiles), $n = 200$ independent positive or negative datasets.

As shown in Panel **a**, the computation time of RSS-NET increases as the number of genome-wide SNPs analyzed increases. On average, when the number of SNPs increases from 348,965 to 1,030,397, the total computation time per dataset is four times longer (one-sided Wilcoxon $P = 8.02 \times 10^{-132}$). Due to the large-scale simulations conducted in this study, we mainly use 348,965 genome-wide SNPs in order to reduce computation.



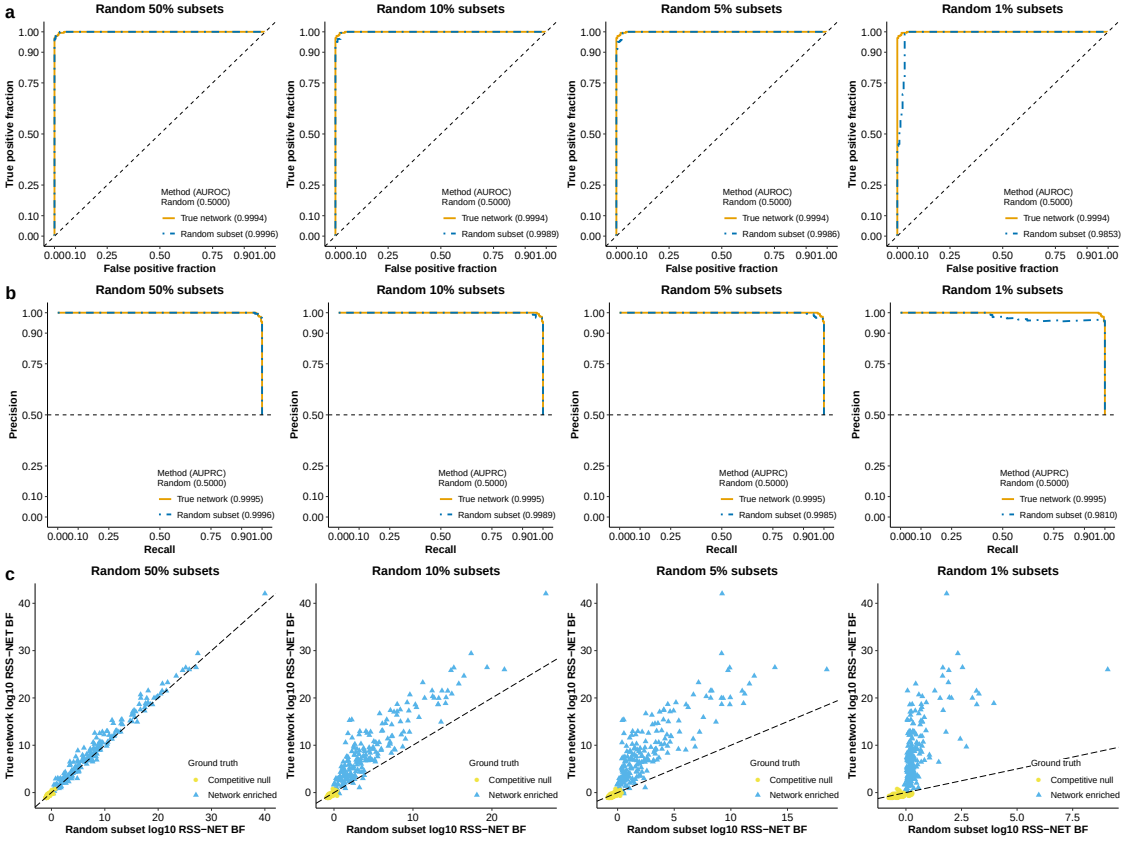
Supplementary Figure 10

Robustness of RSS-NET under noisy networks. This part aims to assess the robustness of RSS-NET to model mis-specification where the GWAS summary data are simulated from a real target network and the enrichment testing is performed on a noisy version of this real target. Details of this part are almost identical to those in Supplementary Figure 1. Here we only highlight the differences.

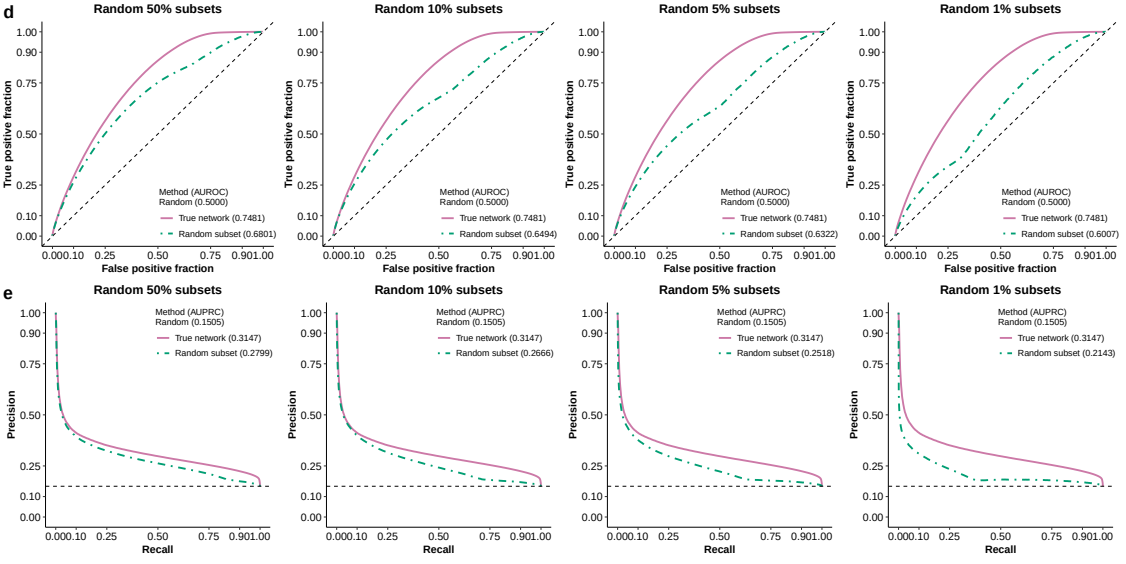
The real target network used to simulate GWAS summary data is the B cell regulatory network in Supplementary Figure 1. We create a noisy network for this target by randomly subsetting a given fraction of its TF-TG edges and the associated REs. Here the fraction of TF-TG edges to subset are $\{1\%, 5\%, 10\%, 50\%\}$. For each pair of negative and positive GWAS datasets, we simulate a new noisy network and use it in RSS-NET.

The GWAS summary data (both negative and positive) used here are the same as those used in Supplementary Figure 1, with true $\theta_0 = -4$, $\theta = 2$, $\sigma_0 = 1$, $\sigma = 2$ and PVE = 0.2. Unlike Supplementary Figure 1, here we do not test the enrichment of B cell network (i.e. the real target network), but test its random subsets instead.

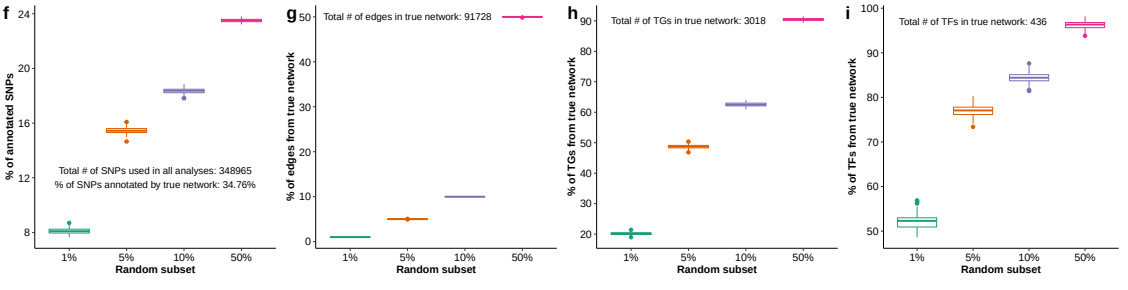
Panels **a**-**c** Comparison of network enrichment results between using the real target network and using its noisy versions in RSS-NET.



Panels **d-e** Comparison of gene-level association results between using the real target network and using its noisy versions in RSS-NET. Note that the AUROC based on true network shown below is slightly different from the AUROC shown in Supplementary Figure 7, although they are generated from the same simulated datasets. This is because AUROC shown in Supplementary Figure 7 is evaluated based on 16,954 autosomal protein-coding genes that are available in Pascal-required database, whereas AUROC shown below is evaluated based on all available autosomal protein-coding genes.



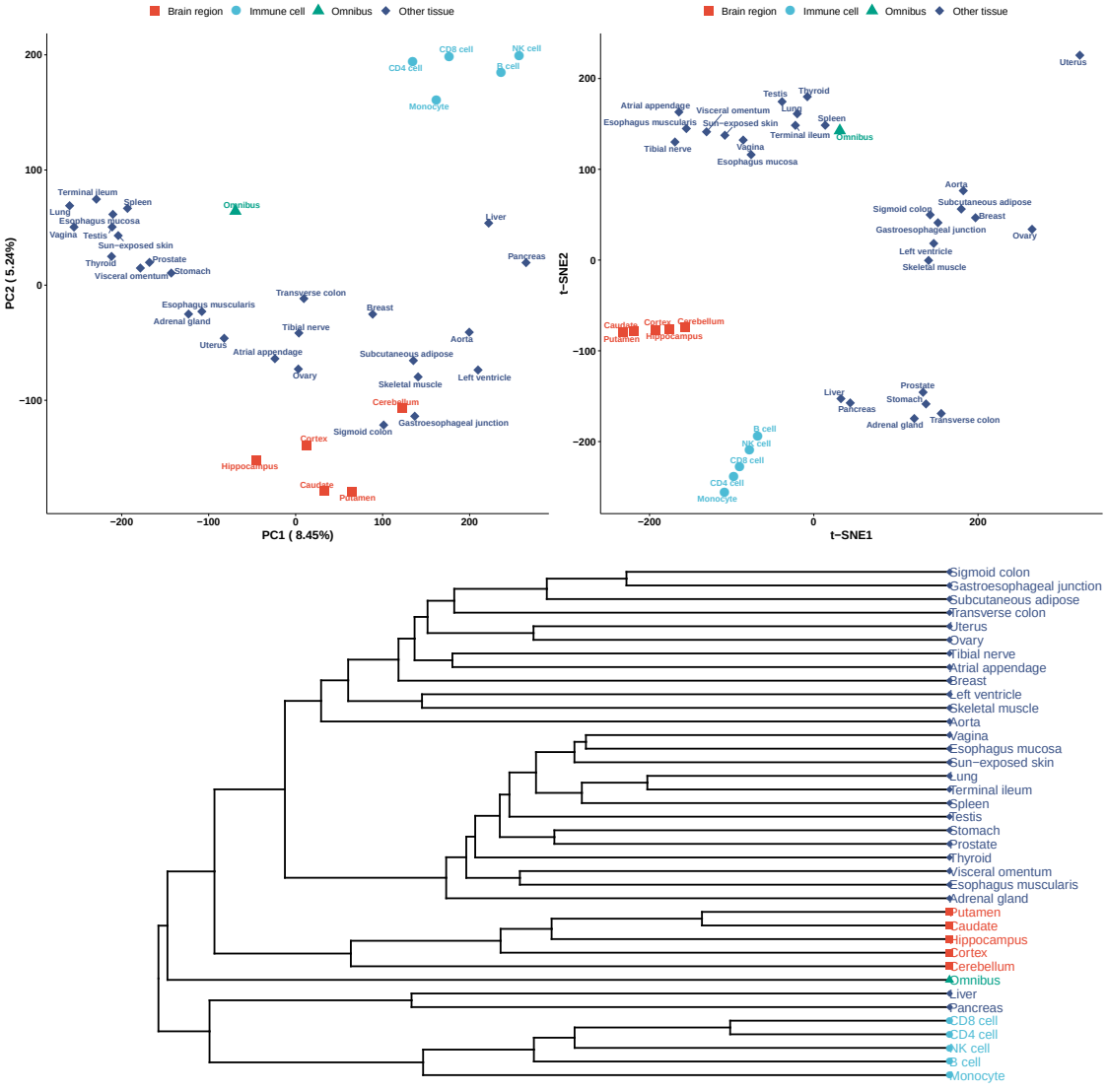
Panels **f-i** Descriptive statistics of noisy networks under different subsetting fractions. Each box plot below is based on 200 independently simulated noisy networks. Box plots indicate median (middle line), 25th, 75th percentile (box) and 5th and 95th percentile (whiskers) as well as outliers (single points).



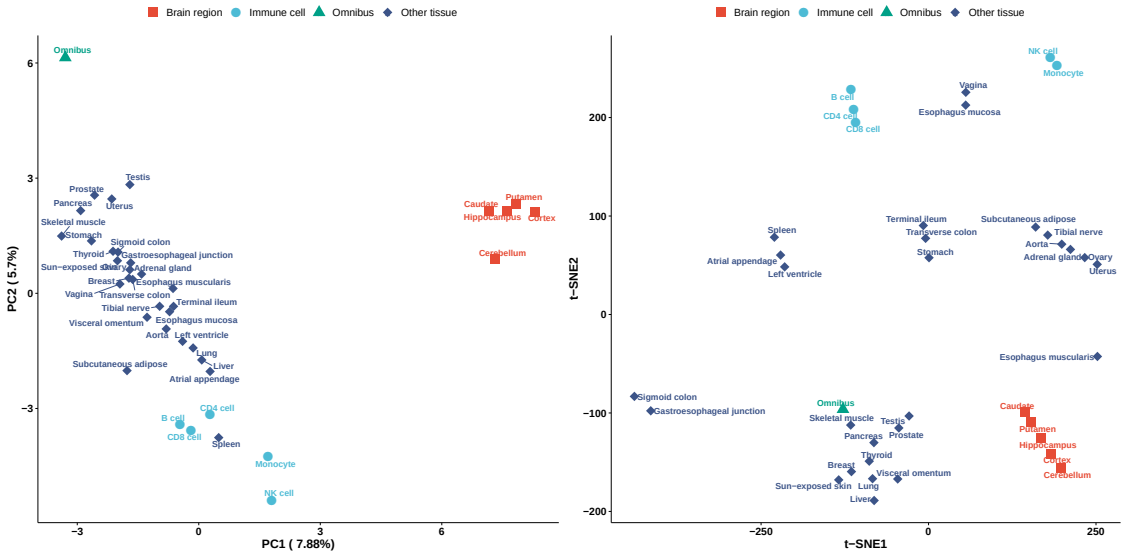
Supplementary Figure 11

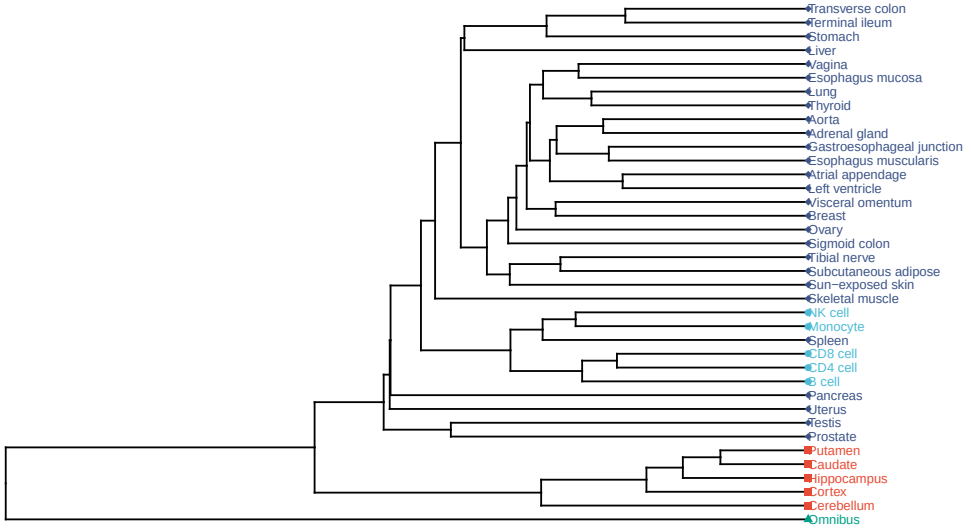
Analysis details and additional results for Figure 5a. Here we perform clustering analysis of 38 regulatory networks based on principal component analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE), and hierarchical clustering. To perform PCA, we use the R built-in function `prcomp`. To perform t-SNE, we use the R package `Rtsne`. To perform hierarchical clustering, we use the R built-in function `hclust`, with distance = 1 – Pearson correlation and the average method. Figure 5a corresponds to the t-SNE plot in Panel **e**.

Panel **a** Here the input data matrix has 1,289,786 rows (SNPs) and 38 columns (networks), where the (i, j)-th entry is 1 if SNP i is within 100 kb of any associated RE or transcribed region of any member gene of network j and 0 otherwise.

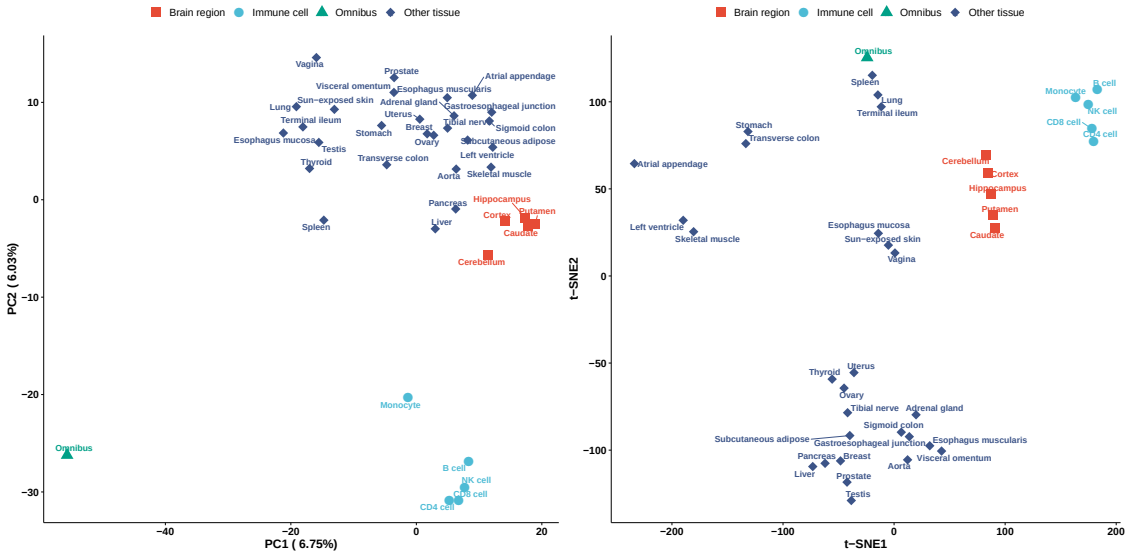


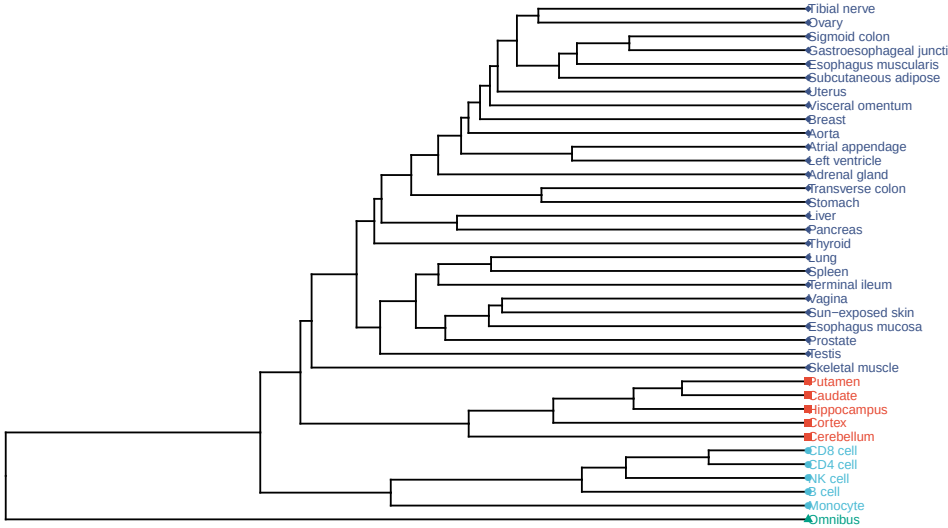
Panel **b** Here the input data matrix has 605 rows (TFs) and 38 columns (networks), where the (i, j)-th entry is 1 if TF i belongs to network j and 0 otherwise.



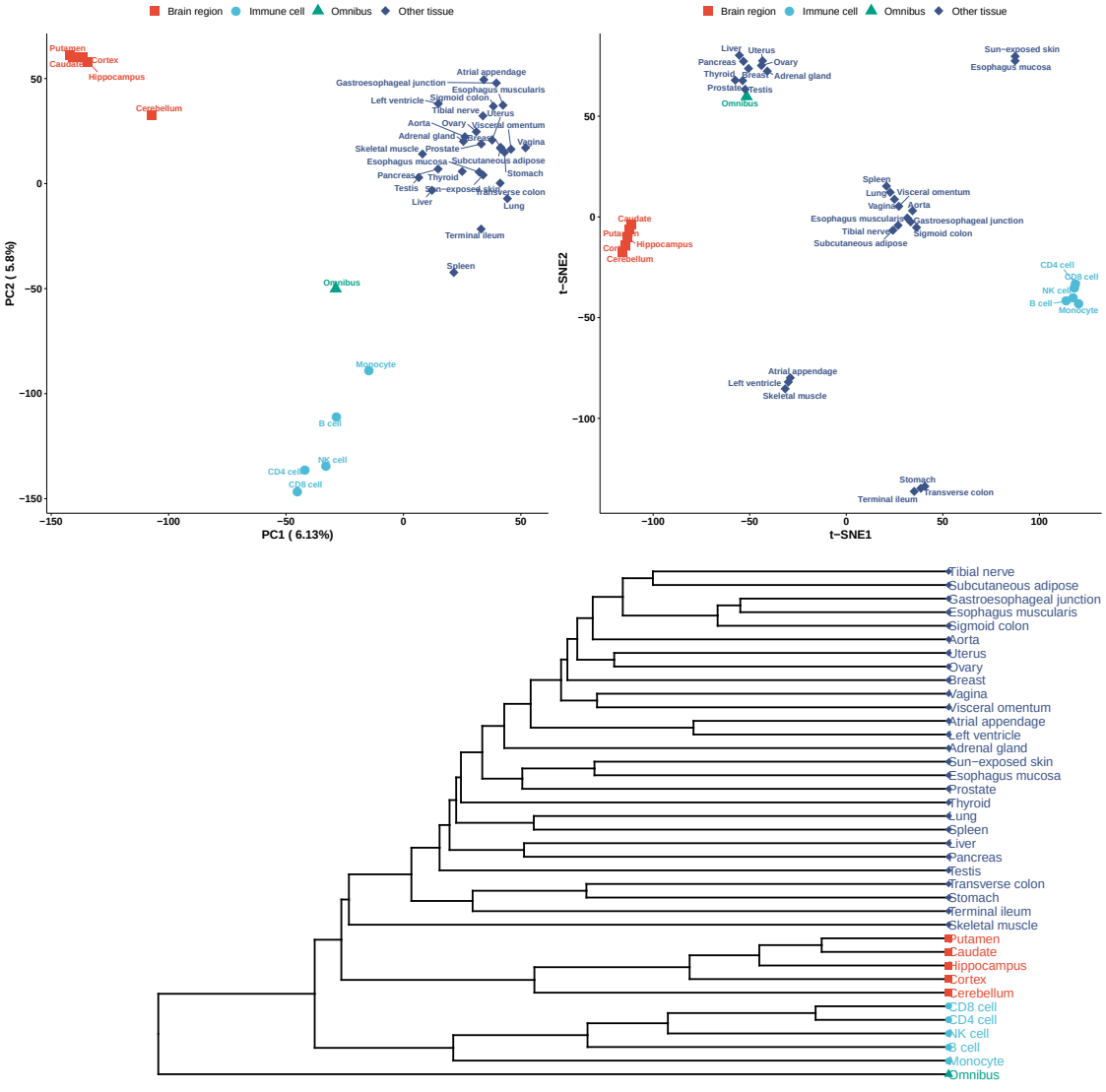


Panel **c** Here the input data matrix has 12,963 rows (TGs) and 38 columns (networks), where the (i, j)-th entry is 1 if TG i belongs to network j and 0 otherwise.

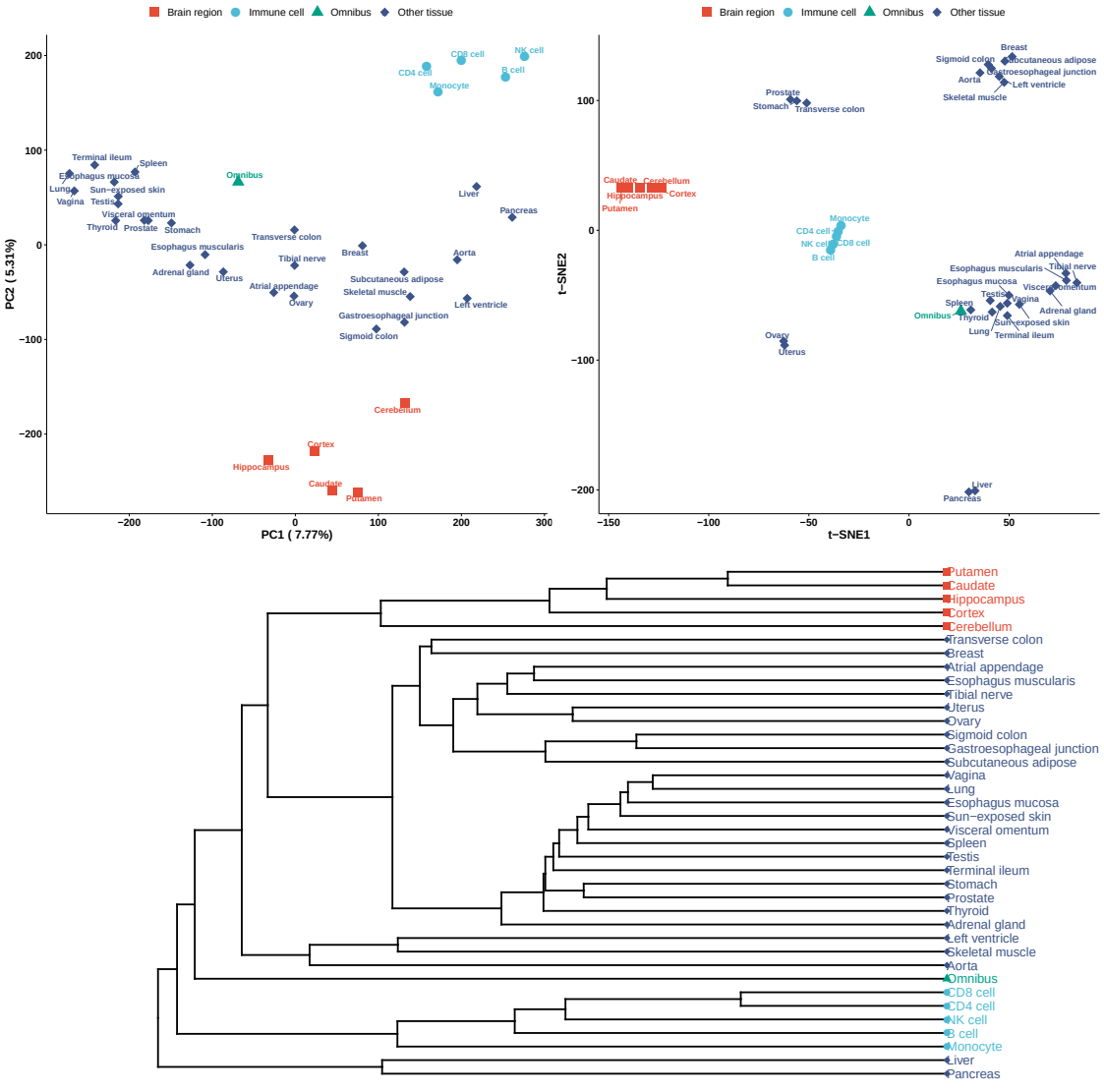




Panel **d** Here the input data matrix has 880,777 rows (TF-TG edges) and 38 columns (networks), where the (i,j)-th entry is the weight of TF-TG edge i in network j, and it is 0 if TF-TG edge i is not available in network j.

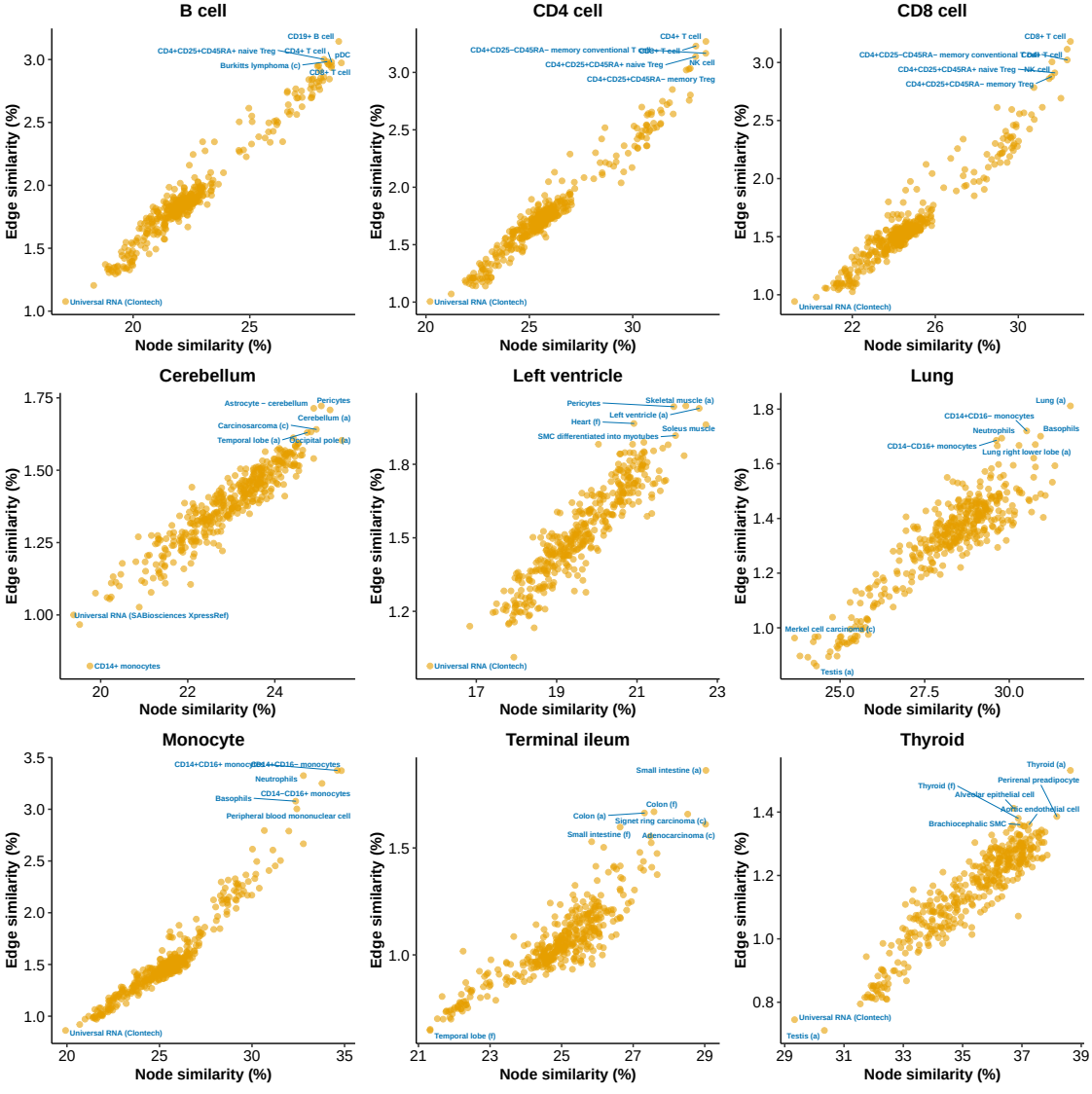


Panel **e** Here the input data matrix is obtained by concatenating the previous 4 data matrices vertically, with a total of 2,184,131 rows and 38 columns (networks).



Supplementary Figure 12

Additional results for Figure 5b. Each panel below shows the node and edge similarities between a given cell type- or tissue-specific PECA-based network (panel caption: cell type or tissue) and CAGE-based networks for 394 cell types and tissues [24] inferred from independent CAGE data [25, 26]. In each panel, each point represents a context-specific CAGE-based network, where *x*- and *y*-axis values indicate its node and edge set similarity with the given PECA network respectively. For both node and edge sets, the similarity is measured by Jaccard index (i.e., the size of the intersection divided by the size of the union).

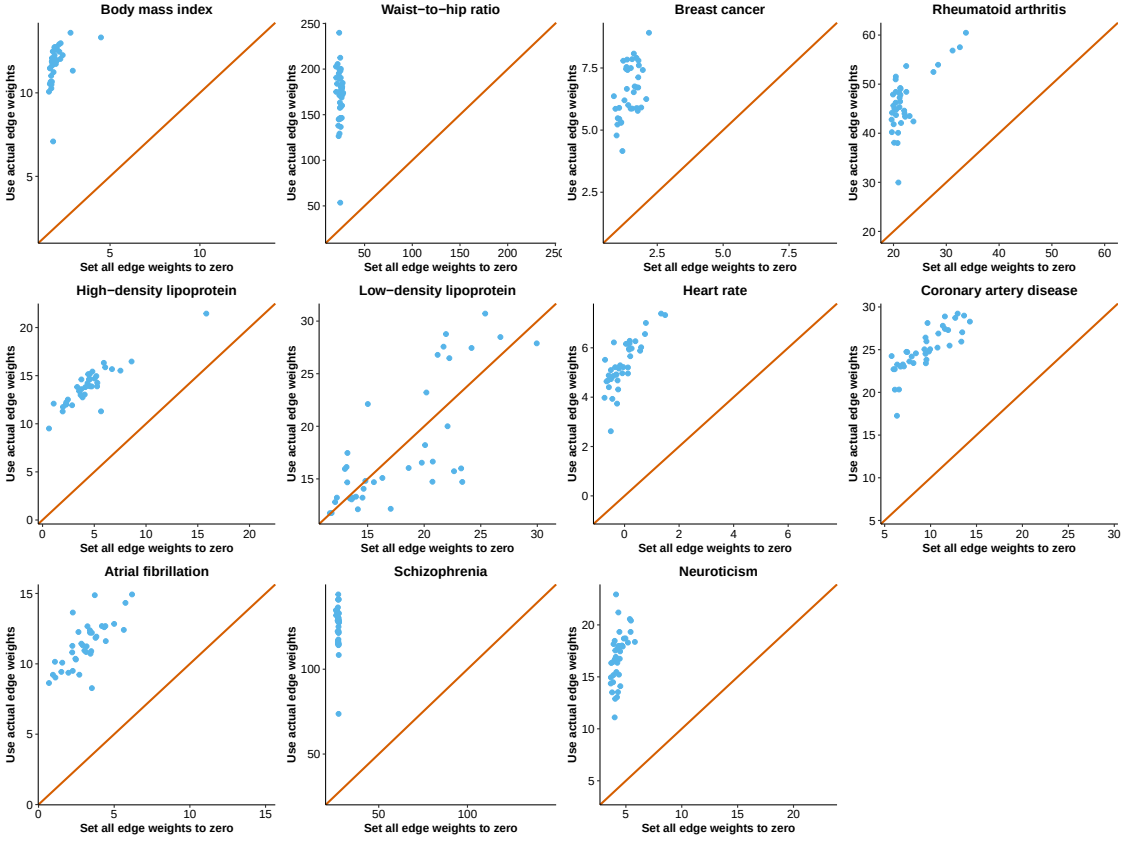


Supplementary Figure 13

RSS-NET results with zero TF-TG edge weights. Here we repeat RSS-NET analyses of 11 GWAS traits on 38 TF-TG regulatory networks by setting all TF-TG edge weights zero (i.e., $v_{gt} = 0$ in main text Equation (5)). Each panel below corresponds to one of 11 GWAS traits that pass the near-gene enrichment control for all 38 networks in our original analyses using actual TF-TG edge weights (Supplementary Table 2). We remove myocardial infarction since it is a subtype of coronary artery disease. In each panel, each point represents one of the 38 regulatory networks. The x- and y-axis values of each point are log 10 BFs (M_1 against M_0) based on zero and actual edge weights of a network respectively. All diagonal lines have slope 1 and intercept 0. Numerical values are available at `https://suwonglab.github.io/rss-net/results.html`.

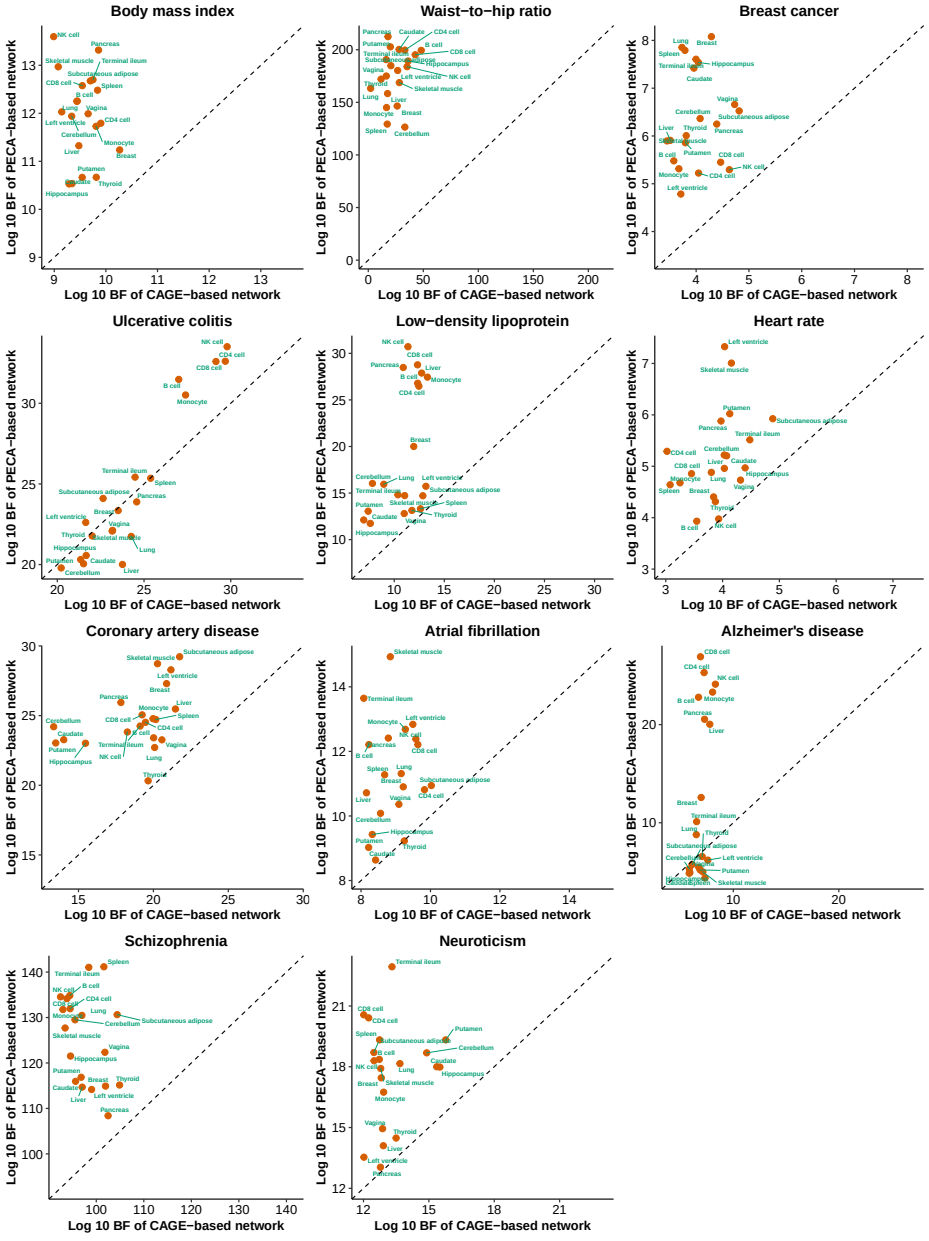
For 10 out 11 traits, BFs based on actual edge weights are unanimously larger than BFs based on zero edge weights across all 38 networks (i.e., all points above the diagonal line). The only exception is low-density lipoprotein cholesterol, for which BFs based on actual edge weights and BFs based on zero edge weights are not significantly different across 38 networks (two-sided Wilcoxon $P = 0.91$).

Of note, setting TF-TG edge weights zero is different from setting the enrichment parameter σ^2 zero; see main text Equations (2), (5) and (6). Setting edge weights zero induces the following effect size distribution: $\beta_j \sim \pi_j \cdot \mathcal{N}(\sum_{g \in \mathbf{G}_j} c_{jg} \cdot \gamma_{jg},\ \sigma_0^2) + (1-\pi_j) \cdot \delta_0$, $\gamma_{jg} \sim \mathcal{N}(0,\ \sigma^2)$. where $\{\mathbf{G}_j, c_{jg}, \gamma_{jg}\}$ are defined in main text Equations (5)-(6). Setting $\sigma^2 = 0$ leads to a different effect size distribution: $\beta_j \sim \pi_j \cdot \mathcal{N}(0,\ \sigma_0^2) + (1-\pi_j) \cdot \delta_0$.



Supplementary Figure 14

Additional results for Figure 5d. Each panel below shows the comparison of PECA- and CAGE-based network enrichments on the same cell types or tissues (labels) and the same GWAS data (caption). In each panel, each point represents a context (cell type or tissue), where x- and y-axis show the RSS-NET log 10 enrichment BFs for the CAGE- and PECA-based networks in this context respectively. Given a context, the analyzed CAGE- and PECA-based networks have the same TF-TG edge counts, and their edge weights are on the same scale (see main text Methods). The RSS-NET analyses of CAGE- and PECA-based networks use the same hyper-parameter grids (Supplementary Table 14). All dashed lines have slope 1 and intercept 0. Numerical values are available at `https://suwonglab.github.io/rss-net/results.html`.

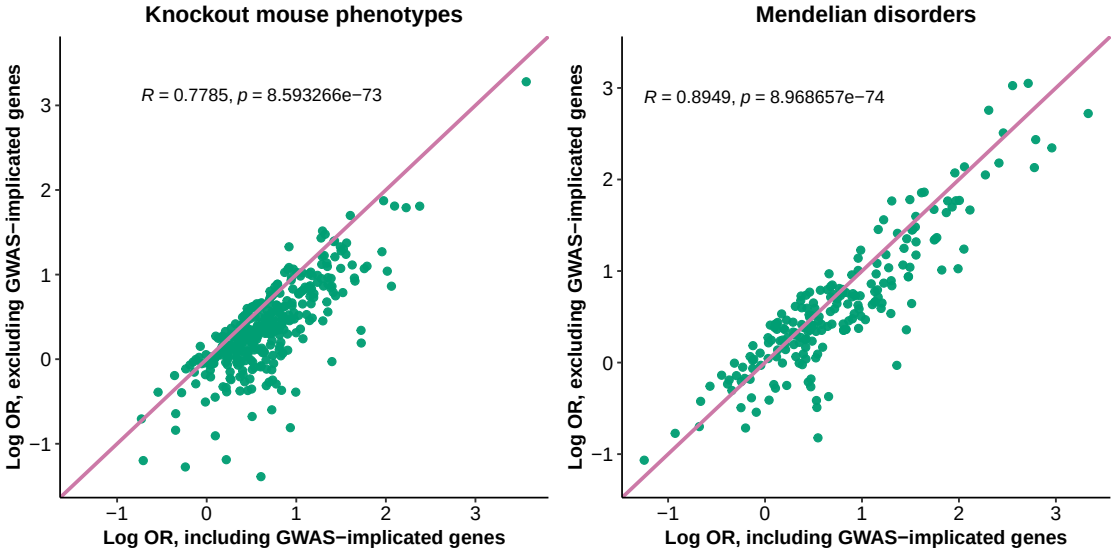


Supplementary Figure 15

Overlap between RSS-NET prioritized genes and genes from external databases. In each panel, each point represents a pair of GWAS trait and knockout mouse phenotype or human Mendelian disorder. For each point, the log odds ratio (OR) is produced by running R built-in function `fisher.test` on the following 2 by 2 table:

	Query genes	Other genes
Genes with $P_1^{\text{bma}} \geq 0.9$		
Genes with $P_1^{\text{bma}} < 0.9$		

For the first panel, "query genes" correspond to genes that are implicated in one of 27 categories of knockout mouse phenotypes [27], and "other genes" correspond to genes that are not implicated in any mouse phenotype. For the second panel, "query genes" correspond to genes that are implicated in one of 19 categories of Mendelian disorders [28, 29], and "other genes" correspond to genes that are not implicated in any Mendelian disorder. For all points, the log ORs shown on the x-axis are estimated from all genes, and the log ORs shown on the y-axis are estimated after excluding genes that are implicated in GWAS of the same trait at the time of analysis (Supplementary Table 7). The OR values and associated P-values are available in Supplementary Data 4-5. The diagonal lines have slope 1 and intercept 0. Pearson correlation and two-sided P-value are shown in each plot.



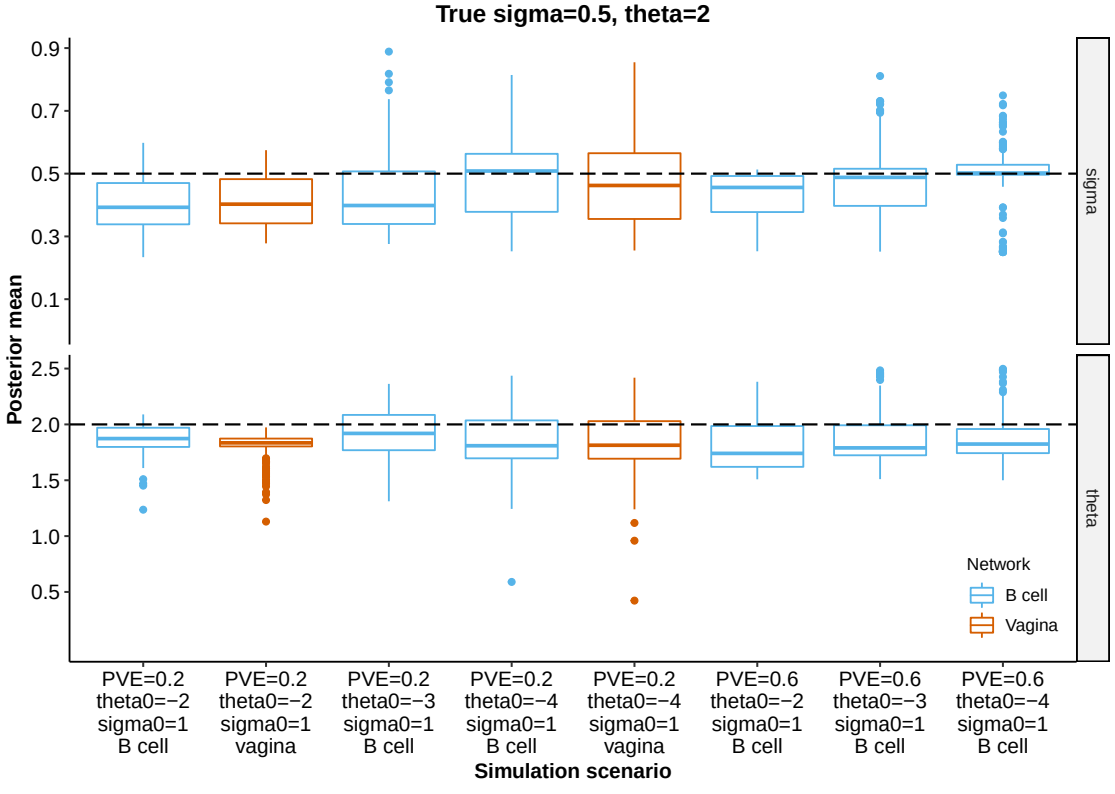
Supplementary Figure 16

Posterior estimation of $\{\sigma, \theta\}$ in simulations. Similar to our calculation of P_1 in Equation (21), we estimate the posterior means of $\{\sigma, \theta\}$ as:

$$\mathrm{E}(\sigma \mid \mathbf{D}) \approx \sum_{h=1}^{H} \sigma^{(h)} \cdot \omega^{\star}(\theta_0^{(h)}, \theta^{(h)}, \sigma_0^{(h)}, \sigma^{(h)}),$$

$$\mathrm{E}(\theta \mid \mathbf{D}) \approx \sum_{h=1}^{H} \theta^{(h)} \cdot \omega^{\star}(\theta_0^{(h)}, \theta^{(h)}, \sigma_0^{(h)}, \sigma^{(h)}),$$

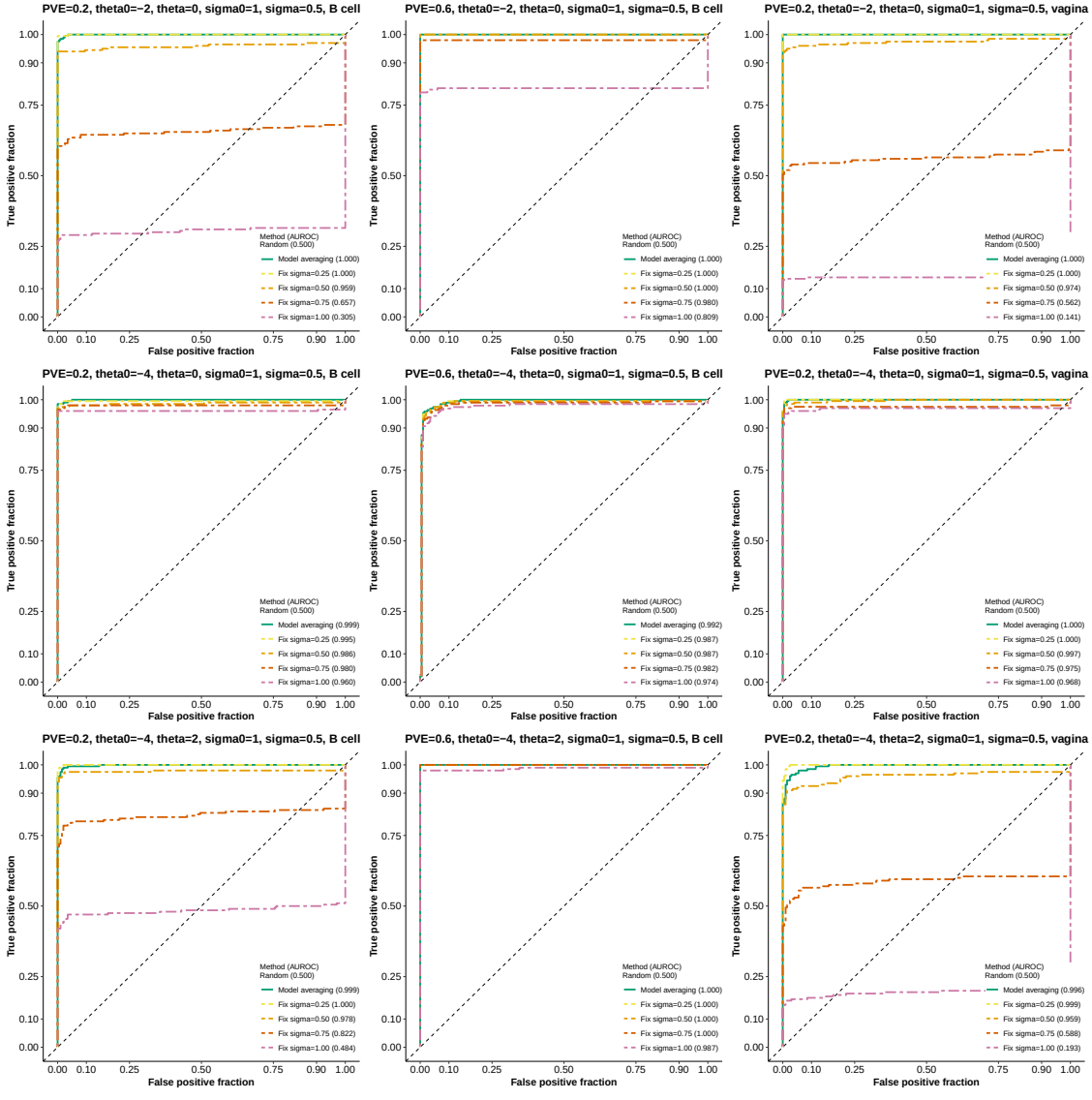
where $\omega^{\star}(\theta_0, \theta, \sigma_0, \sigma)$ is a discrete approximation to the joint posterior of hyper-parameters $\{\theta_0, \theta, \sigma_0, \sigma\}$; see Equation (17). The x-axis indicates the simulation scenarios (true hyper-parameter values of the positive datasets and the target network); see Supplementary Figures 1-2. Each box contains results of 200 independent positive datasets for a given scenario. Box plots indicate median (middle line), 25th, 75th percentile (box) and 5th and 95th percentile (whiskers) as well as outliers (single points). Horizontal dashed lines indicate true values of $\{\sigma, \theta\}$.



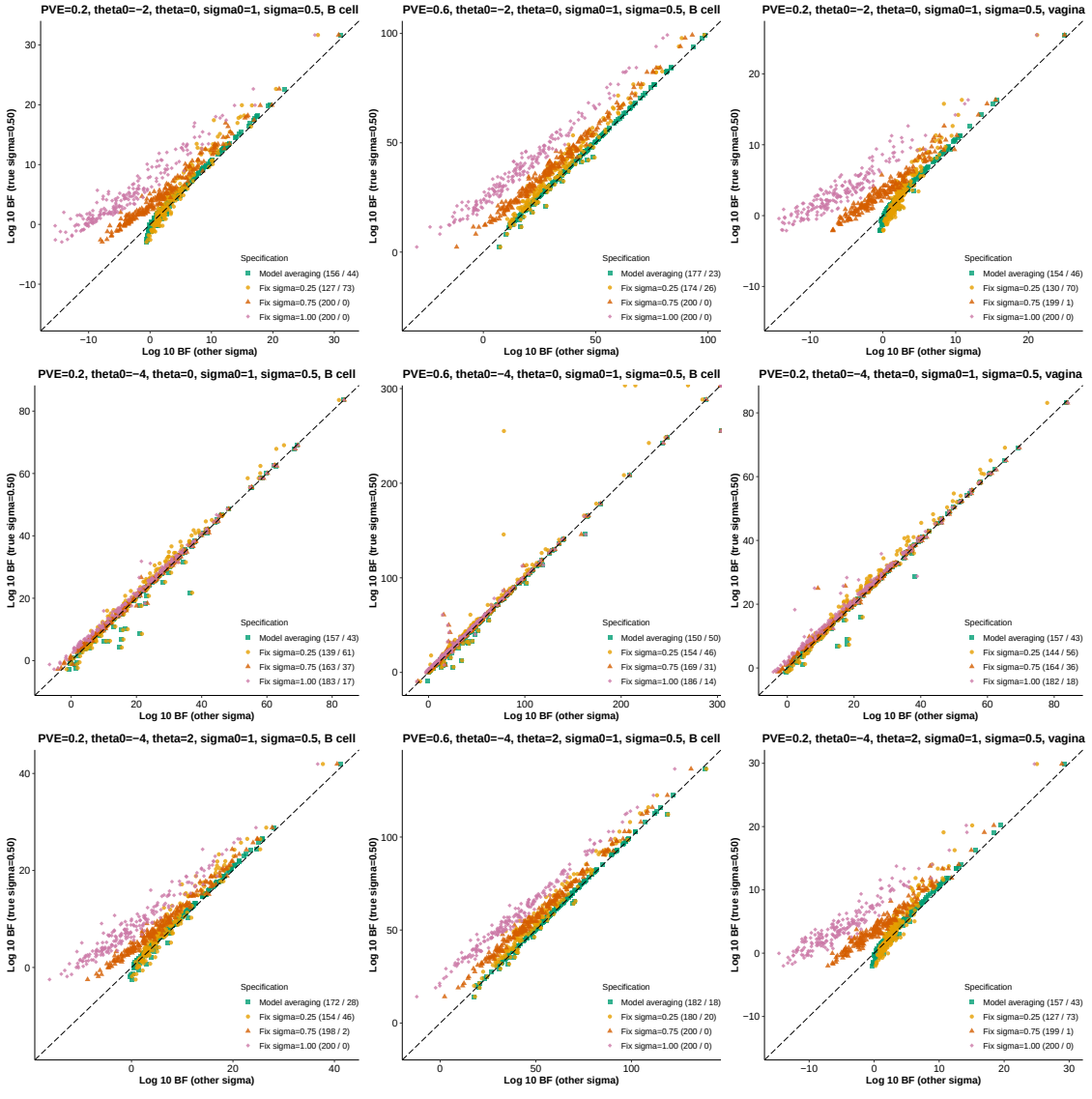
Supplementary Figure 17

Robustness of RSS-NET to σ specification in simulations. Here we repeat RSS-NET analyses of simulated datasets with a fixed value of $\sigma \in \{0.25, 0.5, 0.75, 1\}$. In our original analyses we use a model averaging approach as shown in Equation (19) to account for the uncertainty of hyper-parameters.

Panel **a** The simulated negative and positive GWAS summary data are the same as those from Supplementary Figures 1-2. The caption of each plot shows the true hyper-parameter values of positive datasets and the target network for enrichment testing.



Panel **b** Each plot corresponds to a positive simulation scenario with 200 independent datasets, and its caption indicates the true hyper-parameter values and the target network. In each plot, each point represent a simulated positive dataset. The y-axis value of each point is the log 10 enrichment BF when the value of σ is fixed as 0.5 (ground truth). The x-axis value of each point is the log 10 BF under other σ specifications (differentiated by point shape and color). All dashed lines have slope 1 and intercept 0. The number "$(a \ / \ 200 - a)$" in each legend indicates that BFs based on $\sigma = 0.5$ are higher than BFs based on another σ specification in a simulated datasets for this scenario.

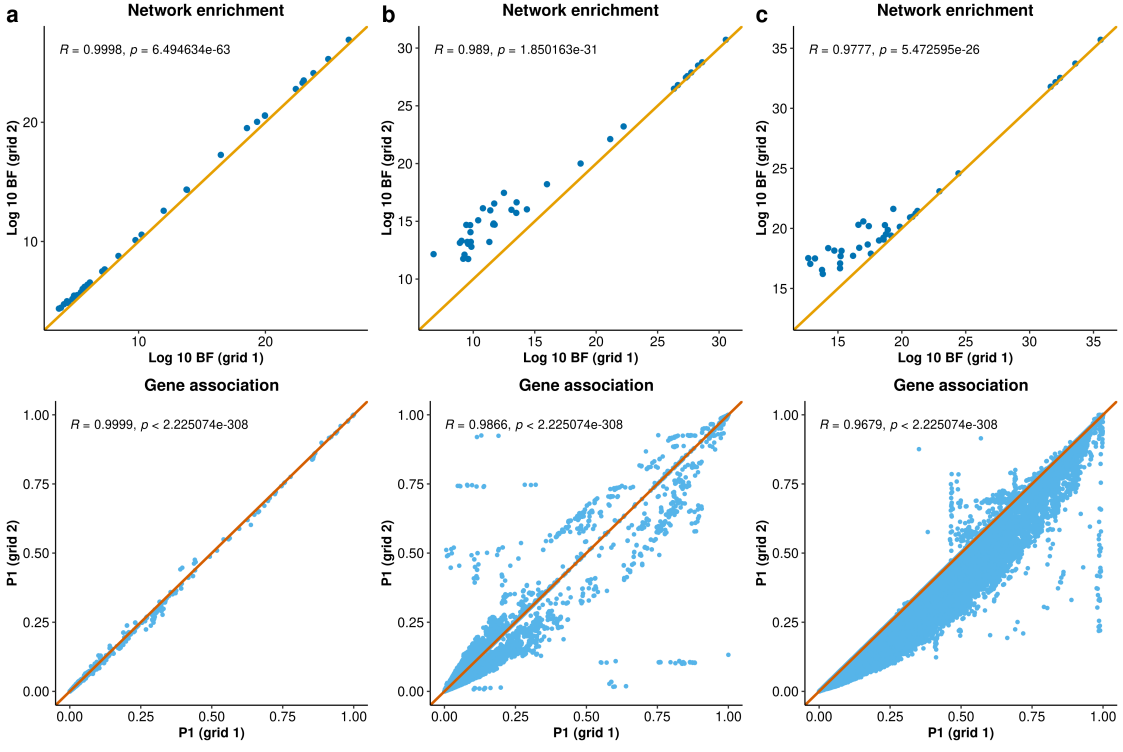


Concluding Remark: Apart from hyper-parameters, RSS-NET uses approximations in the likelihood function [2] and the variational inference [1]. Hence, setting σ as ground truth does not guarantee the optimal outcome in some simulated datasets. Instead we recommend the model averaging approach as shown in Equation (19), which was used in our main analyses.

Supplementary Figure 18

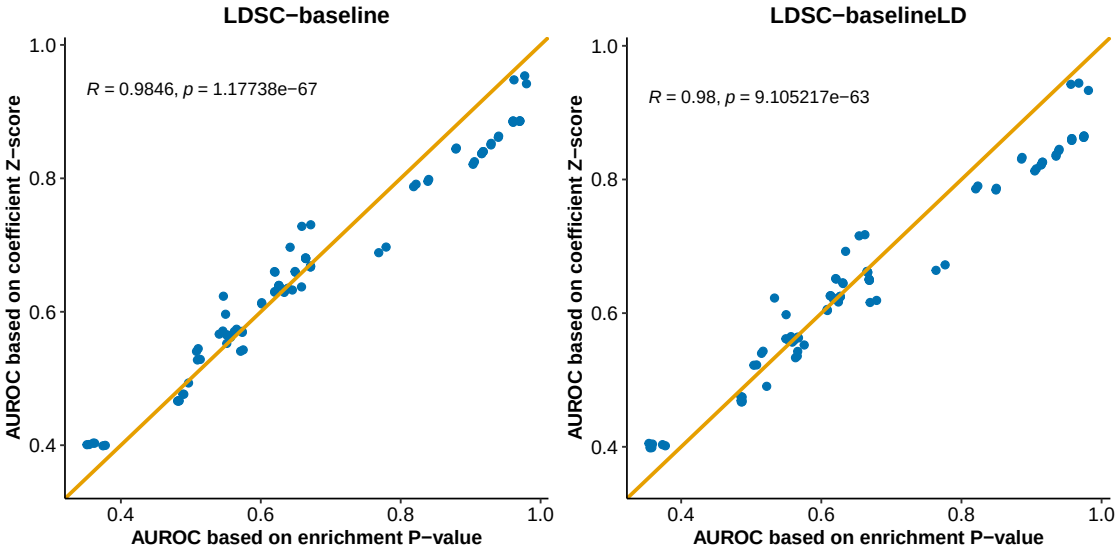
Robustness of RSS-NET to hyper-parameter grid choice in real data. For three GWAS datasets, we run RSS-NET on 38 networks with two different hyper-parameter grids (grid 1: coarse; grid 2: focused), and then compare results of network enrichments (BF, panels in the first row) and gene associations (P_1^{net}, panels in the second row). In each enrichment comparison plot, each point represents a network. In each association comparison plot, each point represent a gene-network pair. All diagonal lines have slope 1 and intercept 0. Pearson correlation and two-sided P-value are shown in each plot.

Panel **a** Alzheimer's disease [10]. In grid 1, $\theta \in (0 : 0.5 : 4)$. In grid 2, $\theta \in (0 : 0.25 : 1)$. For both grids, $\eta = 0.6$, $\rho \in (0 : 0.2 : 0.8)$ and $\theta_0 \in (-5.25 : 0.1 : -4.75)$. Panel **b** low-density lipoprotein [14]. In grid 1, $\theta \in (0 : 0.5 : 3)$. In grid 2, $\theta \in (0 : 0.25 : 1)$. For both grids, $\eta = 0.3$, $\rho \in (0 : 0.2 : 0.8)$ and $\theta_0 \in (-3.75 : 0.05 : -3.5)$. Panel **c** inflammatory bowel disease [7]. In grid 1, $\theta \in (0 : 0.5 : 3)$. In grid 2, $\theta \in (0 : 0.25 : 1)$. For both grids, $\eta = 0.3$, $\rho \in (0 : 0.2 : 0.8)$ and $\theta_0 \in (-3 : 0.05 : -2.8)$.



Supplementary Figure 19

Comparison of two LDSC output statistics in simulations. On each simulated dataset two LDSC methods, LDSC-baseline [30] and LDSC-baselineLD [20], are used to test network enrichments. Each LDSC method produces two statistics, enrichment P-value (denoted as `Enrichment_p` in the software) and enrichment coefficient Z-score (`Coefficient_z-score`), both of which can be used to rank the significance of enrichments. In each panel, each point represent a simulation scenario that contains 200 positive and 200 negative datasets (Supplementary Figures 1-6 and 9). The x- and y-axis of each point show AUROC values evaluated on the basis of enrichment P-value and coefficient Z-score respectively. In total there are 89 simulation scenarios of assessing network enrichments in this study, and thus each panel below consists of 89 points. The diagonal lines in both panels have slope 1 and intercept 0. Pearson correlation and two-sided P-value are shown in each plot.



Supplementary Tables

Supplementary Table 1

GWAS publications, total sample sizes and numbers of genetic variants analyzed by RSS-NET for 18 human traits. See Supplementary Note 8 for data sources and references. Click PMIDs below to view publications online.

Trait (abbreviation)	PMID	# of SNPs (analyzed)	Sample size (cases+controls)
Alzheimer's disease (LOAD)	24162737	1,136,997	17,008+37,154
Neuroticism (NEU)	27089181	1,119,108	170,911
Schizophrenia (SCZ)	25056061	1,113,442	152,805
Body mass index (BMI)	25673413	1,012,465	234,069
Height (HEIGHT)	25282103	1,064,575	253,288
Waist-to-hip ratio (WAIST)	25673412	1,008,898	142,762
Crohn's disease (CD)	26192919	1,064,533	5,956+14,927
Inflammatory bowel disease (IBD)	26192919	1,081,481	12,882+21,770
Rheumatoid arthritis (RA)	24390342	1,158,064	14,361+43,923
Ulcerative colitis (UC)	26192919	1,092,170	6,968+20,464
Breast cancer (BC)	23535729	1,188,902	15,863+40,022
Atrial fibrillation (AF)	28416818	1,175,059	17,931+115,142
Coronary artery disease (CAD)	26343387	1,121,322	60,801+123,504
High-density lipoprotein (HDL)	20686565	1,032,214	99,900
Heart rate (HR)	23583979	1,066,168	92,355
Low-density lipoprotein (LDL)	20686565	1,030,397	95,454
Myocardial infarction (MI)	26343387	1,111,568	42,561+123,504
Type 2 diabetes (T2D)	22885922	1,047,618	12,171+56,862

Supplementary Table 2

To confirm that RSS-NET network enrichment results are unlikely to be driven by generic enrichments harbored in the vicinity of genes, we perform a "near-gene" control analysis. Specifically, we create a "near-gene" control network with 18,334 protein-coding autosomal genes as nodes and no edges, and then analyze this control with RSS-NET on the same GWAS data for 18 traits. The near-gene control analysis is formalized as:

$$\begin{aligned}
\widehat{\boldsymbol{\beta}} &\sim \mathcal{N}(\widehat{\mathbf{S}}\widehat{\mathbf{R}}\widehat{\mathbf{S}}^{-1}\boldsymbol{\beta},\ \widehat{\mathbf{S}}\widehat{\mathbf{R}}\widehat{\mathbf{S}}),\\
\beta_j &\sim \pi_j \cdot \mathcal{N}(\mu_j,\ \sigma_0^2) + (1-\pi_j)\cdot\delta_0,\\
\pi_j &= 1\ /\ [1+10^{-(\theta_0+a_j\theta)}],\\
\mu_j &= \textstyle\sum_{g\in\mathbf{G}_j} c_{jg}\cdot\gamma_{jg},\\
\gamma_{jg} &\sim \mathcal{N}(0,\ \sigma^2),
\end{aligned}$$

where $a_j = 1$ if SNP j is within 100 kb of transcribed region of any protein-coding autosomal gene and 0 otherwise, $\mathbf{G}_j$ denotes the set of all protein-coding genes within 1 Mb window of SNP j, and c_{jg} measures the relative impact of SNP j on gene g (derived from eQTLGen cis-eQTL results; see Supplementary Note 5).

Panel **a** Using the same GWAS data and the same hyper-parameter grids, we compare BFs between each network and the near-gene control under four enrichment models: (1) M_{11}: $\theta > 0$ and $\sigma^2 = 0$; (2) M_{12}: $\theta = 0$ and $\sigma^2 > 0$; (3) M_{13}: $\theta > 0$ and $\sigma^2 > 0$; (4) M_1: $\theta > 0$ or $\sigma^2 > 0$. For all 38 networks and the near-gene control, BF computations are based on the same baseline model (M_0: $\theta = 0$ and $\sigma^2 = 0$). Each row below reports the number of networks with BFs greater than near-gene control BF for each trait and each enrichment model. Trait abbreviations are defined in Supplementary Table 1.

	# of networks			
Trait	M_{11}	M_{12}	M_{13}	M_1
AF	32	38	38	38
BC	33	38	38	38
BMI	9	38	38	38
CAD	38	38	38	38
HDL	38	38	38	38
HR	32	38	38	38
LDL	38	38	38	38
MI	38	38	38	38
NEU	33	38	38	38
RA	38	38	38	38
SCZ	9	38	38	38
WAIST	38	38	38	38
UC	38	22	34	34
LOAD	38	13	11	11
IBD	38	0	6	6
T2D	2	0	21	5
CD	38	0	5	0
HEIGHT	38	0	0	0

Panel **b** Applying the external methods (LDSC, Pascal) to the same GWAS data, we compare enrichment P-values between each network and the near-gene control. Each row below reports the number of networks with enrichment P-value smaller than near-gene control enrichment P-value for each trait and each external method. Numerical values of LDSC and Pascal analyses are available at `https://suwonglab.github.io/rss-net/results.html`.

	# of networks					
	LDSC methods		Pascal methods			
Trait	baseline	baselineLD	max-chi	max-emp	sum-chi	sum-emp
RA	38	38	0	0	2	0
IBD	36	37	0	0	0	0
UC	37	37	1	0	1	0
HR	31	36	0	0	0	0
WAIST	37	36	1	0	27	24
CD	32	34	0	0	0	0
BC	28	32	0	2	1	0
MI	31	30	0	0	15	15
CAD	23	23	0	0	8	0
T2D	25	22	0	0	5	6
LDL	24	18	0	0	1	0
NEU	18	16	0	0	0	0
HEIGHT	6	11	0	0	0	0
HDL	6	10	0	0	1	0
AF	5	5	0	0	7	0
LOAD	1	2	0	0	1	1
SCZ	2	2	0	0	0	0
BMI	0	0	0	0	0	0

Supplementary Table 3

We compute Pearson correlations between five network features and log 10 enrichment BFs, either across 512 trait-network pairs that pass the near-gene control (Supplementary Table 2), or across all 684 trait-network pairs (18 traits and 38 networks). We use R built-in function `cor.test` for all correlation analyses. All tests are two-sided.

	Pearson R (P-value)	
Network feature	512 pairs	684 pairs
% of SNPs in a network	-0.0304 (0.4925)	-0.0403 (0.2924)
# of TF-TG edges	-0.0092 (0.8353)	0.0025 (0.9479)
# of nodes (TF or TG)	-0.0535 (0.2272)	-0.0422 (0.2702)
# of TGs	-0.0531 (0.2308)	-0.0424 (0.2686)
# of TFs	-0.0403 (0.3625)	-0.0184 (0.6312)

Supplementary Table 4

For 512 network-trait pairs passing the near-gene enrichment control (Supplementary Table 2), we compute BFs comparing the baseline model (M_0: $\theta = 0$ and $\sigma^2 = 0$) against three disjoint enrichment models: (1) M_{11}: $\theta > 0$ and $\sigma^2 = 0$; (2) M_{12}: $\theta = 0$ and $\sigma^2 > 0$; (3) M_{13}: $\theta > 0$ and $\sigma^2 > 0$. For each network-trait pair, we compare BFs based on M_{11}, M_{12} and M_{13}, and define the "best" enrichment model as the one with the largest BF. Each row below shows, for each trait, the number of networks with M_{11}, M_{12} and M_{13} being the "best" model, respectively. Trait abbreviations are defined in Supplementary Table 1. Figure 5c is based on the table below.

	# of networks			
Trait	M_{11}	M_{12}	M_{13}	Total
AF	0	0	38	38
BC	0	10	28	38
BMI	0	14	24	38
CAD	0	0	38	38
HDL	0	0	38	38
HR	0	0	38	38
LDL	0	0	38	38
MI	0	0	38	38
NEU	0	0	38	38
RA	0	0	38	38
SCZ	0	38	0	38
WAIST	0	37	1	38
UC	0	0	34	34
LOAD	0	0	11	11
IBD	0	0	6	6
T2D	2	0	3	5
Total	2	99	411	512

Supplementary Table 5

For each trait or trait-network pair, we compute the proportion of genes with higher enrichment P_1 estimates (P_1^{bma} or P_1^{net}) than reference estimates (P_1^{base} or P_1^{near}), among genes with reference P_1 estimates higher than a given cutoff.

Panel **a** Median proportion of genes with P_1^{bma} higher than reference estimates (P_1^{base} or P_1^{near}), among genes with reference estimates higher than a given cutoff. Medians shown below are across 16 traits that have multiple networks more enriched than the near-gene enrichment control (Supplementary Table 2). Figure 5e is based on this panel.

	All genes		Only TFs		Only TGs	
Cutoff	P_1^{base}	P_1^{near}	P_1^{base}	P_1^{near}	P_1^{base}	P_1^{near}
0	0.9828	0.9401	0.9556	0.8829	0.9842	0.9423
0.1	0.9531	0.8811	0.9837	0.8593	0.9499	0.8920
0.2	0.9180	0.8590	0.9514	0.8333	0.9166	0.8707
0.3	0.8856	0.8342	0.9564	0.8000	0.8898	0.8360
0.4	0.8754	0.8061	0.9667	0.7500	0.8784	0.8084
0.5	0.8695	0.7871	0.9444	0.7321	0.8771	0.7979
0.6	0.8499	0.7815	0.9348	0.7321	0.8607	0.7872
0.7	0.8647	0.7795	0.9583	0.7321	0.8624	0.7831
0.8	0.8399	0.7578	1.0000	0.6667	0.8521	0.7657
0.9	0.8167	0.7443	1.0000	0.6870	0.8259	0.7245

Panel **b** Median proportion of genes with P_1^{net} higher than reference estimates (P_1^{base} or P_1^{near}), among genes with reference estimates higher than a given cutoff. Medians shown below are across 512 network-trait pairs passing the near-gene enrichment control (Supplementary Table 2).

	All genes		Only TFs		Only TGs	
Cutoff	P_1^{base}	P_1^{near}	P_1^{base}	P_1^{near}	P_1^{base}	P_1^{near}
0	0.9637	0.9046	0.8926	0.7058	0.9730	0.9325
0.1	0.9203	0.8613	0.8549	0.6667	0.9340	0.8898
0.2	0.9081	0.8494	0.8571	0.6667	0.9114	0.8696
0.3	0.8623	0.8276	0.8313	0.6667	0.8745	0.8537
0.4	0.8622	0.7971	0.8000	0.6000	0.8750	0.8225
0.5	0.8756	0.7891	0.7656	0.6079	0.8827	0.8053
0.6	0.8707	0.7895	0.7895	0.6154	0.8757	0.8000
0.7	0.8750	0.7906	0.7895	0.6206	0.8788	0.8077
0.8	0.8462	0.7679	0.7471	0.5862	0.8583	0.7863
0.9	0.8333	0.7500	0.7230	0.5000	0.8468	0.7619

Supplementary Table 6

For each trait we compute the correlation between RSS-NET enrichment statistic (log 10 BF) and the percentage of network genes with P_1^{net} higher than the reference estimates (P_1^{base} or P_1^{near}), across all networks. Here we only consider 12 traits where all 38 networks show stronger enrichment than the near-gene control (Supplementary Table 2). Rows are ranked by two-sided Pearson P-values. The Bonferroni cutoff is $0.05/12 = 4.2 \times 10^{-3}$. Trait abbreviations are defined in Supplementary Table 1.

Panel **a** The reference P_1 is P_1^{base}.

	Pearson R (- log 10 P-value)		
Trait	All genes	Only TFs	Only TGs
LDL	0.8992 (13.7584)	0.9419 (17.9148)	0.8947 (13.4362)
HR	0.8166 (9.3903)	0.7301 (6.7046)	0.8090 (9.0992)
BMI	0.6810 (5.5959)	0.6628 (5.2378)	0.7271 (6.6306)
AF	0.6687 (5.3518)	0.7041 (6.0910)	0.6492 (4.9850)
MI	0.5851 (3.9420)	0.5452 (3.3972)	0.6125 (4.3594)
SCZ	0.5722 (3.7584)	0.6440 (4.8916)	0.6173 (4.4368)
BC	0.5579 (3.5637)	0.6122 (4.3543)	0.5412 (3.3468)
RA	0.3848 (1.7679)	0.3427 (1.4540)	0.4013 (1.9022)
WAIST	0.3672 (1.6316)	0.4298 (2.1496)	0.3946 (1.8468)
HDL	0.2629 (0.9555)	0.4662 (2.4981)	0.2658 (0.9714)
CAD	0.1948 (0.6175)	0.0261 (0.0574)	0.3409 (1.4413)
NEU	0.1461 (0.4184)	0.0523 (0.1219)	0.2634 (0.9581)

Panel **b** The reference P_1 is P_1^{near}. This panel is reported in the main text.

	Pearson R (- log 10 P-value)		
Trait	All genes	Only TFs	Only TGs
LDL	0.9100 (14.6053)	0.9432 (18.0902)	0.9086 (14.4901)
HR	0.8392 (10.3326)	0.8105 (9.1548)	0.8323 (10.0295)
BMI	0.7247 (6.5719)	0.7711 (7.8321)	0.7873 (8.3440)
MI	0.7079 (6.1755)	0.7797 (8.0986)	0.7372 (6.8854)
AF	0.7069 (6.1523)	0.7894 (8.4116)	0.6956 (5.9022)
NEU	0.5618 (3.6150)	0.6441 (4.8921)	0.6092 (4.3072)
CAD	0.5268 (3.1675)	0.5975 (4.1260)	0.6206 (4.4905)
SCZ	0.5218 (3.1084)	0.6433 (4.8791)	0.5812 (3.8855)
BC	0.5124 (2.9978)	0.5914 (4.0355)	0.4724 (2.5613)
RA	0.5110 (2.9809)	0.5191 (3.0761)	0.5098 (2.9673)
WAIST	0.2877 (1.0977)	0.4315 (2.1654)	0.3785 (1.7184)
HDL	0.2831 (1.0705)	0.4966 (2.8189)	0.3151 (1.2679)

Supplementary Table 7

Here we quantify overlap between RSS-NET prioritized genes ($P_1^{\text{bma}} \geq 0.9$) and genes implicated in the GWAS Catalog [31], for each of 16 traits that pass the near-gene enrichment control (Supplementary Table 2). The "Same GWAS" column reports the overlap between RSS-NET prioritized genes and genes implicated in the same GWAS. The "Later GWAS" column reports the overlap between RSS-NET prioritized genes and genes implicated in a later GWAS of the same trait with increased sample size. The "No GWAS" column reports the number of RSS-NET prioritized genes that were not reported in either the same or the later GWAS at the time of analysis. For example, there are 38 genome-wide significant genes in GWAS Catalog for the same GWAS of HDL [14] that are analyzed by RSS-NET, and 32 of them have $P_1^{\text{bma}} \geq 0.9$; there are 20 new genome-wide significant genes in GWAS Catalog for a later GWAS of HDL [32], and 3 of them have $P_1^{\text{bma}} \geq 0.9$ based on RSS-NET analysis of previous GWAS [14]; there are 142 genes with $P_1^{\text{bma}} \geq 0.9$ that are not reported in either the same or the later GWAS of HDL. All P-values are two-sided and are generated by R built-in function `fisher.test`. Rows are sorted by P-values in the "Same GWAS" column, then P-values in the "Later GWAS" column. Trait abbreviations are defined in Supplementary Table 1.

It is important to note that some published GWAS hits are not necessarily genome-wide significant in the corresponding summary data files. For example, rs34856868 shows $P = 9.80 \times 10^{-9}$ for association with IBD in Table 2 of [7]; however, the P-value of the same SNP is 0.27 in the corresponding summary data file (`EUR.IBD.gwas.assoc.gz`). For this SNP, the result in Table 2 of [7] was indeed obtained from a combined analysis of data on both GWAS and Immunochip arrays, whereas the result in the summary data file was only based on GWAS arrays. Because of this type of potential discrepancy between summary data files and corresponding publications, the overlap fractions shown in the "Same GWAS" column are not very high for certain traits.

	Fraction (- log 10 P-value)		
Trait	Same GWAS	Later GWAS	No GWAS
RA	44 / 61 (63.75)	4 / 7 (5.79)	190
HDL	32 / 38 (54.66)	3 / 20 (2.87)	142
LDL	26 / 27 (48.62)	1 / 18 (0.75)	141
SCZ	47 / 58 (45.93)	138 / 330 (82.51)	638
AF	16 / 18 (36.05)	10 / 83 (12.79)	35
CAD	18 / 20 (35.64)	21 / 140 (21.75)	72
IBD	33 / 80 (35.50)	4 / 12 (4.60)	101
WAIST	16 / 33 (29.73)	5 / 233 (3.11)	36
HR	10 / 15 (22.08)	2 / 43 (2.33)	29
BC	8 / 9 (21.03)	4 / 107 (4.82)	15
BMI	13 / 59 (19.61)	4 / 461 (1.39)	33
MI	8 / 9 (17.45)	NA	61
T2D	6 / 16 (12.26)	5 / 110 (6.29)	8
LOAD	6 / 18 (11.12)	11 / 29 (22.14)	18
UC	6 / 17 (7.92)	2 / 5 (3.02)	126
NEU	3 / 5 (6.58)	9 / 122 (10.40)	28

Supplementary Table 8

Examples of RSS-NET highlighted genes that were reported in GWAS of the same data. The "mouse trait" column is based on the Mouse Genome Informatics [27]. The "therapeutic/clinical evidence" column is based on the Online Mendelian Inheritance in Man [29] and Therapeutic Target Database [33]. Click entries in these two columns to view details online. CS: cardiovascular system; Metab.: metabolism; NS: nervous system.

Trait	Gene (Role)	P_1^{base}	P_1^{near}	P_1^{bma}	P_1^{net} (Network, BF)	Mouse trait	Therapeutic/clinical evidence
BMI	*ADCY3* (TG)	1.00	1.00	1.00	1.00 (Ileum, 4.99×10^{12})	Growth	Severe obesity
	NPC1 (TG)	0.72	0.93	0.97	0.97 (Colon, 5.49×10^{12})	Growth	Niemann-Pick disease
	RPL27A (TG)	0.66	0.70	0.86	0.90 (Pancreas, 2.07×10^{13})	Growth	
WAIST	*PIGC* (TG)	1.00	1.00	1.00	1.00 (Esophagus, 6.78×10^{239})		GPIBD16
	TBX15 (TF)	1.00	1.00	1.00	1.00 (Pancreas, 2.72×10^{212})	Growth	Cousin syndrome
	PPARG (TF)	0.94	0.94	0.96	0.96 (Esophagus, 6.78×10^{239})	Metab.	Insulin resistance, Obesity
BC	*FGFR2* (TG)	1.00	1.00	1.00	1.00 (Lung, 7.14×10^{7})	Growth	Pfeiffer syndrome
	PTHLH (TG)	1.00	1.00	1.00	1.00 (Aorta, 8.27×10^{8})	Growth	Brachydactyly E2
	MKL1 (TG)	0.80	0.80	0.91	0.92 (Spleen, 6.19×10^{7})		AMKL
RA	*CD40* (TG)	1.00	1.00	1.00	1.00 (B cell, 3.31×10^{57})	Immune	HIGM3
	CTLA4 (TG)	1.00	1.00	1.00	1.00 (CD8, 6.96×10^{56})	Immune	ALPS5
	ICOS (TG)	1.00	1.00	1.00	1.00 (CD8, 6.96×10^{56})	Immune	Immunodeficiency CV1
	IL2RA (TG)	1.00	1.00	1.00	1.00 (NK cell, 2.95×10^{60})	Immune	Immunodeficiency 41
	IL6R (TG)	1.00	1.00	1.00	1.00 (Monocyte, 8.91×10^{53})	Immune	
	RASGRP1 (TG)	1.00	1.00	1.00	1.00 (NK cell, 2.95×10^{60})	Immune	Immunodeficiency 64
	STAT4 (TF)	1.00	1.00	1.00	1.00 (NK cell, 2.95×10^{60})	Immune	
	TYK2 (TG)	1.00	1.00	1.00	1.00 (NK cell, 2.95×10^{60})	Immune	Immunodeficiency 35
IBD	*ATG16L1* (TG)	1.00	1.00	1.00	1.00 (NK cell, 5.07×10^{35})	Immune	
	BACH2 (TF)	1.00	1.00	1.00	1.00 (NK cell, 5.07×10^{35})	Immune	Immunodeficiency 60
	FCGR2A (TG)	1.00	1.00	1.00	1.00 (Monocyte, 6.28×10^{31})	Immune	Lupus nephritis
	JAK2 (TG)	1.00	1.00	1.00	1.00 (Monocyte, 6.28×10^{31})	Immune	Baricitinib
	STAT3 (TF)	1.00	1.00	1.00	1.00 (NK cell, 5.07×10^{35})	Immune	ADMIO1
	TYK2 (TG)	0.96	0.98	0.98	0.98 (NK cell, 5.07×10^{35})	Immune	Immunodeficiency 35
	STAT4 (TF)	0.77	0.82	0.91	0.91 (NK cell, 5.07×10^{35})	Immune	Susceptibility to SLE
HDL	*ABCA1* (TG)	1.00	1.00	1.00	1.00 (Liver, 2.81×10^{21})	Metab.	Tangier disease, Probucol
	APOB (TG)	1.00	1.00	1.00	1.00 (Liver, 2.81×10^{21})	Metab.	FCHL2, FHBL1, Mipomersen
	GALNT2 (TG)	1.00	1.00	1.00	1.00 (Liver, 2.81×10^{21})	Metab.	
	LIPG (TG)	1.00	1.00	1.00	1.00 (Liver, 2.81×10^{21})	Metab.	GSK-264220A
	LPL (TG)	1.00	1.00	1.00	1.00 (Omentum, 8.70×10^{13})	Metab.	Clofibrate Gemfibrozil
	SCARB1 (TG)	1.00	1.00	1.00	1.00 (Liver, 2.81×10^{21})	Metab.	
LDL	*APOB* (TG)	1.00	1.00	1.00	1.00 (Liver, 7.66×10^{27})	Metab.	FCHL2, FHBL1, Mipomersen
	HMGCR (TG)	1.00	1.00	1.00	1.00 (CD8, 5.86×10^{28})	Metab.	
	HNF1A (TF)	1.00	1.00	1.00	1.00 (CD8, 5.86×10^{28})	Metab.	MODY3, Hepatic adenomas
	LDLR (TG)	1.00	1.00	1.00	1.00 (Pancreas, 3.04×10^{28})	Metab.	FHCL1
	NPC1L1 (TG)	1.00	1.00	1.00	1.00 (Liver, 7.66×10^{27})	Metab.	Response to ezetimibe
T2D	*TCF7L2* (TF)	1.00	1.00	1.00	1.00 (NK cell, 1.49×10^{77})	Metab.	Susceptibility to T2D
	IGF2BP2 (TG)	0.99	1.00	1.00	1.00 (Ileum, 4.52×10^{62})	Metab.	Susceptibility to T2D
	PPARG (TF)	0.98	1.00	1.00	1.00 (Prostate, 5.64×10^{66})	Metab.	Insulin resistance, Obesity
	PROX1 (TG)	0.32	0.92	0.92	0.92 (Thyroid, 3.61×10^{61})	CS, Metab.	
HR	*MYH6* (TG)	1.00	1.00	1.00	1.00 (Heart, 2.12×10^{7})	CS, Muscle	Cardiomyopathy D1EE, H14
	GJA1 (TG)	1.00	1.00	1.00	1.00 (Uterus, 1.51×10^{6})	CS, Muscle	AVSD3, HLHS1
	EPHB4 (TG)	1.00	1.00	1.00	1.00 (Aorta, 2.43×10^{7})	CS, Muscle	CMAVM2, LMPHM7
	TTN (TG)	1.00	1.00	1.00	1.00 (Aorta, 2.43×10^{7})	CS, Muscle	Cardiomyopathy D1G, H9
	GNB4 (TG)	0.55	0.64	0.91	0.91 (Aorta, 2.43×10^{7})	CS	CMTDIF
CAD	*SMAD3* (TF)	0.99	1.00	1.00	1.00 (Adipose, 1.67×10^{29})	CS, Blood	Loeys-Dietz syndrome 3
	APOB (TG)	0.94	0.98	0.99	0.99 (Liver, 2.99×10^{25})	CS, Metab.	FCHL2, FHBL1, Mipomersen
	ATP2B1 (TG)	0.91	0.96	0.98	0.99 (Heart, 1.93×10^{28})	CS, Muscle	
	MRAS (TG)	0.89	0.96	0.98	0.98 (Adipose, 1.67×10^{29})	Immune	Noonan syndrome 11
	APOE (TG)	0.80	0.92	0.96	0.96 (Adipose, 1.67×10^{29})	CS, Metab.	Hyperlipoproteinemia 3
	ABHD2 (TG)	0.81	0.92	0.95	0.95 (Adipose, 1.67×10^{29})	CS	
	IL6R (TG)	0.77	0.91	0.93	0.94 (Heart, 1.93×10^{28})	Immune	Serum level of IL6
	FURIN (TG)	0.69	0.86	0.93	0.95 (Heart, 1.93×10^{28})	CS	
AF	*TBX5* (TF)	1.00	1.00	1.00	1.00 (Heart, 2.15×10^{14})	CS, Muscle	Holt-Oram syndrome
	PITX2 (TF)	1.00	1.00	1.00	1.00 (Muscle, 8.55×10^{14})	CS, Muscle	
	CAV1 (TG)	1.00	1.00	1.00	1.00 (Muscle, 8.55×10^{14})	CS, Muscle	PP hypertension 3
	SH3PXD2A (TG)	1.00	1.00	1.00	1.00 (Muscle, 8.55×10^{14})	CS, Muscle	
	ASAH1 (TG)	1.00	1.00	1.00	1.00 (Muscle, 8.55×10^{14})	Metab.	Farber lipogranulomatosis
	TTN (TG)	0.95	0.97	0.99	0.99 (Muscle, 8.55×10^{14})	CS, Muscle	Cardiomyopathy D1G, H9
LOAD	*APOE* (TG)	1.00	1.00	1.00	1.00 (Liver, 1.09×10^{20})	Liver, NS	Hyperlipoproteinemia 3
	BIN1 (TG)	1.00	1.00	1.00	1.00 (CD8, 8.31×10^{26})	Muscle	Centronuclear myopathy 2

(continued)

Trait	Gene (Role)	P_1^{base}	P_1^{near}	P_1^{bma}	P_1^{net} (Network, BF)	Mouse trait	Therapeutic/clinical evidence
	CLU (TG)	1.00	1.00	1.00	1.00 (Aorta, 3.24×10^{23})	CS, Muscle	
	EPHA1 (TG)	0.99	1.00	1.00	1.00 (CD8, 8.31×10^{26})	Immune	
	CD2AP (TG)	0.70	0.95	0.98	0.98 (Pancreas, 3.53×10^{20})	Metab.	FSGS3
SCZ	*CACNA1C* (TG)	1.00	1.00	1.00	1.00 (Spleen, 1.44×10^{141})	Immune, NS	Timothy syndrome
	TCF4 (TF)	1.00	1.00	1.00	1.00 (Colon, 1.20×10^{144})	Immune, NS	Pitt-Hopkins syndrome
	DPYD (TG)	1.00	1.00	1.00	1.00 (Spleen, 1.44×10^{141})	Neuron	DPD deficiency
NEU	*LINGO1* (TG)	0.99	1.00	1.00	1.00 (Putamen, 2.12×10^{19})	NS	Mental retardation
	SBF2 (TG)	0.94	0.97	0.99	0.99 (Lung, 1.42×10^{18})	Neuron, NS	CMT4B2
	PAFAH1B1 (TG)	0.77	0.93	1.00	1.00 (Cerebellum, 4.85×10^{18})	Neuron, NS	Lissencephaly

Supplementary Table 9

Descriptive statistics of proportion of GWAS SNPs inside a network ($a_j = 1$) over 18 traits for each of 38 regulatory networks. Rows are sorted by the "Median" column.

	Summary of $\sum_{j=1}^{p} a_j/p$				
Network	Min	Q1	Median	Q3	Max
Vagina	0.6400	0.6423	0.6438	0.6481	0.6505
Lung	0.6377	0.6399	0.6414	0.6460	0.6485
Thyroid	0.6287	0.6305	0.6324	0.6368	0.6394
Terminal ileum	0.6145	0.6170	0.6188	0.6230	0.6256
Testis	0.6109	0.6129	0.6149	0.6193	0.6220
Esophagus mucosa	0.6068	0.6090	0.6107	0.6153	0.6180
Sun-exposed skin	0.6014	0.6037	0.6056	0.6101	0.6129
Spleen	0.5965	0.5988	0.6003	0.6056	0.6085
Visceral omentum	0.5875	0.5901	0.5912	0.5963	0.5989
Prostate	0.5738	0.5762	0.5788	0.5828	0.5856
Stomach	0.5632	0.5655	0.5677	0.5725	0.5753
Adrenal gland	0.5576	0.5605	0.5621	0.5673	0.5699
Uterus	0.5508	0.5534	0.5551	0.5602	0.5631
Omnibus	0.5475	0.5494	0.5516	0.5567	0.5597
Esophagus muscularis	0.5442	0.5469	0.5489	0.5535	0.5561
Hippocampus	0.5429	0.5450	0.5463	0.5516	0.5545
Cortex	0.5068	0.5083	0.5101	0.5164	0.5197
Ovary	0.4999	0.5028	0.5049	0.5103	0.5135
Atrial appendage	0.4976	0.5001	0.5017	0.5070	0.5099
Transverse colon	0.4930	0.4956	0.4981	0.5034	0.5067
Caudate	0.4914	0.4930	0.4946	0.5001	0.5033
Tibial nerve	0.4876	0.4900	0.4918	0.4971	0.5002
Putamen	0.4701	0.4721	0.4743	0.4793	0.4824
CD4 cell	0.4468	0.4487	0.4521	0.4578	0.4614
Breast	0.4426	0.4451	0.4477	0.4532	0.4564
Sigmoid colon	0.4397	0.4425	0.4446	0.4501	0.4533
Cerebellum	0.4366	0.4385	0.4405	0.4467	0.4503
CD8 cell	0.4198	0.4213	0.4249	0.4307	0.4345
Subcutaneous adipose	0.4170	0.4193	0.4219	0.4275	0.4309
Gastroesophageal junction	0.4150	0.4180	0.4202	0.4260	0.4294
Skeletal muscle	0.4116	0.4145	0.4169	0.4225	0.4258
Monocyte	0.4093	0.4111	0.4148	0.4202	0.4236
Aorta	0.3766	0.3786	0.3817	0.3869	0.3901
Left ventricle	0.3713	0.3742	0.3768	0.3826	0.3859
B cell	0.3669	0.3691	0.3730	0.3786	0.3822
Liver	0.3611	0.3640	0.3676	0.3730	0.3763
NK cell	0.3494	0.3508	0.3549	0.3605	0.3643
Pancreas	0.3241	0.3262	0.3300	0.3355	0.3390

Supplementary Table 10

Descriptive statistics of 38 regulatory networks. Rows are sorted by "# of TF-TG edges".

Network	# of TF-TG edges	# of TGs	# of TFs
Stomach	96135	3727	450
Omnibus	96008	9317	571
Esophagus mucosa	95797	4891	419
Terminal ileum	95596	3700	408
Pancreas	95446	2941	481
Transverse colon	95446	3696	435
Lung	95361	4123	415
Breast	95355	3260	442
Left ventricle	95030	2546	424
Sun-exposed skin	94988	4144	448
Gastroesophageal junction	94838	2380	443
Atrial appendage	94749	2564	399
Vagina	94592	4325	431
Sigmoid colon	94549	2545	460
Esophagus muscularis	94408	2993	429
Visceral omentum	94334	3701	428
Tibial nerve	94133	3214	431
Subcutaneous adipose	94005	2718	434
Ovary	93822	3659	430
Liver	93706	3398	412
Spleen	93494	4358	397
Aorta	93408	3295	414
Uterus	93057	3569	453
Skeletal muscle	92832	2308	490
NK cell	92399	3105	376
Thyroid	92263	5369	450
CD8 cell	91925	3439	424
B cell	91728	3018	436
Prostate	91722	3165	474
CD4 cell	91569	3643	433
Monocyte	91401	3702	384
Testis	90804	4646	474
Cortex	90579	2972	333
Adrenal gland	90401	2758	434
Caudate	89754	2547	374
Cerebellum	89371	3301	336
Hippocampus	88948	2592	366
Putamen	88634	2380	365

Supplementary Table 11

Descriptive statistics of TF-TG edge weights $\{v_{gt}\}$ over all TF-TG edges available in each of 38 regulatory networks. Rows are sorted by the "Median" column.

	Summary of v_{gt}				
Network	Min	Q1	Median	Q3	Max
Cortex	0.6946	0.7105	0.7315	0.7636	1.0000
Putamen	0.6797	0.6972	0.7201	0.7560	0.9944
Cerebellum	0.6817	0.6970	0.7170	0.7484	1.0000
Caudate	0.6704	0.6880	0.7113	0.7469	1.0000
Hippocampus	0.6678	0.6843	0.7063	0.7403	0.9977
Atrial appendage	0.6489	0.6674	0.6920	0.7314	1.0000
CD8 cell	0.6385	0.6553	0.6775	0.7125	1.0000
Subcutaneous adipose	0.6339	0.6520	0.6769	0.7160	1.0000
Omnibus	0.6278	0.6484	0.6748	0.7143	1.0000
Transverse colon	0.6353	0.6518	0.6740	0.7099	1.0000
Terminal ileum	0.6362	0.6522	0.6736	0.7082	1.0000
Adrenal gland	0.6296	0.6445	0.6650	0.6987	1.0000
CD4 cell	0.6266	0.6430	0.6647	0.6989	1.0000
Gastroesophageal junction	0.6233	0.6413	0.6645	0.7020	1.0000
Stomach	0.6276	0.6431	0.6640	0.6984	1.0000
Skeletal muscle	0.6187	0.6366	0.6609	0.6998	1.0000
Lung	0.6243	0.6393	0.6597	0.6927	1.0000
NK cell	0.6138	0.6324	0.6568	0.6949	1.0000
B cell	0.6101	0.6278	0.6515	0.6885	1.0000
Breast	0.6087	0.6263	0.6497	0.6870	1.0000
Spleen	0.6122	0.6279	0.6492	0.6831	1.0000
Monocyte	0.6088	0.6259	0.6488	0.6844	1.0000
Sigmoid colon	0.6079	0.6250	0.6482	0.6849	1.0000
Prostate	0.6047	0.6210	0.6427	0.6780	1.0000
Left ventricle	0.6006	0.6185	0.6418	0.6788	1.0000
Esophagus muscularis	0.5972	0.6157	0.6405	0.6797	1.0000
Uterus	0.5977	0.6142	0.6369	0.6728	1.0000
Sun-exposed skin	0.5984	0.6149	0.6368	0.6724	1.0000
Ovary	0.5938	0.6122	0.6368	0.6763	1.0000
Tibial nerve	0.5920	0.6104	0.6350	0.6740	1.0000
Liver	0.5867	0.6060	0.6312	0.6707	1.0000
Pancreas	0.5847	0.6046	0.6305	0.6722	1.0000
Vagina	0.5852	0.6010	0.6228	0.6581	1.0000
Visceral omentum	0.5814	0.5977	0.6198	0.6552	1.0000
Aorta	0.5646	0.5826	0.6067	0.6438	1.0000
Esophagus mucosa	0.5643	0.5809	0.6034	0.6400	1.0000
Thyroid	0.5621	0.5766	0.5966	0.6288	0.9587
Testis	0.5512	0.5675	0.5896	0.6252	1.0000

Supplementary Table 12

Descriptive statistics of cis-eQTL databases used in this study. Rows are sorted by the "# of SNP-gene pairs" column.

Source of eQTL	# of SNP-gene pairs	# of genes	# of SNPs
Testis, GTEx (V7)	15,759,533	17,366	1,230,746
Prostate, GTEx (V7)	14,585,826	16,085	1,219,002
Terminal ileum, GTEx (V7)	14,546,128	16,041	1,215,324
Lung, GTEx (V7)	14,483,585	15,973	1,214,950
Sun-exposed skin, GTEx (V7)	14,404,353	15,869	1,213,959
Transverse colon, GTEx (V7)	14,402,835	15,875	1,215,596
Breast, GTEx (V7)	14,392,706	15,872	1,212,389
Vagina, GTEx (V7)	14,380,972	15,869	1,217,316
Thyroid, GTEx (V7)	14,336,449	15,810	1,214,645
Tibial nerve, GTEx (V7)	14,288,824	15,761	1,217,709
Cortex, GTEx (V7)	14,261,947	15,745	1,226,087
Hippocampus, GTEx (V7)	14,233,460	15,719	1,224,980
Stomach, GTEx (V7)	14,220,507	15,684	1,213,739
Caudate, GTEx (V7)	14,209,653	15,689	1,226,945
Sigmoid colon, GTEx (V7)	14,202,126	15,668	1,219,231
Cerebellum, GTEx (V7)	14,192,621	15,679	1,224,520
Visceral omentum, GTEx (V7)	14,151,283	15,611	1,210,453
Uterus, GTEx (V7)	14,144,127	15,599	1,208,996
Spleen, GTEx (V7)	14,088,656	15,568	1,203,063
Ovary, GTEx (V7)	14,083,563	15,543	1,211,776
Subcutaneous adipose, GTEx (V7)	14,074,828	15,528	1,213,026
Esophagus mucosa, GTEx (V7)	14,050,266	15,504	1,201,919
Putamen, GTEx (V7)	14,036,948	15,503	1,226,334
Gastroesophageal junction, GTEx (V7)	13,998,654	15,446	1,217,148
Esophagus muscularis, GTEx (V7)	13,940,389	15,390	1,211,409
Adrenal gland, GTEx (V7)	13,918,735	15,368	1,215,183
Aorta, GTEx (V7)	13,913,807	15,356	1,207,409
Pancreas, GTEx (V7)	13,707,801	15,139	1,200,808
Atrial appendage, GTEx (V7)	13,683,429	15,122	1,207,966
Liver, GTEx (V7)	13,481,989	14,908	1,190,839
Left ventricle, GTEx (V7)	13,120,834	14,507	1,195,904
Skeletal muscle, GTEx (V7)	13,077,954	14,458	1,193,409
CD8 cell, DICE (DB1)	12,467,472	18,258	944,172
CD4 cell, DICE (DB1)	12,464,676	18,258	943,869
NK cell, DICE (DB1)	12,379,934	18,258	937,629
Blood, eQTLGen	12,372,754	13,588	1,173,301
B cell, DICE (DB1)	12,344,429	18,258	934,954
Monocyte, DICE (DB1)	12,344,429	18,258	934,954

Supplementary Table 13

Descriptive statistics of relative cis impact of SNPs on genes $\{c_{jg}\}$ over all SNP-gene pairs available in cis-eQTL databases. Rows are sorted by the "Median" column.

	Summary of $(c_{jg} - 1)$				
Source of eQTL	Min	Q1	Median	Q3	Max
Uterus, GTEx (V7)	0	0.033	0.069	0.118	0.935
Vagina, GTEx (V7)	0	0.032	0.067	0.115	0.932
Putamen, GTEx (V7)	0	0.032	0.067	0.114	0.923
Hippocampus, GTEx (V7)	0	0.031	0.066	0.114	0.912
CD8 cell, DICE (DB1)	0	0.024	0.065	0.118	0.774
CD4 cell, DICE (DB1)	0	0.023	0.064	0.118	0.771
Monocyte, DICE (DB1)	0	0.023	0.064	0.118	0.792
Terminal ileum, GTEx (V7)	0	0.030	0.064	0.109	0.935
NK cell, DICE (DB1)	0	0.023	0.064	0.117	0.762
Ovary, GTEx (V7)	0	0.030	0.063	0.108	0.931
B cell, DICE (DB1)	0	0.021	0.063	0.117	0.777
Cortex, GTEx (V7)	0	0.029	0.061	0.105	0.915
Prostate, GTEx (V7)	0	0.029	0.061	0.104	0.943
Spleen, GTEx (V7)	0	0.028	0.059	0.101	0.942
Caudate, GTEx (V7)	0	0.028	0.059	0.101	0.908
Cerebellum, GTEx (V7)	0	0.027	0.058	0.101	0.952
Liver, GTEx (V7)	0	0.027	0.057	0.098	0.921
Adrenal gland, GTEx (V7)	0	0.026	0.054	0.093	0.945
Sigmoid colon, GTEx (V7)	0	0.024	0.050	0.087	0.942
Gastroesophageal junction, GTEx (V7)	0	0.023	0.049	0.085	0.948
Pancreas, GTEx (V7)	0	0.023	0.049	0.084	0.939
Testis, GTEx (V7)	0	0.023	0.049	0.084	0.927
Stomach, GTEx (V7)	0	0.022	0.046	0.080	0.942
Transverse colon, GTEx (V7)	0	0.022	0.046	0.079	0.943
Breast, GTEx (V7)	0	0.021	0.045	0.078	0.949
Atrial appendage, GTEx (V7)	0	0.021	0.045	0.077	0.943
Aorta, GTEx (V7)	0	0.021	0.045	0.077	0.948
Left ventricle, GTEx (V7)	0	0.021	0.044	0.076	0.943
Visceral omentum, GTEx (V7)	0	0.019	0.041	0.071	0.939
Esophagus muscularis, GTEx (V7)	0	0.019	0.040	0.070	0.955
Tibial nerve, GTEx (V7)	0	0.019	0.040	0.069	0.964
Esophagus mucosa, GTEx (V7)	0	0.018	0.039	0.068	0.938
Subcutaneous adipose, GTEx (V7)	0	0.018	0.038	0.066	0.952
Thyroid, GTEx (V7)	0	0.018	0.038	0.065	0.955
Lung, GTEx (V7)	0	0.018	0.038	0.065	0.928
Sun-exposed skin, GTEx (V7)	0	0.017	0.037	0.064	0.948
Skeletal muscle, GTEx (V7)	0	0.016	0.034	0.058	0.918
Blood, eQTLGen	0	0.003	0.007	0.014	0.855

Supplementary Table 14

RSS-NET hyper-parameter grids used in the present study. For all traits, the grid for θ is $(0 : 0.25 : 1)$ and the grid for ρ is $(0 : 0.2 : 0.8)$. Here $(j : i : k)$ denotes a regularly-spaced vector that starts at j, uses i as the increment, and (roughly) stops at k. RSS-NET analyses of the same trait across different networks use the same hyper-parameter grid specified here. Trait abbreviations are defined in Supplementary Table 1.

Trait	θ_0	η
LOAD	$(-5.25 : 0.1 : -4.75)$	0.6
NEU	$(-4.4 : 0.05 : -4)$	0.3
SCZ	$(-2.175 : 0.025 : -2.05)$	0.3
BMI	$(-4.1 : 0.025 : -3.95)$	0.3
HEIGHT	$(-2.05 : 0.05 : -1.95)$	0.3
WAIST	$(-3 : 0.025 : -2.95)$	0.3
CD	$(-3 : 0.05 : -2.9)$	0.3
IBD	$(-3 : 0.05 : -2.8)$	0.3
RA	$(-3.25 : 0.025 : -3.125)$	0.3
UC	$(-3.05 : 0.025 : -2.95)$	0.3
BC	$(-4.25 : 0.05 : -4)$	0.2
AF	$(-4.5 : 0.25 : -4)$	$0.2, 0.3$
CAD	$(-4 : 0.025 : -3.85)$	0.3
HDL	$(-3.575 : 0.025 : -3.45)$	0.3
HR	$(-4.45 : 0.05 : -4.25)$	$0.3, 0.4$
LDL	$(-3.75 : 0.05 : -3.5)$	0.3
MI	$(-4.3 : 0.025 : -4.1)$	0.3
T2D	$(-4.65 : 0.05 : -4.55)$	$0.4, 0.5, 0.6$

Supplementary Table 15

False positive rates of RSS-NET under different log 10 enrichment BF cutoffs across 47 simulation scenarios of negative datasets.

Panel **a** Descriptive statistics of RSS-NET false positive rates across 17 negative scenarios in Supplementary Figures 1-2 and 9.

Cutoff	Min	Q1	Median	Q3	Max
0	0.0000	0.0000	0.0200	0.1250	0.3800
1	0.0000	0.0000	0.0000	0.0250	0.1800
2	0.0000	0.0000	0.0000	0.0000	0.1150
3	0.0000	0.0000	0.0000	0.0000	0.0900
4	0.0000	0.0000	0.0000	0.0000	0.0600
5	0.0000	0.0000	0.0000	0.0000	0.0550
6	0.0000	0.0000	0.0000	0.0000	0.0450
7	0.0000	0.0000	0.0000	0.0000	0.0400
8	0.0000	0.0000	0.0000	0.0000	0.0400
9	0.0000	0.0000	0.0000	0.0000	0.0350

Panel **b** Descriptive statistics of RSS-NET false positive rates across 6 negative scenarios in Supplementary Figure 3.

Cutoff	Min	Q1	Median	Q3	Max
0	0.0150	0.0438	0.3300	0.6050	0.6300
1	0.0000	0.0000	0.0475	0.1362	0.2000
2	0.0000	0.0000	0.0075	0.0300	0.0750
3	0.0000	0.0000	0.0000	0.0037	0.0150
4	0.0000	0.0000	0.0000	0.0000	0.0050
5	0.0000	0.0000	0.0000	0.0000	0.0050
6	0.0000	0.0000	0.0000	0.0000	0.0000
7	0.0000	0.0000	0.0000	0.0000	0.0000
8	0.0000	0.0000	0.0000	0.0000	0.0000
9	0.0000	0.0000	0.0000	0.0000	0.0000

Panel **c** Descriptive statistics of RSS-NET false positive rates across 6 negative scenarios in Supplementary Figure 4.

Cutoff	Min	Q1	Median	Q3	Max
0	0.0200	0.0563	0.4225	0.8337	0.8700
1	0.0000	0.0000	0.1450	0.3013	0.4200
2	0.0000	0.0000	0.0100	0.0650	0.1400
3	0.0000	0.0000	0.0025	0.0125	0.0200
4	0.0000	0.0000	0.0000	0.0000	0.0050
5	0.0000	0.0000	0.0000	0.0000	0.0000
6	0.0000	0.0000	0.0000	0.0000	0.0000
7	0.0000	0.0000	0.0000	0.0000	0.0000
8	0.0000	0.0000	0.0000	0.0000	0.0000
9	0.0000	0.0000	0.0000	0.0000	0.0000

Panel **d** Descriptive statistics of RSS-NET false positive rates across 6 negative scenarios in Supplementary Figure 5.

Cutoff	Min	Q1	Median	Q3	Max
0	0.0150	0.0287	0.0950	0.1462	0.2350
1	0.0000	0.0000	0.0125	0.0287	0.0350
2	0.0000	0.0000	0.0000	0.0000	0.0050
3	0.0000	0.0000	0.0000	0.0000	0.0000
4	0.0000	0.0000	0.0000	0.0000	0.0000
5	0.0000	0.0000	0.0000	0.0000	0.0000
6	0.0000	0.0000	0.0000	0.0000	0.0000
7	0.0000	0.0000	0.0000	0.0000	0.0000
8	0.0000	0.0000	0.0000	0.0000	0.0000
9	0.0000	0.0000	0.0000	0.0000	0.0000

Panel **e** Descriptive statistics of RSS-NET false positive rates across 12 negative scenarios in Supplementary Figure 6.

Cutoff	Min	Q1	Median	Q3	Max
0	0.2100	0.4625	0.8075	0.9950	1.0000
1	0.0400	0.0900	0.5675	0.9413	1.0000
2	0.0050	0.0138	0.3825	0.8712	1.0000
3	0.0000	0.0050	0.2675	0.7650	0.9900
4	0.0000	0.0050	0.2050	0.6262	0.9800
5	0.0000	0.0050	0.1425	0.4888	0.9700
6	0.0000	0.0050	0.0775	0.3700	0.9550
7	0.0000	0.0000	0.0400	0.2500	0.9250
8	0.0000	0.0000	0.0275	0.1675	0.8900
9	0.0000	0.0000	0.0225	0.1038	0.8550

Panel **f** Descriptive statistics of RSS-NET false positive rates across 47 negative scenarios from Panels **a** to **e** above.

Cutoff	Min	Q1	Median	Q3	Max
0	0.0000	0.0200	0.1350	0.6150	1.0000
1	0.0000	0.0000	0.0300	0.1900	1.0000
2	0.0000	0.0000	0.0000	0.0750	1.0000
3	0.0000	0.0000	0.0000	0.0150	0.9900
4	0.0000	0.0000	0.0000	0.0050	0.9800
5	0.0000	0.0000	0.0000	0.0050	0.9700
6	0.0000	0.0000	0.0000	0.0025	0.9550
7	0.0000	0.0000	0.0000	0.0000	0.9250
8	0.0000	0.0000	0.0000	0.0000	0.8900
9	0.0000	0.0000	0.0000	0.0000	0.8550

Supplementary Table 16

False positive rates of RSS-NET under different P_1 cutoffs across 49 simulation scenarios in Supplementary Figures 7-9.

Panel **a** Descriptive statistics of RSS-NET false positive rates across 8 scenarios with $\theta = 0$ and $\sigma^2 = 0$.

Cutoff	Min	Q1	Median	Q3	Max
0.50	1.9158e-05	4.3152e-05	1.3280e-04	3.1445e-02	6.6337e-02
0.55	1.9158e-06	3.9181e-05	1.1041e-04	1.7910e-02	3.7976e-02
0.60	9.5791e-07	3.7196e-05	9.1061e-05	1.0309e-02	2.1019e-02
0.65	7.1843e-07	3.2524e-05	8.2365e-05	5.7166e-03	1.1068e-02
0.70	7.1843e-07	2.8846e-05	7.1198e-05	3.0852e-03	5.6593e-03
0.75	0.0000e+00	2.5692e-05	6.2483e-05	1.5269e-03	2.9909e-03
0.80	0.0000e+00	2.3357e-05	5.5691e-05	6.9265e-04	1.4404e-03
0.85	0.0000e+00	1.9970e-05	5.2003e-05	2.7399e-04	6.2126e-04
0.90	0.0000e+00	1.7868e-05	4.4425e-05	8.9064e-05	1.8040e-04
0.95	0.0000e+00	1.4094e-05	1.5193e-05	2.1983e-05	6.0961e-05

Panel **b** Descriptive statistics of RSS-NET false positive rates across 16 scenarios with $\theta = 0$ and $\sigma^2 > 0$.

Cutoff	Min	Q1	Median	Q3	Max
0.50	3.4568e-05	7.9590e-05	1.8260e-04	3.9777e-02	1.3713e-01
0.55	3.1765e-05	7.3369e-05	1.3957e-04	2.4210e-02	9.0496e-02
0.60	2.8028e-05	6.3212e-05	1.3148e-04	1.4952e-02	5.6791e-02
0.65	2.5926e-05	5.2579e-05	1.2136e-04	8.2891e-03	3.3777e-02
0.70	2.5225e-05	4.7513e-05	1.1207e-04	4.4175e-03	1.8742e-02
0.75	2.4525e-05	4.1115e-05	1.0169e-04	2.1514e-03	9.8169e-03
0.80	2.1722e-05	3.7478e-05	9.0797e-05	9.1384e-04	4.7398e-03
0.85	2.1255e-05	3.2583e-05	7.9758e-05	3.8602e-04	2.0535e-03
0.90	1.9620e-05	2.8757e-05	7.1112e-05	1.4587e-04	6.6205e-04
0.95	1.7518e-05	2.1299e-05	3.8781e-05	8.3458e-05	1.1029e-04

Panel **c** Descriptive statistics of RSS-NET false positive rates across 8 scenarios with $\theta > 0$ and $\sigma^2 = 0$.

Cutoff	Min	Q1	Median	Q3	Max
0.50	3.1391e-04	2.7308e-03	8.4924e-02	1.0921e-01	1.6779e-01
0.55	1.4283e-04	1.3149e-03	6.0600e-02	8.4700e-02	1.3078e-01
0.60	5.7572e-05	6.2062e-04	4.6203e-02	6.5058e-02	9.8028e-02
0.65	2.6867e-05	2.9597e-04	3.4367e-02	4.9285e-02	6.9542e-02
0.70	1.5078e-05	1.3498e-04	2.4858e-02	3.5227e-02	4.6563e-02
0.75	8.2246e-06	5.4906e-05	1.7402e-02	2.3246e-02	2.8511e-02
0.80	3.0157e-06	2.1513e-05	1.1155e-02	1.3445e-02	1.6516e-02
0.85	5.4831e-07	5.1493e-06	6.1871e-03	6.8245e-03	1.0139e-02
0.90	0.0000e+00	1.0088e-06	2.1403e-03	3.0779e-03	4.6428e-03
0.95	0.0000e+00	2.8646e-07	2.5865e-04	8.9441e-04	1.1598e-03

Panel **d** Descriptive statistics of RSS-NET false positive rates across 17 scenarios with $\theta > 0$ and $\sigma^2 > 0$.

Cutoff	Min	Q1	Median	Q3	Max
0.50	3.6369e-03	1.4967e-02	9.4407e-02	1.1433e-01	2.8226e-01
0.55	1.8042e-03	9.9008e-03	7.1041e-02	9.6921e-02	2.4531e-01
0.60	8.8415e-04	6.4193e-03	5.1986e-02	8.7748e-02	2.0722e-01
0.65	4.2384e-04	3.9577e-03	4.0081e-02	6.6512e-02	1.6879e-01
0.70	2.0918e-04	2.2908e-03	3.0449e-02	5.0644e-02	1.3095e-01
0.75	1.1240e-04	1.1916e-03	2.1810e-02	3.5411e-02	9.5022e-02
0.80	6.4700e-05	5.0790e-04	1.4876e-02	2.5534e-02	6.2574e-02
0.85	3.1671e-05	1.8625e-04	8.6225e-03	1.2890e-02	3.6571e-02
0.90	1.7724e-05	9.7325e-05	4.3930e-03	6.0918e-03	2.0671e-02
0.95	1.1623e-05	4.5801e-05	7.0200e-04	1.6252e-03	8.4746e-03

Panel **e** Descriptive statistics of RSS-NET false positive rates across 49 scenarios from Panels **a** to **d** above.

Cutoff	Min	Q1	Median	Q3	Max
0.50	1.9158e-05	1.5303e-04	2.3029e-02	9.9199e-02	2.8226e-01
0.55	1.9158e-06	1.1734e-04	1.3126e-02	7.1041e-02	2.4531e-01
0.60	9.5791e-07	8.9924e-05	7.9050e-03	5.1986e-02	2.0722e-01
0.65	7.1843e-07	7.9649e-05	4.4754e-03	3.7725e-02	1.6879e-01
0.70	7.1843e-07	7.3107e-05	2.5532e-03	2.8071e-02	1.3095e-01
0.75	0.0000e+00	6.8903e-05	1.3102e-03	2.0162e-02	9.5022e-02
0.80	0.0000e+00	4.7177e-05	6.6729e-04	1.2758e-02	6.2574e-02
0.85	0.0000e+00	3.4724e-05	3.1254e-04	6.7839e-03	3.6571e-02
0.90	0.0000e+00	2.6867e-05	1.2389e-04	2.6863e-03	2.0671e-02
0.95	0.0000e+00	1.8642e-05	4.5801e-05	3.8126e-04	8.4746e-03

Supplementary Table 17

False discovery rates of RSS-NET under different P_1 cutoffs across 49 simulation scenarios in Supplementary Figures 7-9.

Panel **a** Descriptive statistics of RSS-NET false discovery rates across 8 scenarios with $\theta = 0$ and $\sigma^2 = 0$.

Cutoff	Min	Q1	Median	Q3	Max
0.50	1.1188e-01	1.2069e-01	2.8257e-01	4.2073e-01	5.1834e-01
0.55	1.0503e-01	1.1364e-01	1.6283e-01	3.6907e-01	4.7243e-01
0.60	9.3834e-02	1.0941e-01	1.2964e-01	3.2187e-01	4.1940e-01
0.65	9.1880e-02	1.0105e-01	1.1862e-01	2.7451e-01	3.6253e-01
0.70	8.6757e-02	9.2940e-02	1.2742e-01	2.2929e-01	3.0309e-01
0.75	0.0000e+00	8.5148e-02	9.2273e-02	1.8561e-01	2.4523e-01
0.80	0.0000e+00	8.2356e-02	8.4653e-02	1.4323e-01	1.9181e-01
0.85	0.0000e+00	7.5615e-02	8.2129e-02	9.7863e-02	1.4419e-01
0.90	0.0000e+00	6.5377e-02	7.2441e-02	7.9936e-02	8.9158e-02
0.95	0.0000e+00	2.8272e-02	4.4659e-02	6.9743e-02	7.8781e-02

Panel **b** Descriptive statistics of RSS-NET false discovery rates across 16 scenarios with $\theta = 0$ and $\sigma^2 > 0$.

Cutoff	Min	Q1	Median	Q3	Max
0.50	9.3552e-02	1.1167e-01	1.6365e-01	4.6974e-01	5.8608e-01
0.55	8.8140e-02	1.0565e-01	1.3374e-01	4.2736e-01	5.5283e-01
0.60	7.9947e-02	1.0227e-01	1.2180e-01	3.8241e-01	5.1369e-01
0.65	7.5203e-02	9.9072e-02	1.1368e-01	3.3464e-01	4.7198e-01
0.70	7.4227e-02	9.2284e-02	1.0672e-01	2.7871e-01	4.2216e-01
0.75	6.9693e-02	8.8877e-02	9.7413e-02	2.2139e-01	3.6598e-01
0.80	6.5062e-02	8.2524e-02	9.2572e-02	1.5970e-01	3.0792e-01
0.85	6.2286e-02	7.6151e-02	8.2741e-02	1.1029e-01	2.4134e-01
0.90	5.8094e-02	6.7992e-02	7.1887e-02	8.2338e-02	1.6747e-01
0.95	3.3389e-02	5.6885e-02	6.2543e-02	7.3687e-02	7.9775e-02

Panel **c** Descriptive statistics of RSS-NET false discovery rates across 8 scenarios with $\theta > 0$ and $\sigma^2 = 0$.

Cutoff	Min	Q1	Median	Q3	Max
0.50	7.5518e-03	1.5761e-02	1.1404e-01	2.7079e-01	4.1668e-01
0.55	6.0640e-03	1.1565e-02	1.0039e-01	2.2832e-01	3.6661e-01
0.60	4.8994e-03	9.0095e-03	8.7617e-02	1.8033e-01	3.2287e-01
0.65	3.9766e-03	6.8870e-03	7.5335e-02	1.6327e-01	2.8728e-01
0.70	3.0792e-03	5.1992e-03	6.4138e-02	1.6055e-01	2.5701e-01
0.75	2.5024e-03	3.8653e-03	5.3417e-02	1.4547e-01	2.1409e-01
0.80	1.8573e-03	2.6962e-03	4.3209e-02	1.0037e-01	2.3810e-01
0.85	1.3463e-03	1.8628e-03	2.2601e-02	5.9213e-02	1.2366e-01
0.90	0.0000e+00	9.1602e-04	7.7182e-03	2.9053e-02	8.5106e-02
0.95	0.0000e+00	4.2681e-04	2.9493e-03	1.9917e-02	1.1111e-01

Panel **d** Descriptive statistics of RSS-NET false discovery rates across 17 scenarios with $\theta > 0$ and $\sigma^2 > 0$.

Cutoff	Min	Q1	Median	Q3	Max
0.50	7.6112e-03	1.6971e-02	1.4316e-01	3.6950e-01	5.8305e-01
0.55	6.3926e-03	1.2899e-02	1.3107e-01	3.1458e-01	5.4295e-01
0.60	5.4230e-03	1.0222e-02	1.1972e-01	2.8583e-01	4.9586e-01
0.65	4.4578e-03	8.0614e-03	1.0798e-01	2.4972e-01	4.4053e-01
0.70	3.4374e-03	6.1538e-03	9.7503e-02	1.8479e-01	3.8368e-01
0.75	2.6539e-03	4.5910e-03	8.6778e-02	1.4767e-01	3.2701e-01
0.80	1.9719e-03	3.3894e-03	7.1166e-02	1.1008e-01	2.6341e-01
0.85	1.2512e-03	2.3037e-03	5.7580e-02	7.6599e-02	1.9226e-01
0.90	6.2747e-04	1.2001e-03	4.1285e-02	5.9720e-02	1.3092e-01
0.95	1.6947e-04	3.4233e-04	1.8744e-02	4.7921e-02	6.8044e-02

Panel **e** Descriptive statistics of RSS-NET false discovery rates across 49 scenarios from Panels **a** to **d** above.

Cutoff	Min	Q1	Median	Q3	Max
0.50	7.5518e-03	1.1161e-01	1.4864e-01	4.1003e-01	5.8608e-01
0.55	6.0640e-03	9.9029e-02	1.3107e-01	3.5226e-01	5.5283e-01
0.60	4.8994e-03	9.0636e-02	1.2062e-01	2.9651e-01	5.1369e-01
0.65	3.9766e-03	8.0559e-02	1.1214e-01	2.5767e-01	4.7198e-01
0.70	3.0792e-03	7.1282e-02	1.0602e-01	2.2890e-01	4.2216e-01
0.75	0.0000e+00	5.6612e-02	9.3486e-02	1.9724e-01	3.6598e-01
0.80	0.0000e+00	4.3114e-02	8.3675e-02	1.5600e-01	3.0792e-01
0.85	0.0000e+00	2.6688e-02	7.6517e-02	1.0708e-01	2.4134e-01
0.90	0.0000e+00	1.5997e-02	6.4300e-02	7.8121e-02	1.6747e-01
0.95	0.0000e+00	7.2447e-03	4.2460e-02	6.3459e-02	1.1111e-01

References

[1] Zhu, X. & Stephens, M. Large-scale genome-wide enrichment analyses identify new trait-associated genes and pathways across 31 human phenotypes. *Nat. Commun.* **9**, 4361 (2018).

[2] Zhu, X. & Stephens, M. Bayesian large-scale multiple regression with summary statistics from genome-wide association studies. *Ann. Appl. Stat.* **11**, 1561–1592 (2017).

[3] Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014).

[4] Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).

[5] Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187–196 (2015).

[6] Ripke, S. *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).

[7] Liu, J. Z. *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* **47**, 979–986 (2015).

[8] Nikpay, M. *et al.* A comprehensive 1000 genomes–based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121–1130 (2015).

[9] Den Hoed, M. *et al.* Identification of heart rate-associated loci and their effects on cardiac conduction and rhythm disorders. *Nat. Genet.* **45**, 621–631 (2013).

[10] Lambert, J.-C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452–1458 (2013).

[11] Okbay, A. *et al.* Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.* **48**, 624–633 (2016).

[12] Okada, Y. *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376–381 (2014).

[13] Morris, A. P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44**, 981–990 (2012).

[14] Teslovich, T. M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).

[15] Christophersen, I. E. *et al.* Large-scale analyses of common and rare variants identify 12 new loci associated with atrial fibrillation. *Nat. Genet.* **49**, 946 (2017).

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

[17] Duren, Z., Chen, X., Jiang, R., Wang, Y. & Wong, W. H. Modeling gene regulation from paired expression and chromatin accessibility data. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E4914–E4923 (2017).

[18] Luo, Y. *et al.* New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Res.* **48**, D882–D889 (2020).

[19] Sachs, M. C. plotROC: a tool for plotting ROC curves. *J. Stat. Softw.* **79**, 1–19 (2017).

[20] Gazal, S. *et al.* Linkage disequilibrium–dependent architecture of human complex traits shows action of negative selection. *Nat. Genet.* **49**, 1421–1427 (2017).

[21] Lamparter, D., Marbach, D., Rueedi, R., Kutalik, Z. & Bergmann, S. Fast and rigorous computation of gene and pathway scores from SNP-based summary statistics. *PLoS Comput. Biol.* **12**, e1004714 (2016).

[22] Saito, T. & Rehmsmeier, M. precrec: fast and accurate precision-recall and ROC curve calculations in R. *Bioinformatics* **33**, 145–147 (2017).

[23] 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

[24] Marbach, D. *et al.* Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases. *Nat. Methods* **13**, 366–370 (2016).

[25] Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).

[26] The FANTOM Consortium and the RIKEN PMI and CLST (DGT). A promoter-level mammalian expression atlas. *Nature* **507**, 462–470 (2014).

[27] Bult, C. J., Blake, J. A., Smith, C. L., Kadin, J. A. & Richardson, J. E. Mouse genome database (MGD) 2019. *Nucleic Acids Res.* **47**, D801–D806 (2019).

[28] Freund, M. K. *et al.* Phenotype-specific enrichment of Mendelian disorder genes near GWAS regions across 62 complex traits. *Am. J. Hum. Genet.* **103**, 535–552 (2018).

[29] Amberger, J., Bocchini, C., Scott, A. & Hamosh, A. OMIM.org: leveraging knowledge across phenotype-gene relationships. *Nucleic Acids Res.* **47**, D1038–D1043 (2019).

[30] Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).

[31] Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).

[32] Global Lipids Genetics Consortium. Discovery and refinement of loci associated with lipids levels. *Nat. Genet.* **45**, 1274–1283 (2013).

[33] Wang, Y. *et al.* Therapeutic target database 2020: enriched resource for facilitating research and early development of targeted therapeutics. *Nucleic Acids Res.* **48**, D1031–D1041 (2020).