The supplementary figures and text of "IGREX for quantifying the impact of genetically regulated expression on phenotypes"

Mingxuan Cai[1], Lin S. Chen[2], Jin Liu[3*]& Can Yang[1*]

[1]Department of Mathematics, The Hong Kong University of Science and Technology

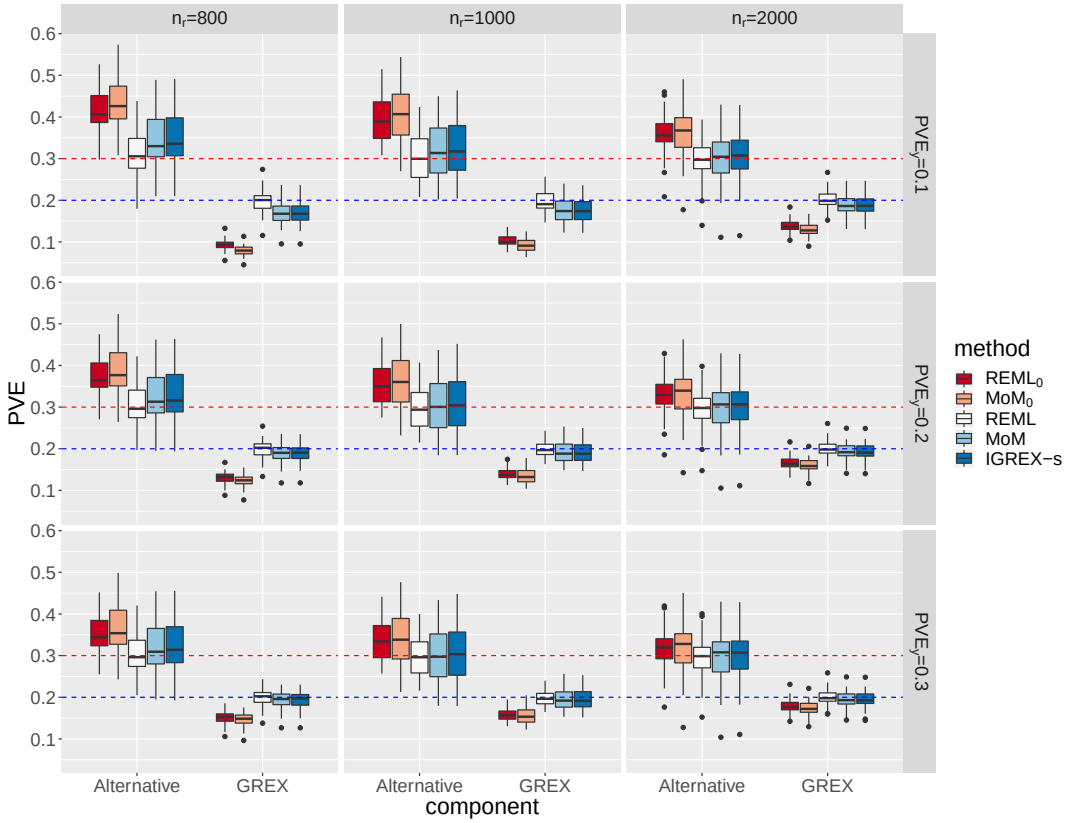
[2]Department of Public Health Sciences, The University of Chicago

[3]Center for Quantitative Medicine, Duke-NUS Medical School

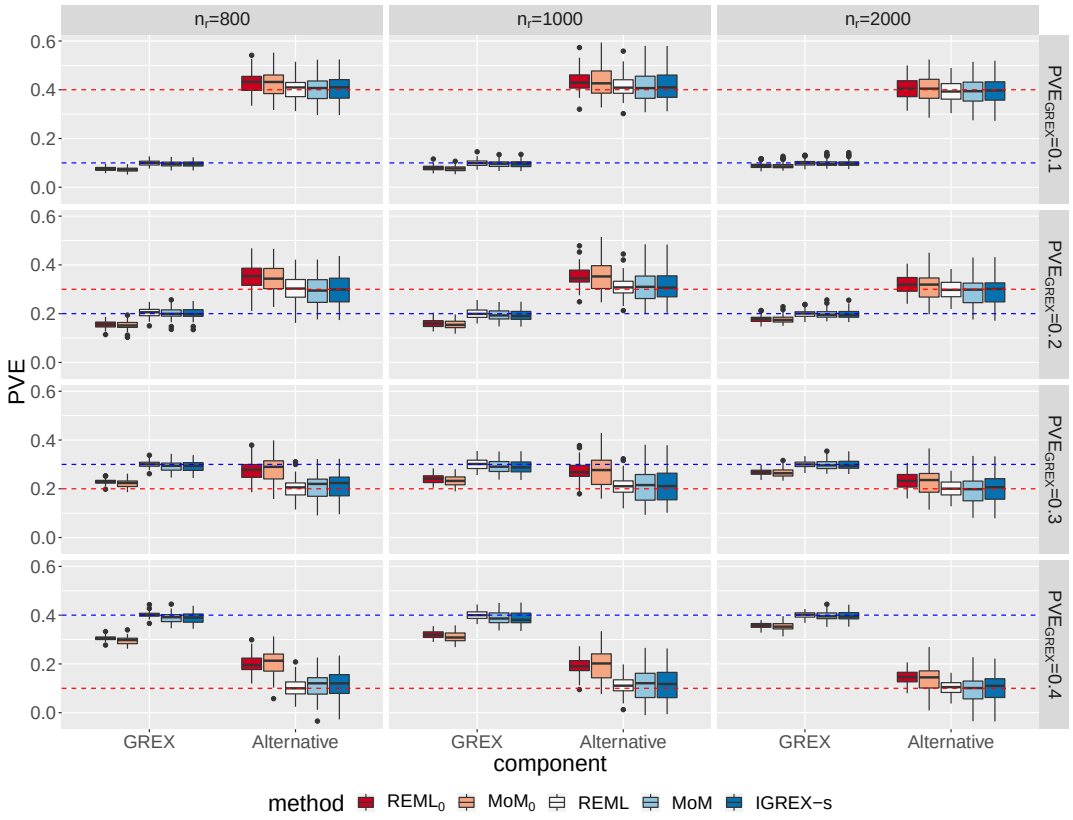
Contents

*Correspondence should be addressed to Jin Liu (jin.liu@duke-nus.edu.sg) and Can Yang (macyang@ust.hk)

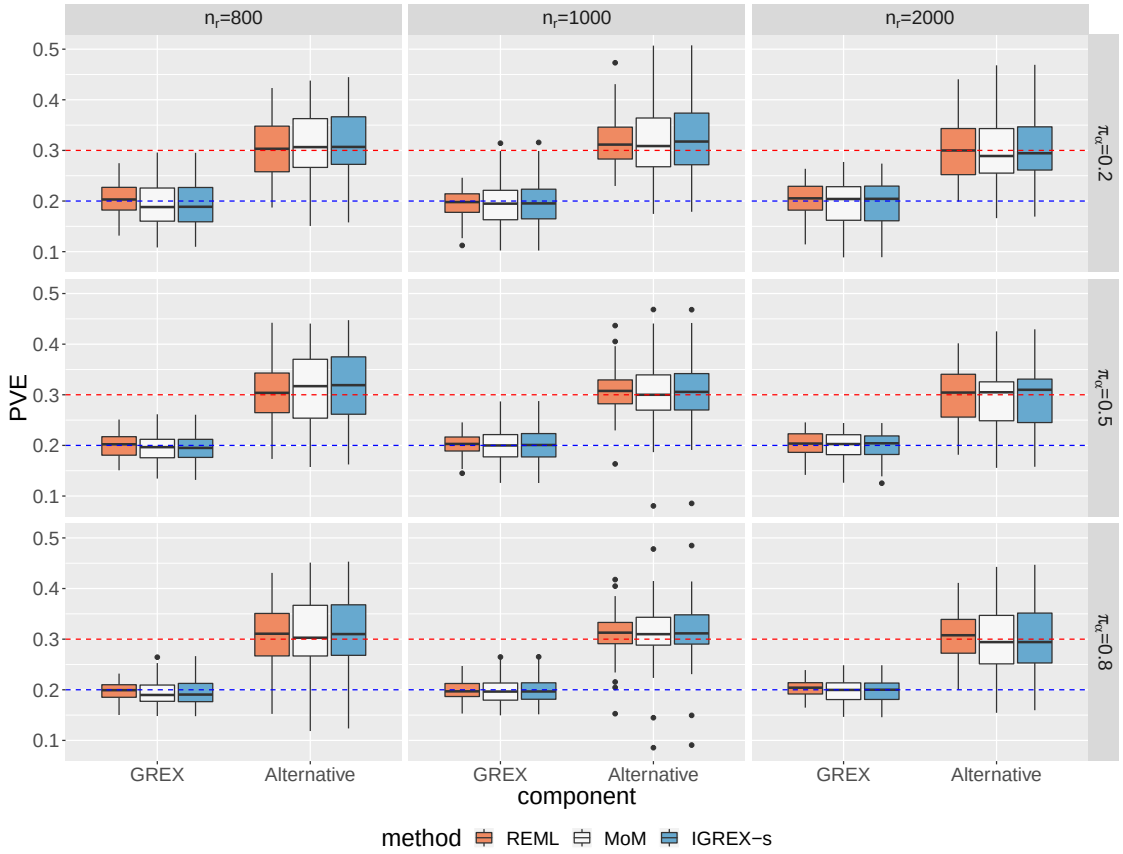
1 Supplementary Figures



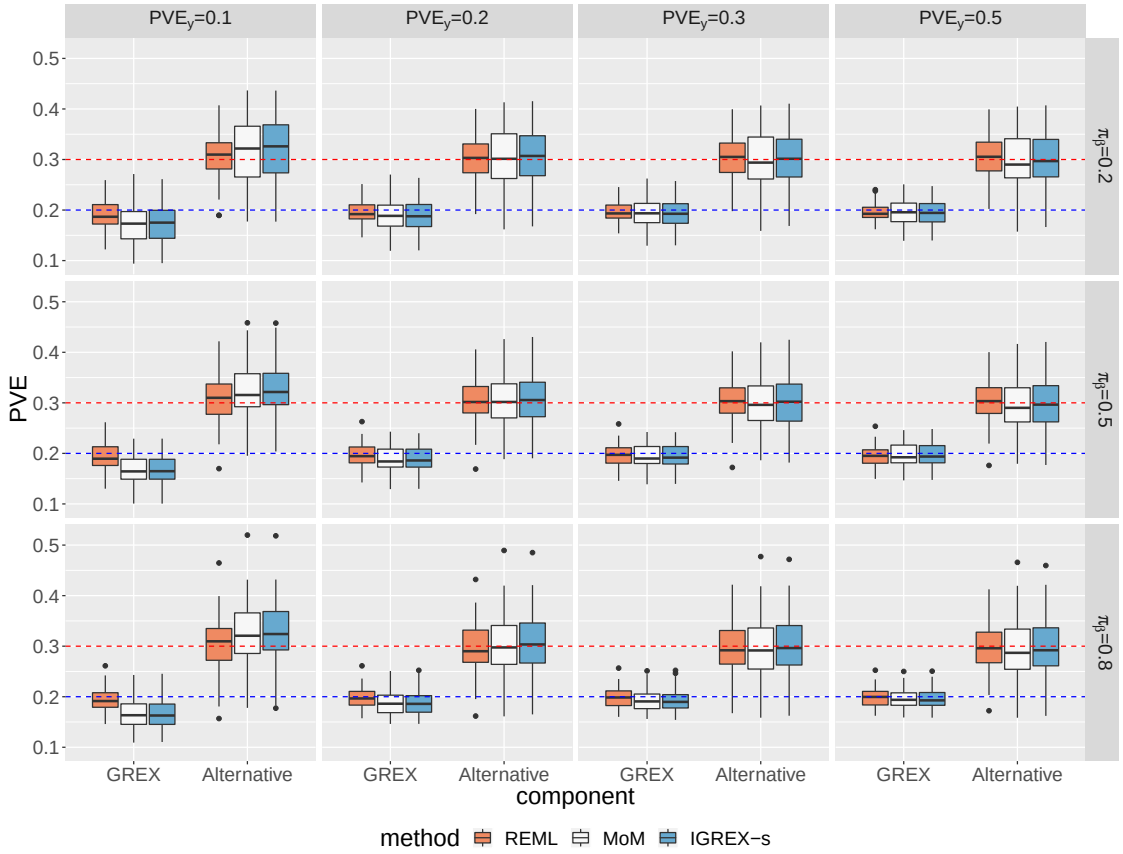
Supplementary Figure S1: Estimates of PVE_{GREX} and $\text{PVE}_{\text{Alternative}}$ for REML_0, MoM_0, IGREX-REML, IGREX-MoM and IGREX-s with $n = 4000$, $h_t^2 = \text{PVE}_{\text{GREX}} + \text{PVE}_{\text{Alternative}} = 0.5$ and $\text{PVE}_y = 0.3$. n_r is varied at $\{800, 1000, 2000\}$ and PVE_y is varied at $\{0.1, 0.2, 0.3\}$. The blue dashed line and red dashed line represent true values of PVE_{GREX} and $\text{PVE}_{\text{Alternative}}$, respectively. The results are summarized from 50 replications.



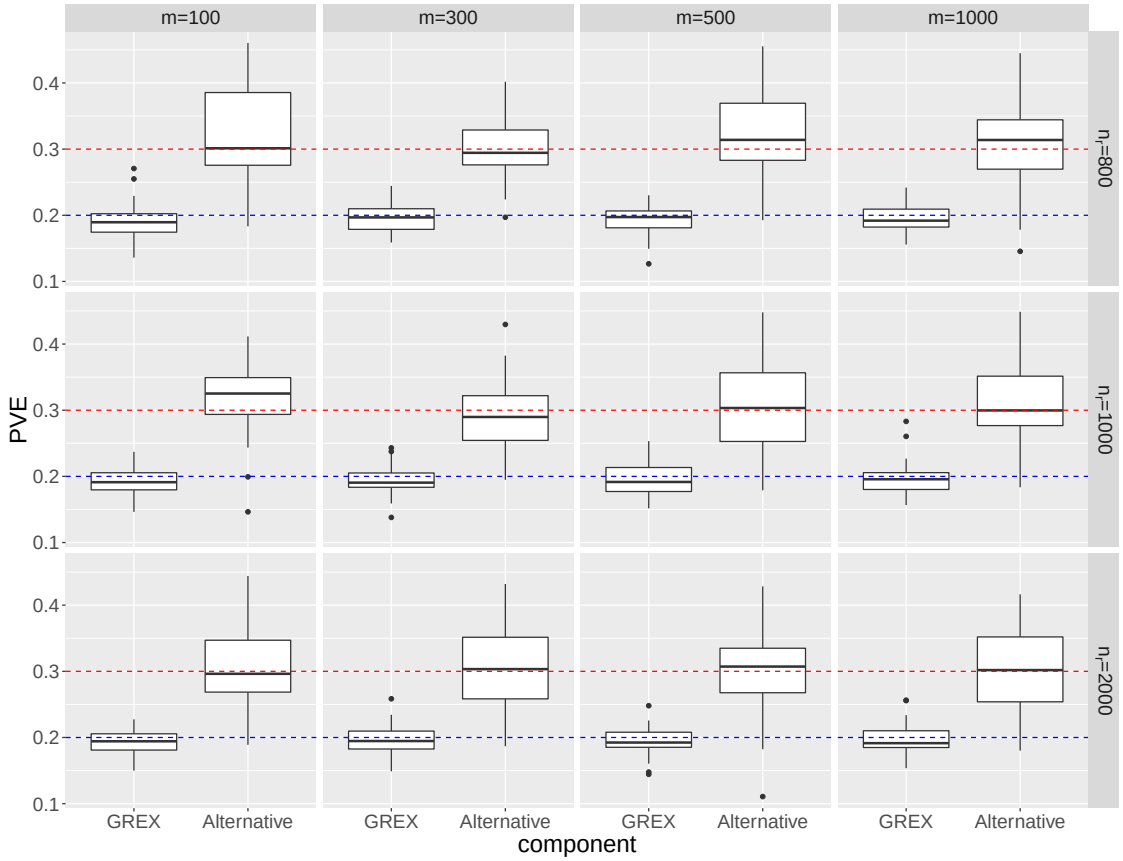
Supplementary Figure S2: Estimates of PVE_{GREX} and $\text{PVE}_{\text{Alternative}}$ for REML_0, MoM_0, IGREX-REML, IGREX-MoM and IGREX-s with $n = 4000$, $h_t^2 = \text{PVE}_{\text{GREX}} + \text{PVE}_{\text{Alternative}} = 0.5$ and $\text{PVE}_y = 0.3$. n_r is varied at $\{800, 1000, 2000\}$ and PVE_{GREX} is varied at $\{0.1, 0.2, 0.3, 0.4\}$. The blue dashed line and red dashed line represent true values of PVE_{GREX} and $\text{PVE}_{\text{Alternative}}$, respectively. The results are summarized from 50 replications.



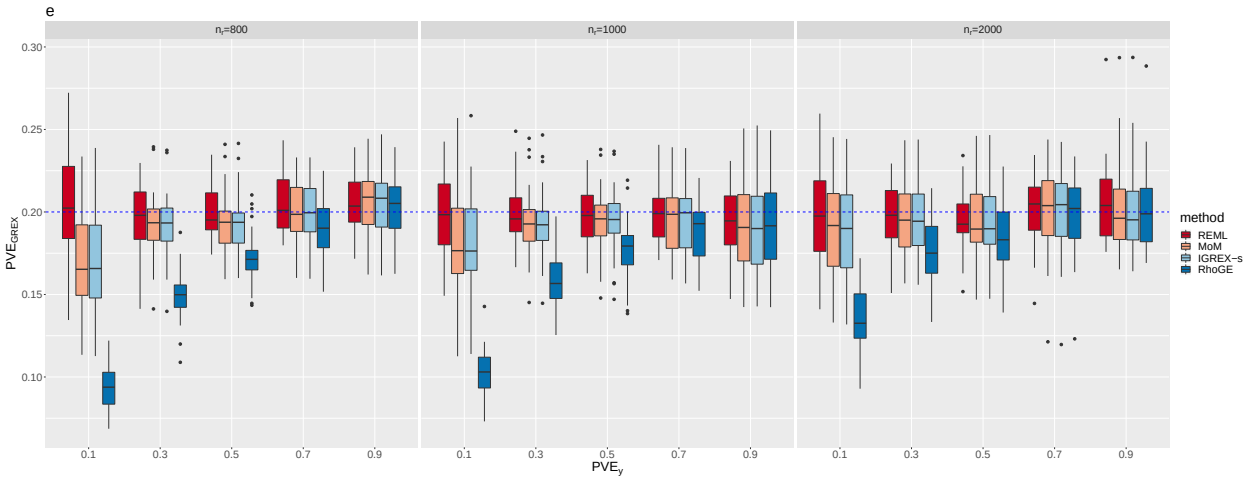
Supplementary Figure S3: Estimates of PVE_{GREX} and $\text{PVE}_{\text{Alternative}}$ for IGREX-REML, IGREX-MoM and IGREX-s with $n = 4000$, $\text{PVE}_{\text{GREX}} = 0.2$, $\text{PVE}_{\text{Alternative}} = 0.3$, $\text{PVE}_y = 0.3$ and $\pi_\beta = 0.2$. n_r is varied at $\{800, 1000, 2000\}$ and π_α is varied at $\{0.2, 0.5, 0.8\}$. The blue dashed line and red dashed line represent true values of PVE_{GREX} and $\text{PVE}_{\text{Alternative}}$, respectively. The results are summarized from 50 replications.



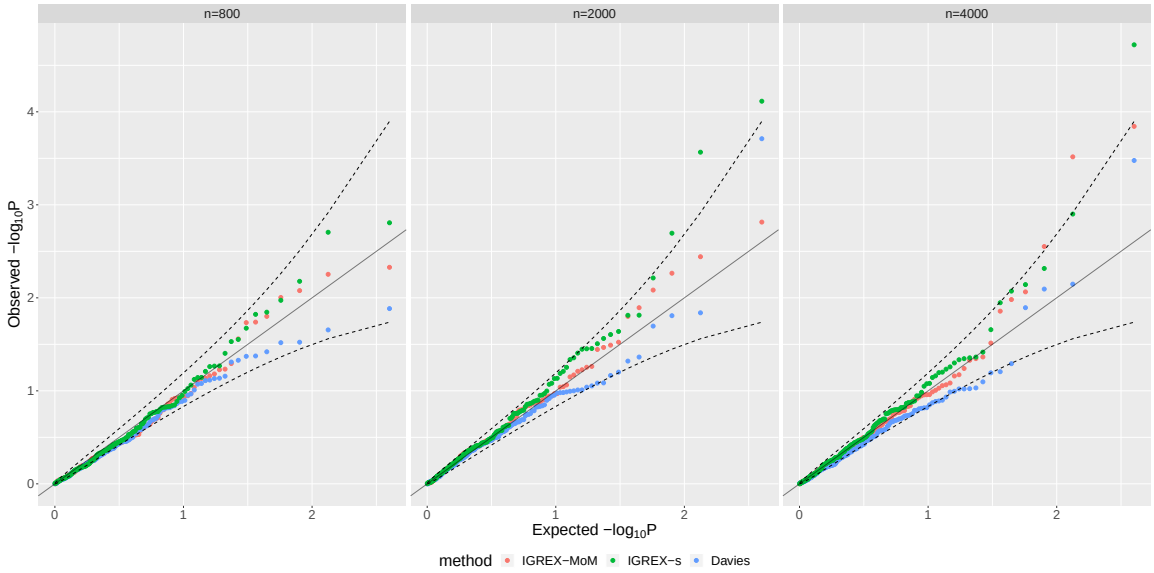
Supplementary Figure S4: Estimates of PVE_{GREX} and $\text{PVE}_{\text{Alternative}}$ for IGREX-REML, IGREX-MoM and IGREX-s with $n = 4000$, $n_r = 800$, $\text{PVE}_{\text{GREX}} = 0.2$, $\text{PVE}_{\text{Alternative}} = 0.3$ and $\pi_\beta = 0.2$. PVE_y is varied at $\{0.1, 0.2, 0.3, 0.5\}$ and π_β is varied at $\{0.2, 0.5, 0.8\}$. The blue dashed line and red dashed line represent true values of PVE_{GREX} and $\text{PVE}_{\text{Alternative}}$, respectively. The results are summarized from 50 replications.



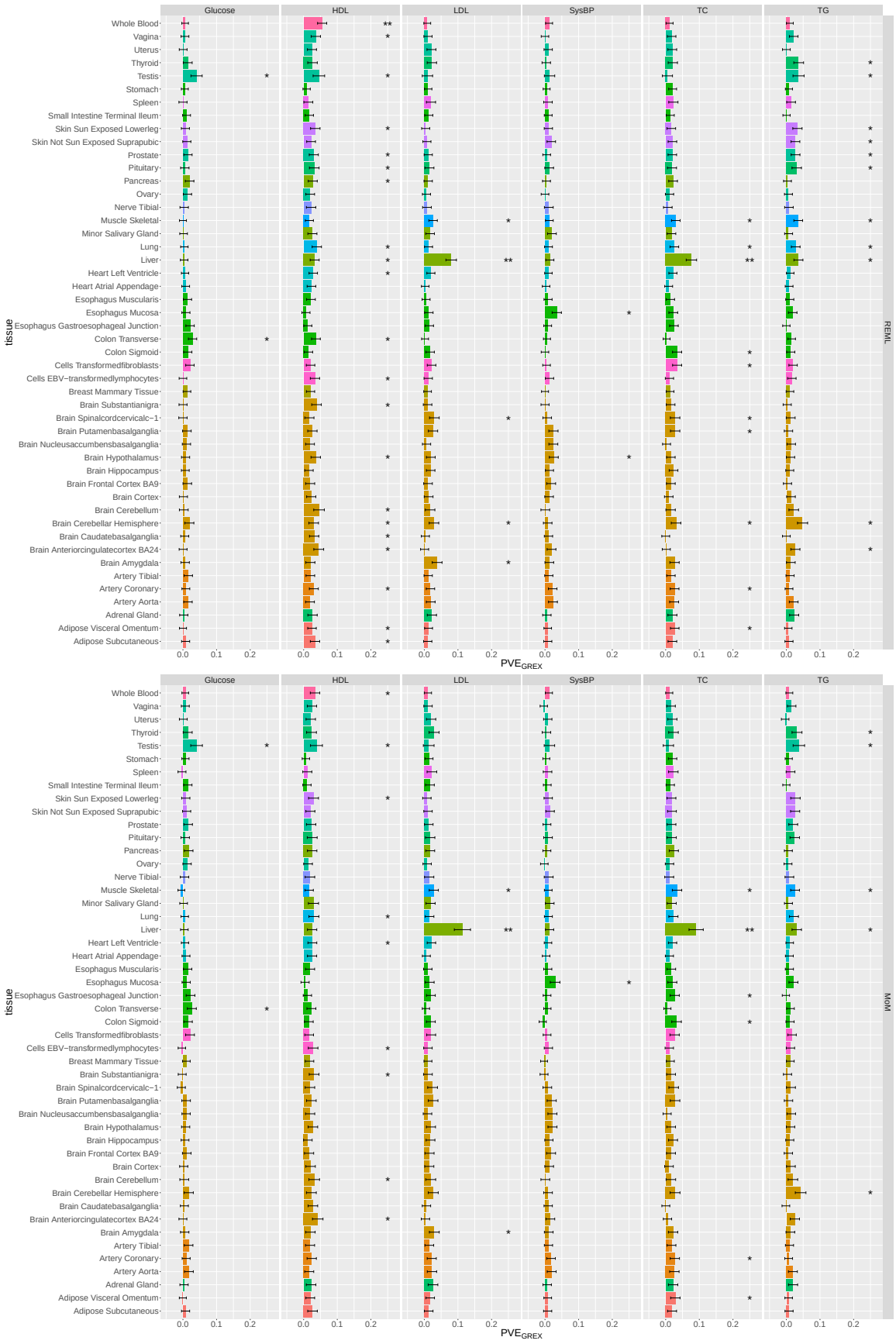
Supplementary Figure S5: Estimates of PVE_{GREX} and $\text{PVE}_{\text{Alternative}}$ for IGREX-s with $n = 4000$, $\text{PVE}_{\text{GREX}} = 0.2$, $\text{PVE}_{\text{Alternative}} = 0.3$ and $\text{PVE}_y = 0.3$. n_r is varied at $\{800, 1000, 2000\}$ and m is varied at $\{100, 300, 500, 1000\}$. The blue dashed line and red dashed line represent true values of PVE_{GREX} and $\text{PVE}_{\text{Alternative}}$, respectively. The results are summarized from 50 replications.



Supplementary Figure S6: Simulation studies to compare estimation accuracies of IGREX with other methods. The blue and red dashed lines represent the true values of PVE_{GREX} and $\text{PVE}_{\text{Alternative}}$, respectively. We averaged the results over 30 replications and generated box plots for evaluating the estimation performance ofthe three models of IGREX and RhoGE when n_r was varied at $\{800, 1000, 2000\}$.

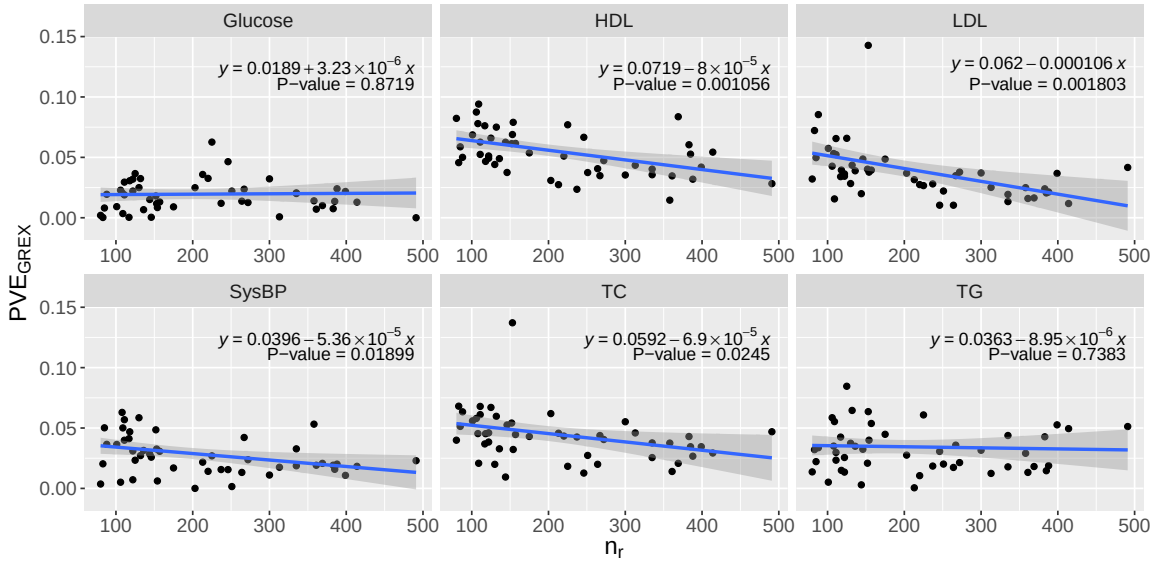


Supplementary Figure S7: The QQ plots of p-values obtained by Davies method and normal tests for IGREX-MoM and IGREX-s with n varied at $\{800, 2000, 4000\}$. Between the dashed lines are the 95% confidence intervals under the null hypothesis. Each sub-plot is generated by 200 replications.

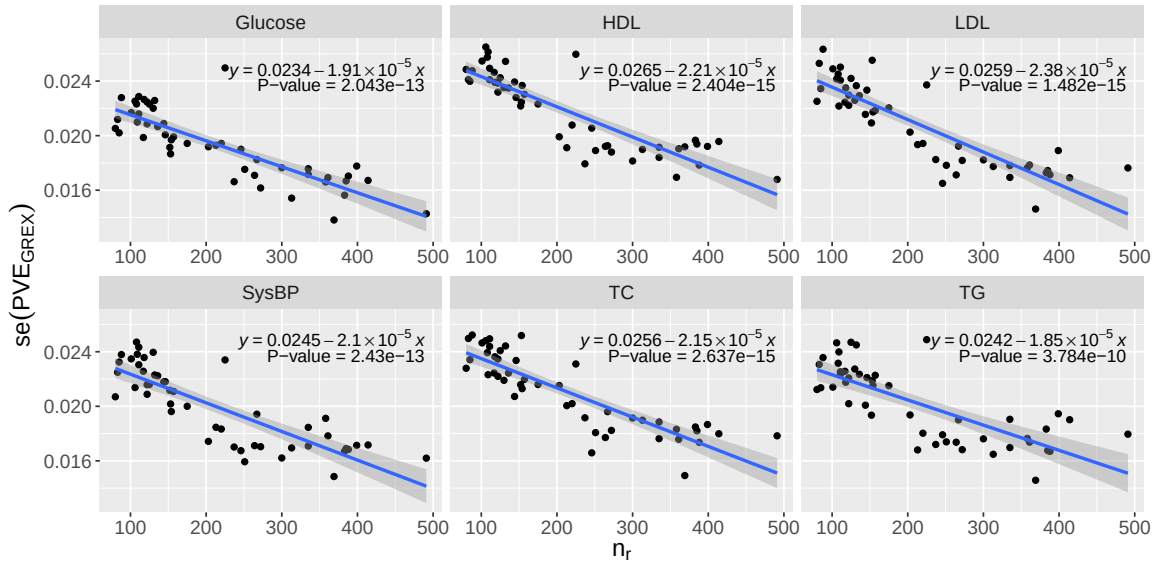


Supplementary Figure S8: REML and MoM estimates of PVE_{GREX} for NFBC data without accounting for uncertainty.

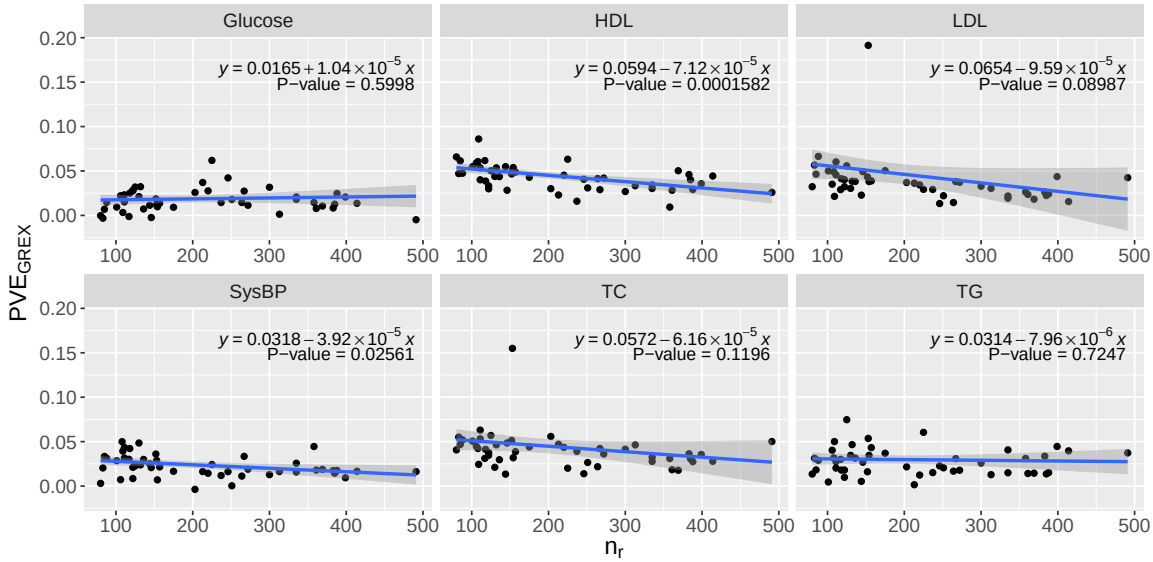
Supplementary Figure S9: REML and MoM estimates of $\text{PVE}_{\text{GREX}}/h_t^2$ for NFBC data.



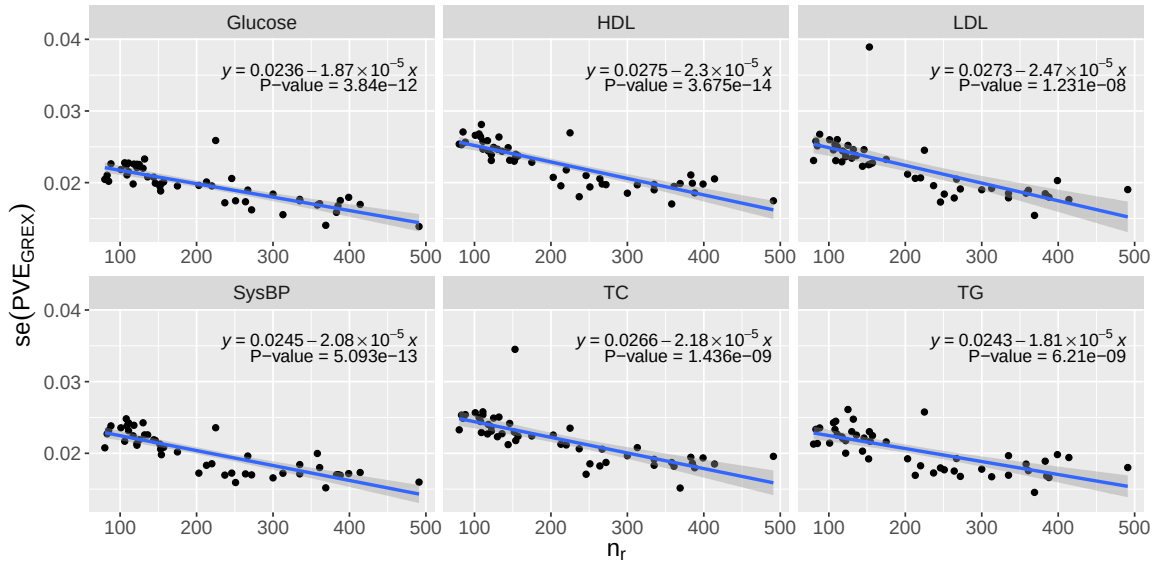
Supplementary Figure S10: $\widehat{\text{PVE}}_{\text{GREX}}$ obtained by REML against n_r in the NFBC dataset.



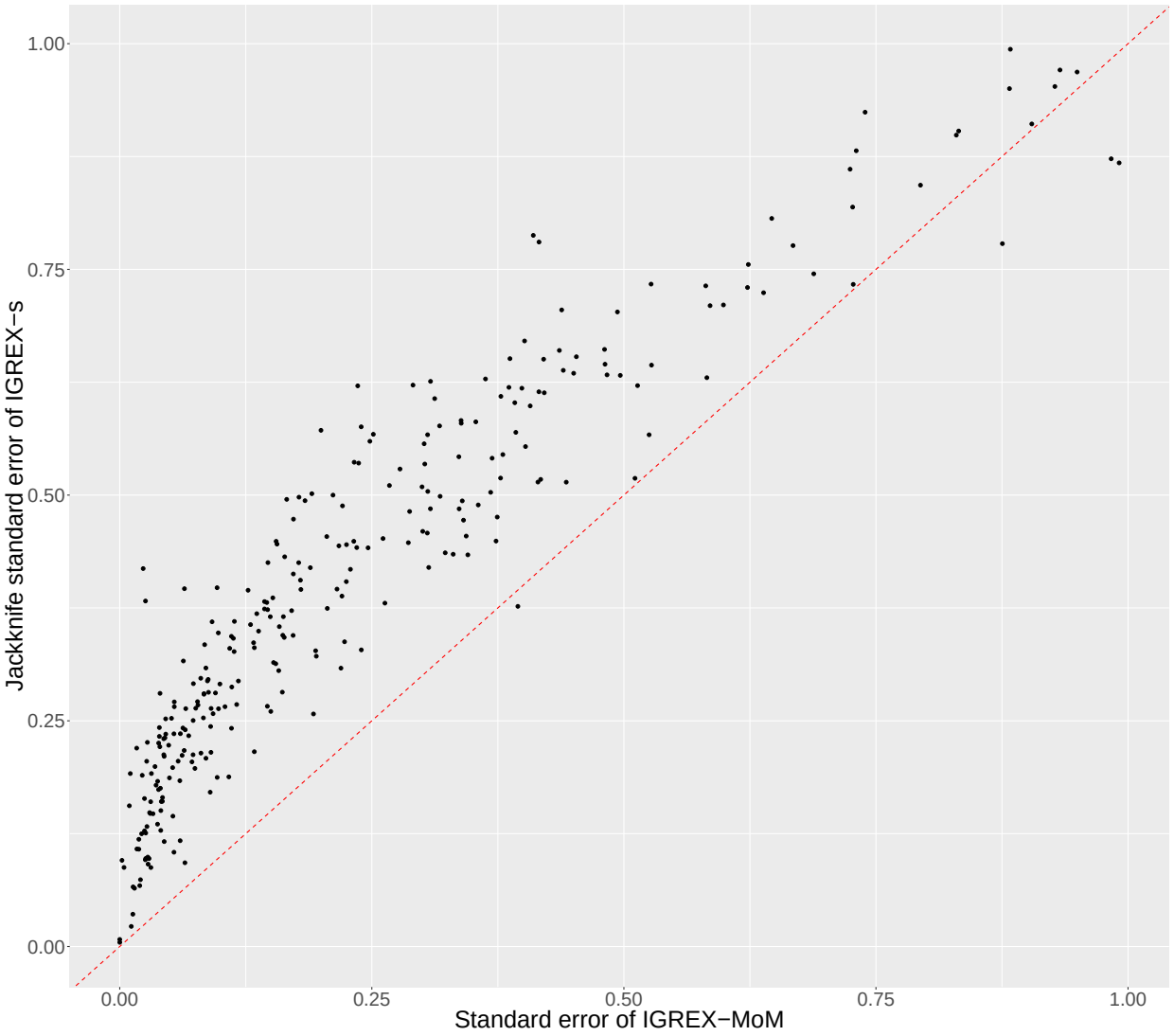
Supplementary Figure S11: Standard errors of $\widehat{\text{PVE}}_{\text{GREX}}$ obtained by REML against n_r in the NFBC dataset.



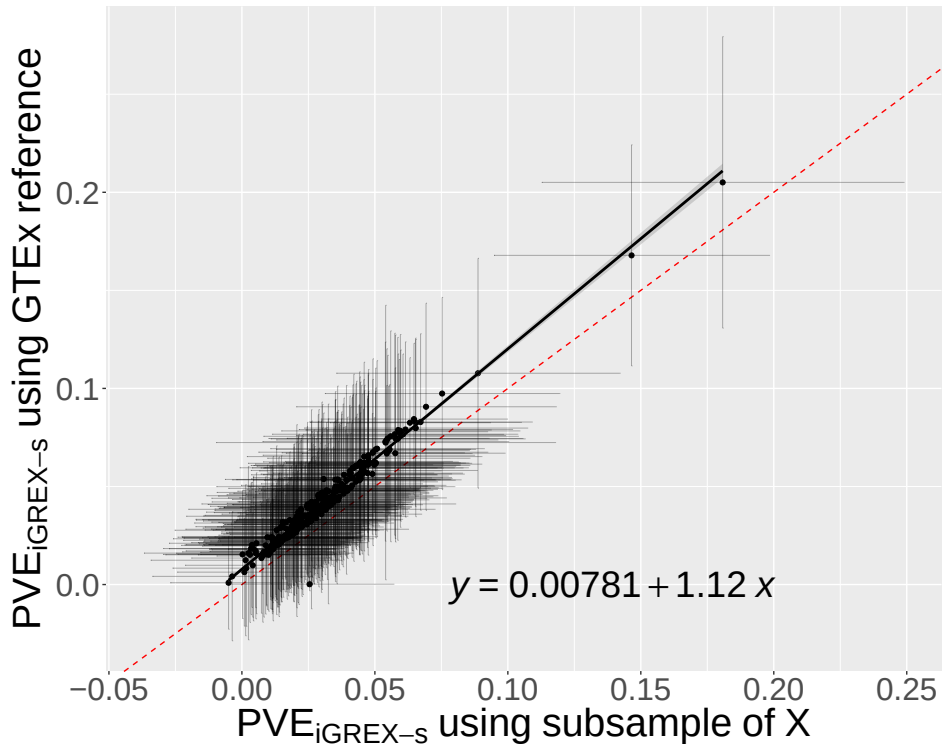
Supplementary Figure S12: $\widehat{\text{PVE}}_{\text{GREX}}$ obtained by MoM against n_r in the NFBC dataset.



Supplementary Figure S13: Standard errors of $\widehat{\text{PVE}}_{\text{GREX}}$ obtained by MoM against n_r in the NFBC dataset.

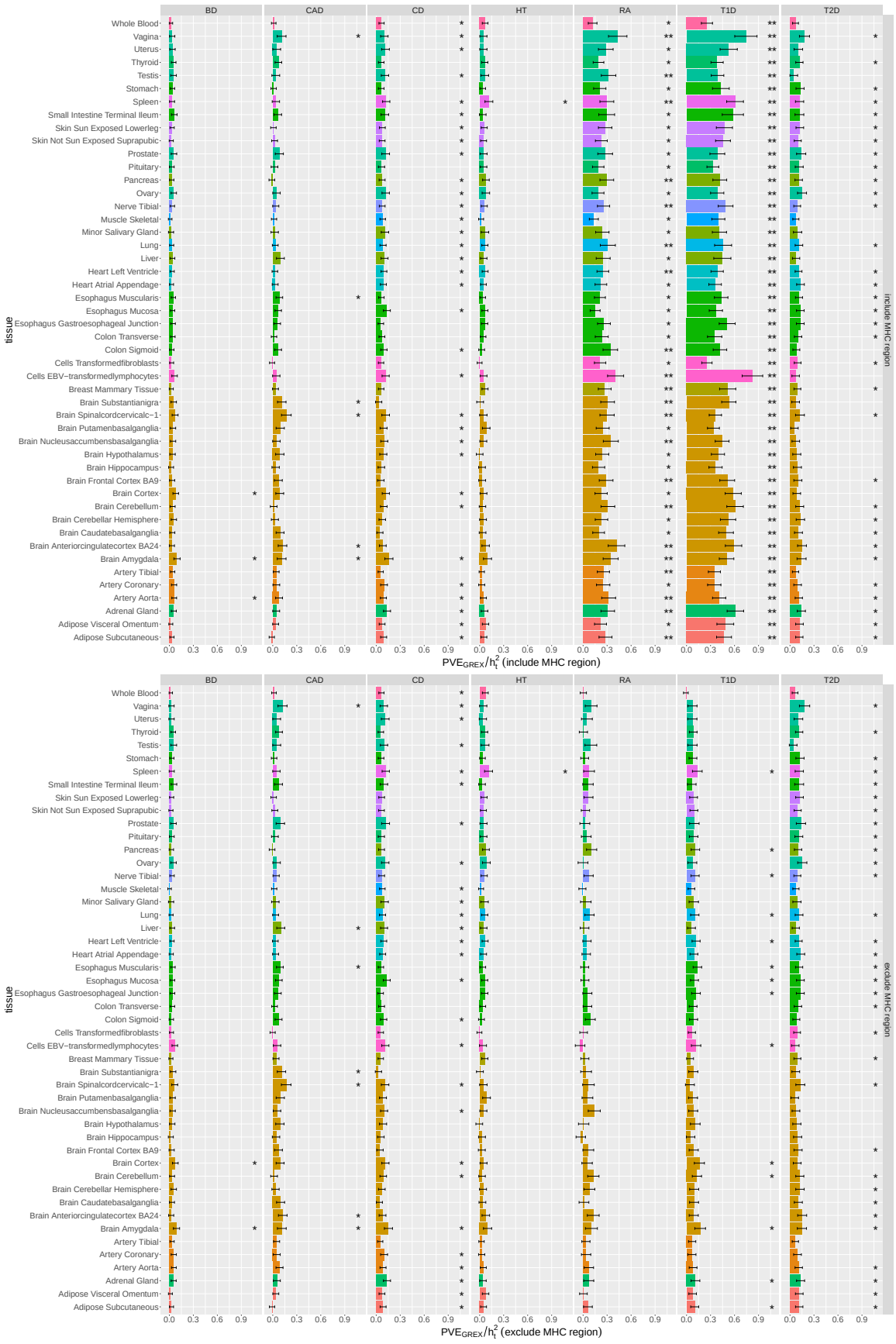


Supplementary Figure S14: Standard errors of PVE_{GREX} for NFBC data estimated using block Jackknife and sandwich estimator.

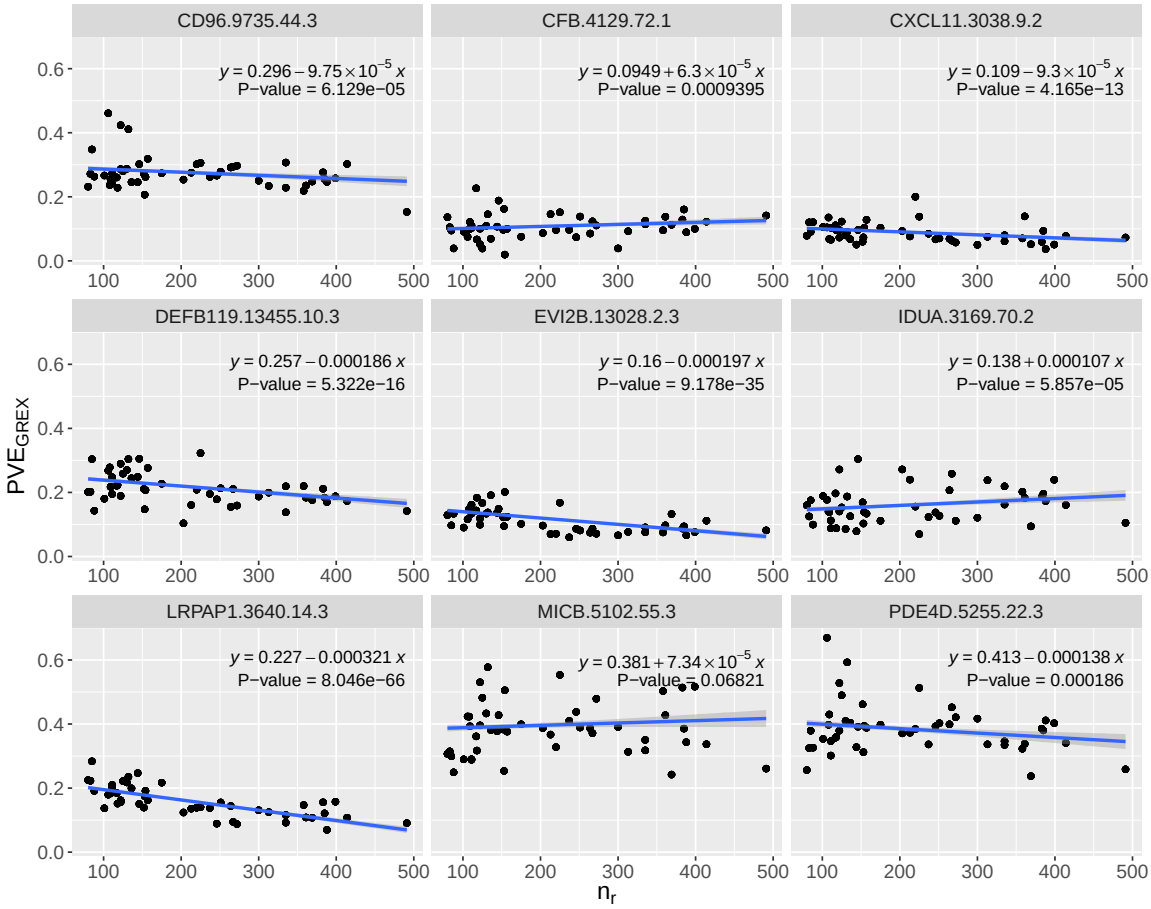


Supplementary Figure S15: Comparison of $\widehat{\text{PVE}}_{\text{GREX}}$ obtained by IGREX-s using subsample of NFBC individuals against those using GTEx reference. The red dashed line represents the identity line. The $\widehat{\text{PVE}}_{\text{GREX}}$ obtained by using GTEx reference panel is slightly larger than those obtained by the subsampling. After considering the standard errors of the estimates, however, the results given by the GTEx samples and the subsamples of original population are quite consistent with each other. Therefore, we expect that GTEx reference panel can produce satisfactory IGREX-s approximation in practice.

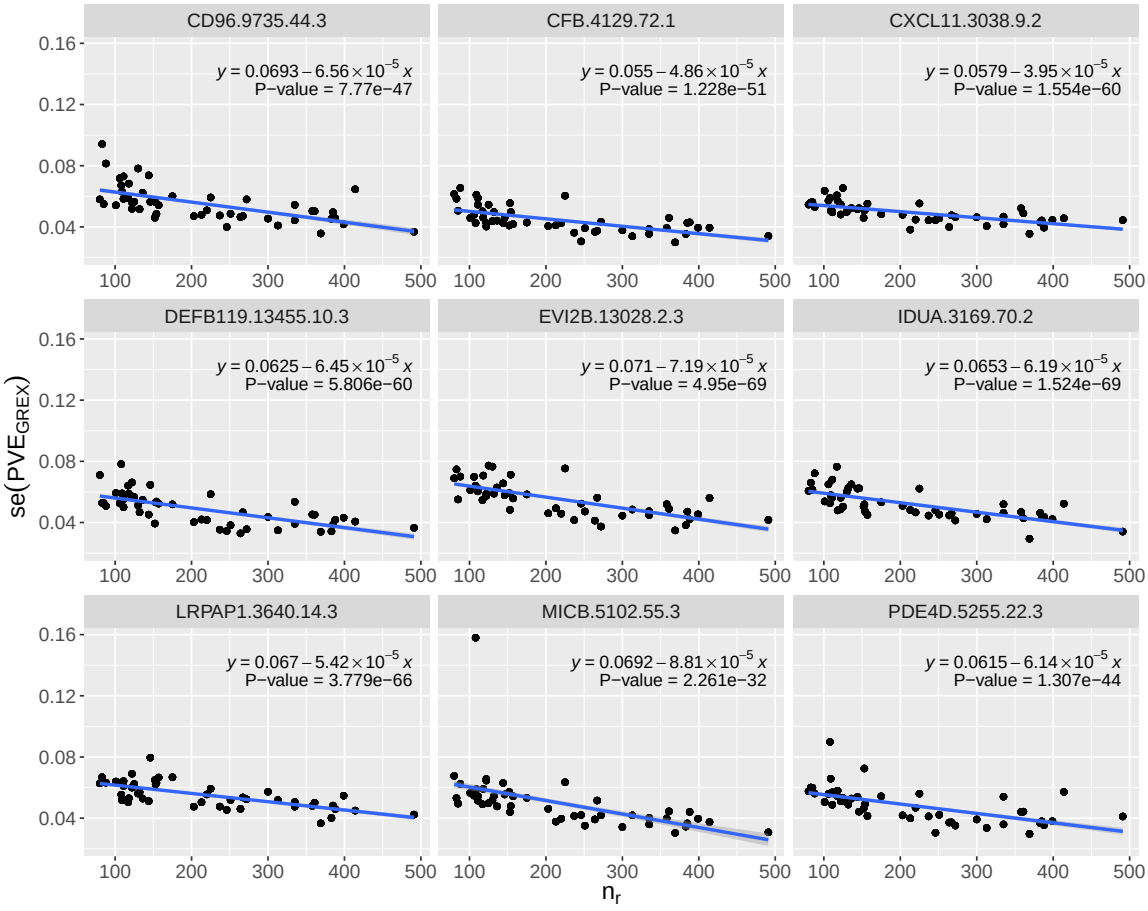
Supplementary Figure S16: REML estimates of $\text{PVE}_{\text{GREX}}/h_t^2$ for WTCCC data.



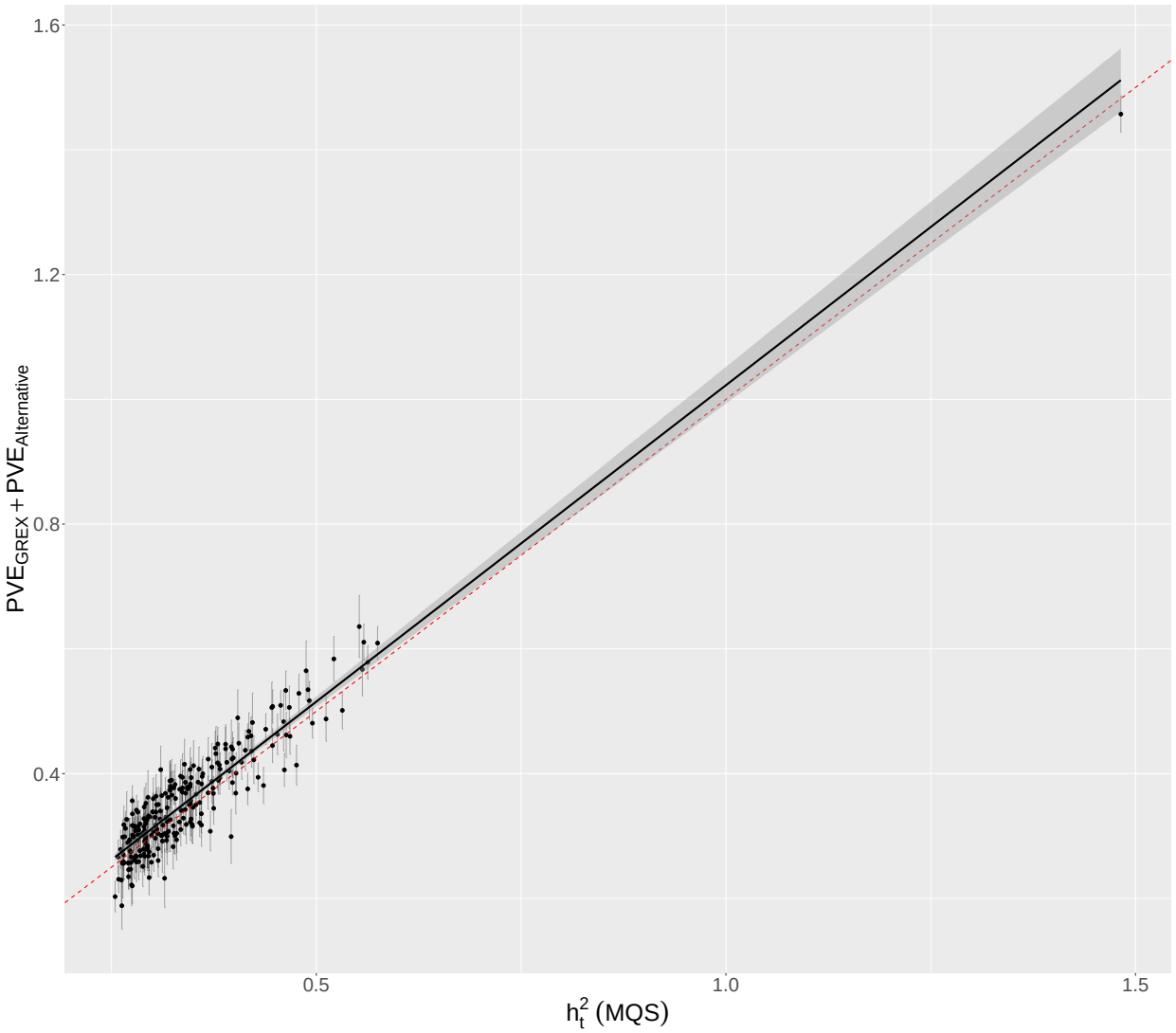
Supplementary Figure S17: MoM estimates of $\text{PVE}_{\text{GREX}}/h_t^2$ for WTCCC data.



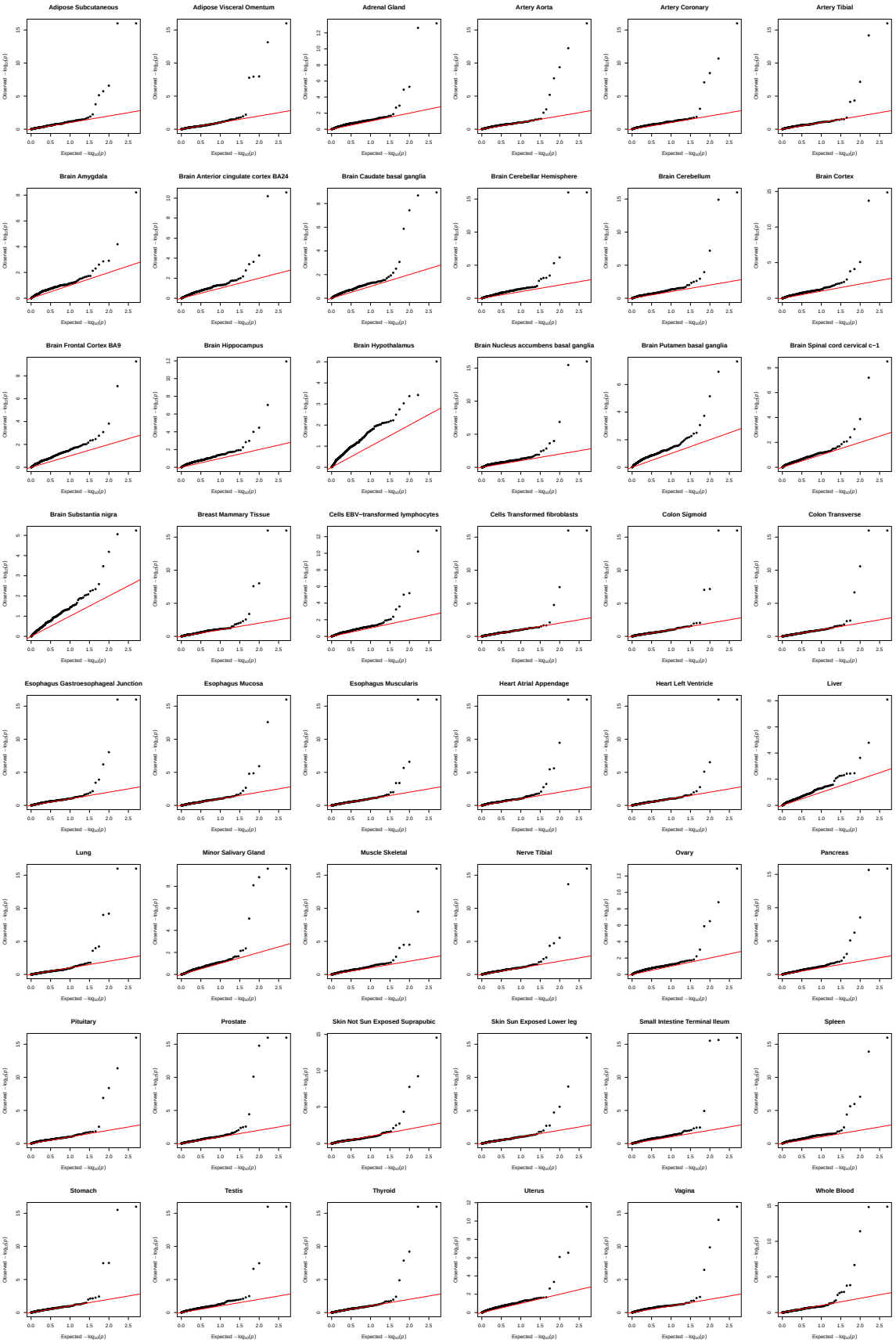
Supplementary Figure S18: $\widehat{\text{PVE}}_{\text{GREX}}$ obtained by IGREX-s against n_r in the pQTL dataset.



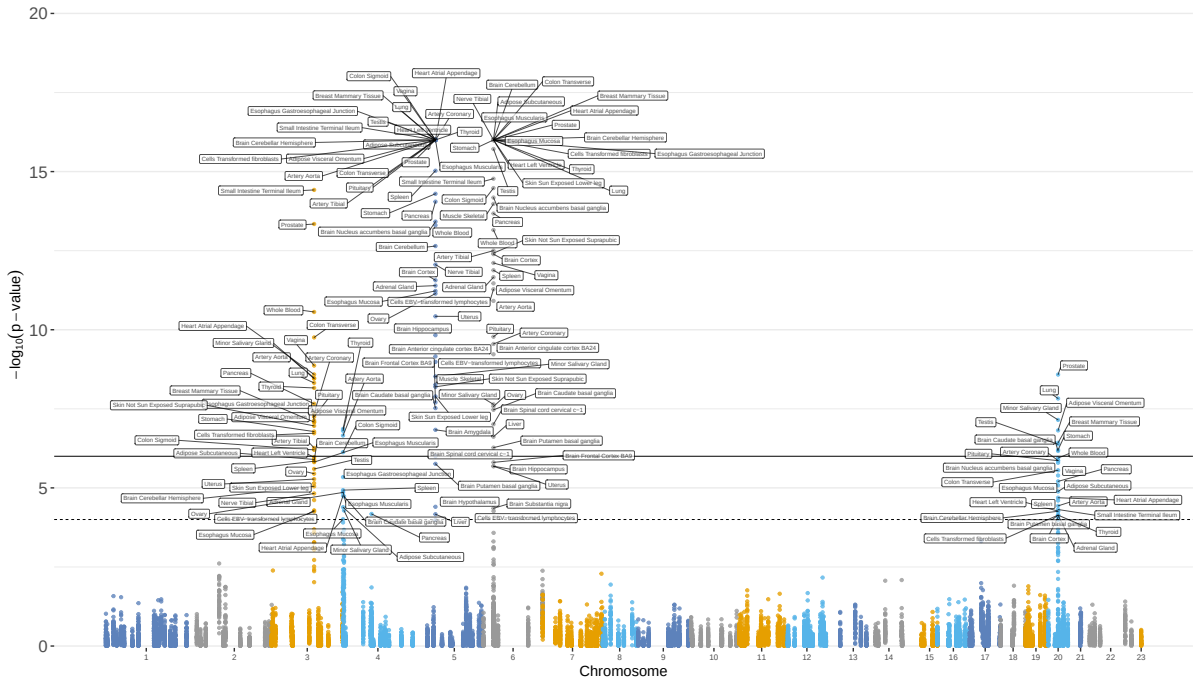
Supplementary Figure S19: Standard errors of $\widehat{\text{PVE}}_{\text{GREX}}$ obtained by IGREX-s against n_r in the pQTL dataset.



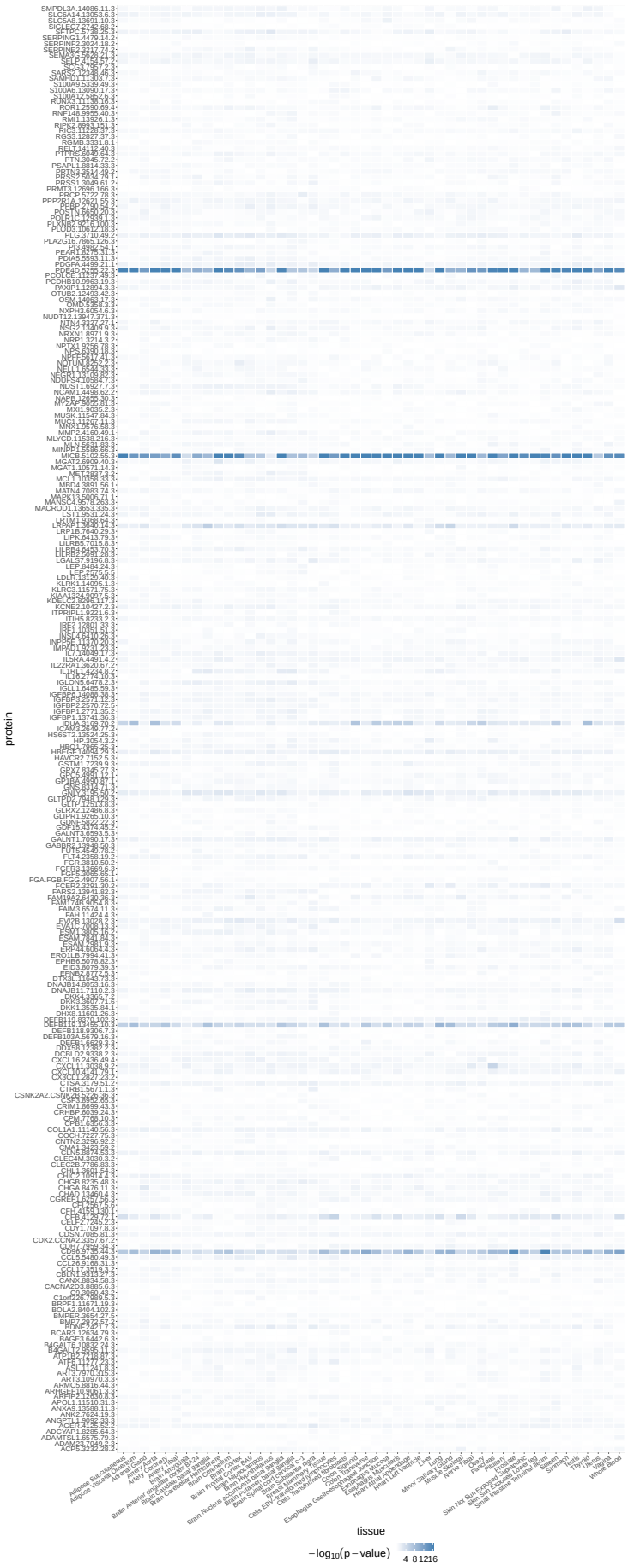
Supplementary Figure S20: Comparison of $\mathrm{PVE}_{\mathrm{GREX}} + \mathrm{PVE}_{\mathrm{Alternative}}$ estimated by IGREX and h_t^2 estimated by [1] in pQTL dataset.



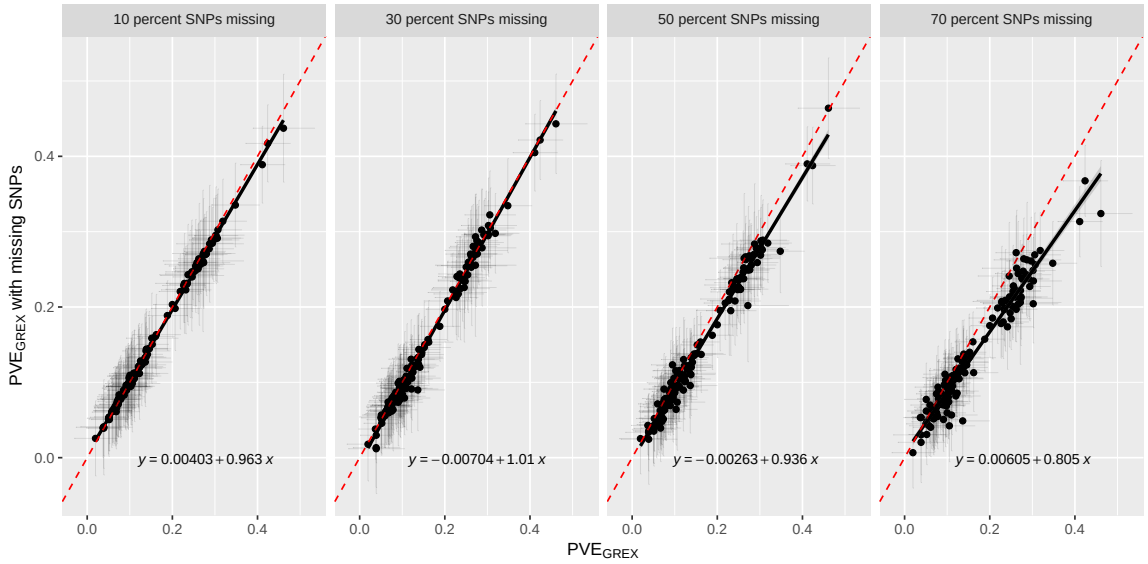
Supplementary Figure S21: QQ plot of p-values of PVE_{GREX} for proteins in each of GTEx tissue.



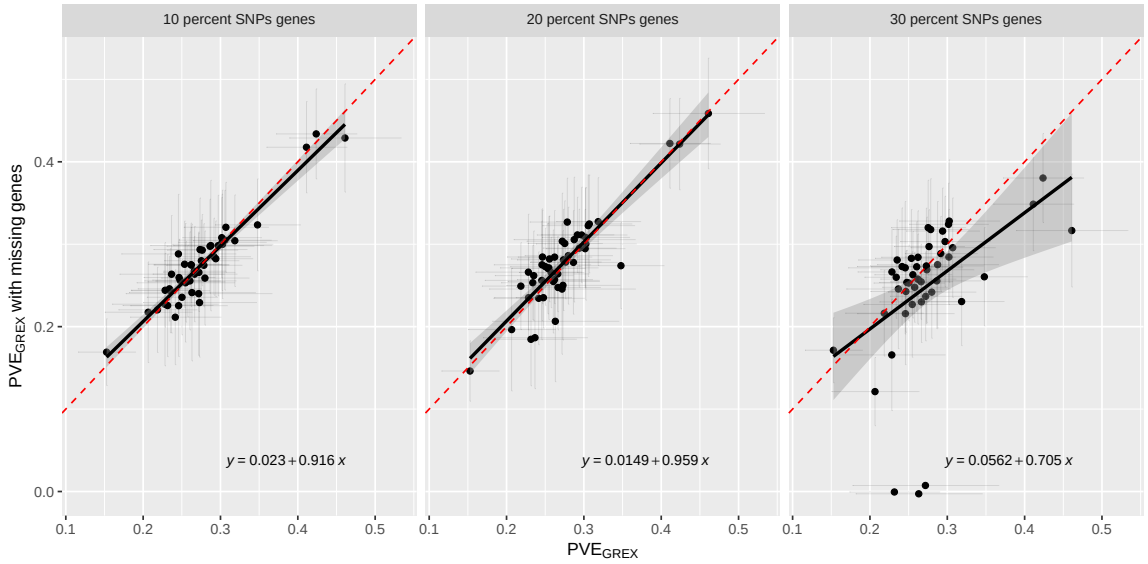
Supplementary Figure S22: Manhattan plot of protein-coding genes for pQTL study. p-values are computed for testing $H_0 : \mathrm{PVE}_{\mathrm{GREX}} = 0$



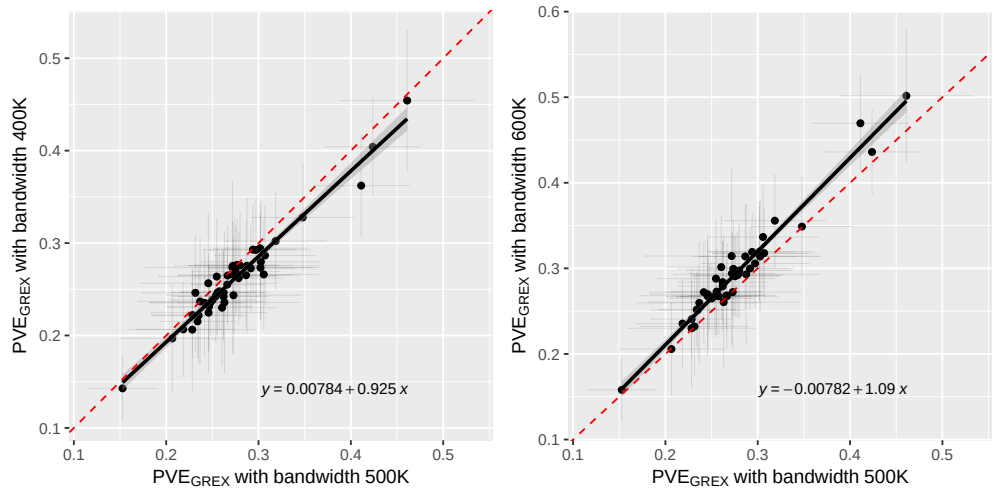
Supplementary Figure S23: Heatmap of p-values of PVE_{GREX} for 249 proteins.



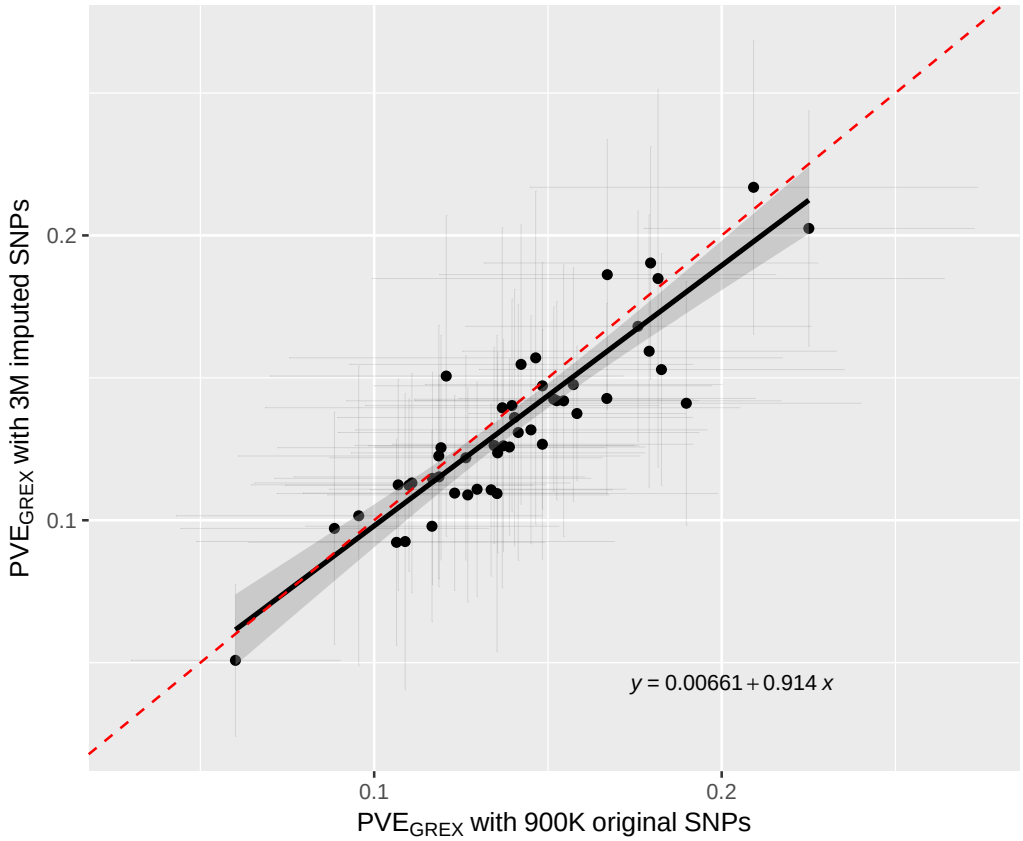
Supplementary Figure S24: Sensitivity analysis for mismatch between GTEx reference and GWAS dataset. PVE_{GREX}'s are estimated for the 9 proteins across 48 tissues. The x-axis is the $\widehat{\text{PVE}}_{\text{GREX}}$ using all overlapped SNPs; the y-axis is the $\widehat{\text{PVE}}_{\text{GREX}}$ obtained by randomly removing SNPs in GTEx data. We can see that the slope remains close to one until 50% of SNPs are randomly removed, and the intercept always remain close to zero. The slope drops to 0.805 when 70% of SNPs are randomly removed. The results indicate that the estimate of IGREX is insensitive to mismatch of SNPs.



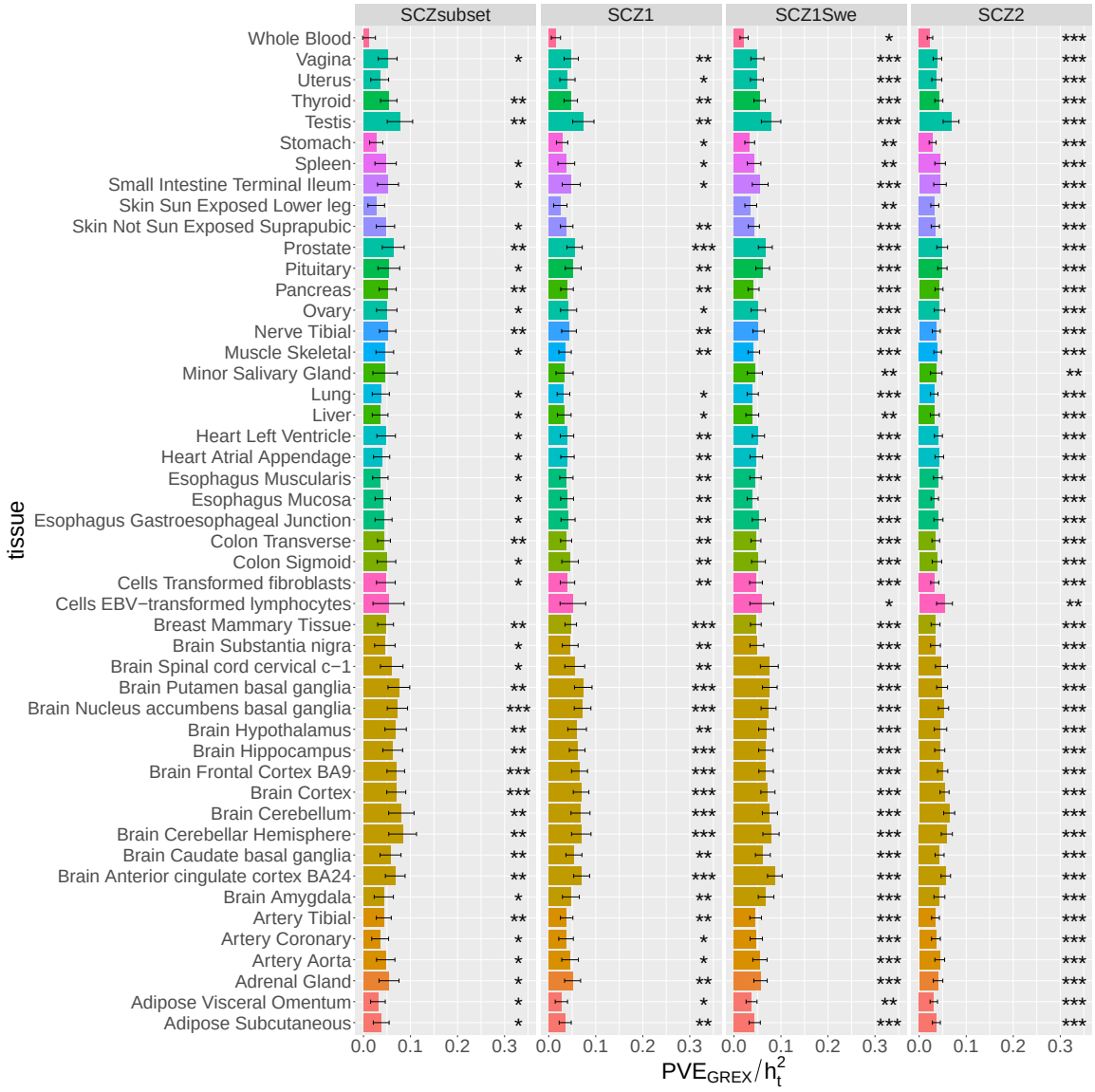
Supplementary Figure S25: Sensitivity analysis for mismatch between GTEx reference and GWAS dataset. PVE_{GREX}'s are estimated for protein CD96 across 48 tissues. The x-axis is the $\widehat{\text{PVE}}_{\text{GREX}}$ using all genes; the y-axis is the $\widehat{\text{PVE}}_{\text{GREX}}$ obtained by randomly removing genes in GTEx data. The results remain nearly the same until 30% of genes are randomly removed.



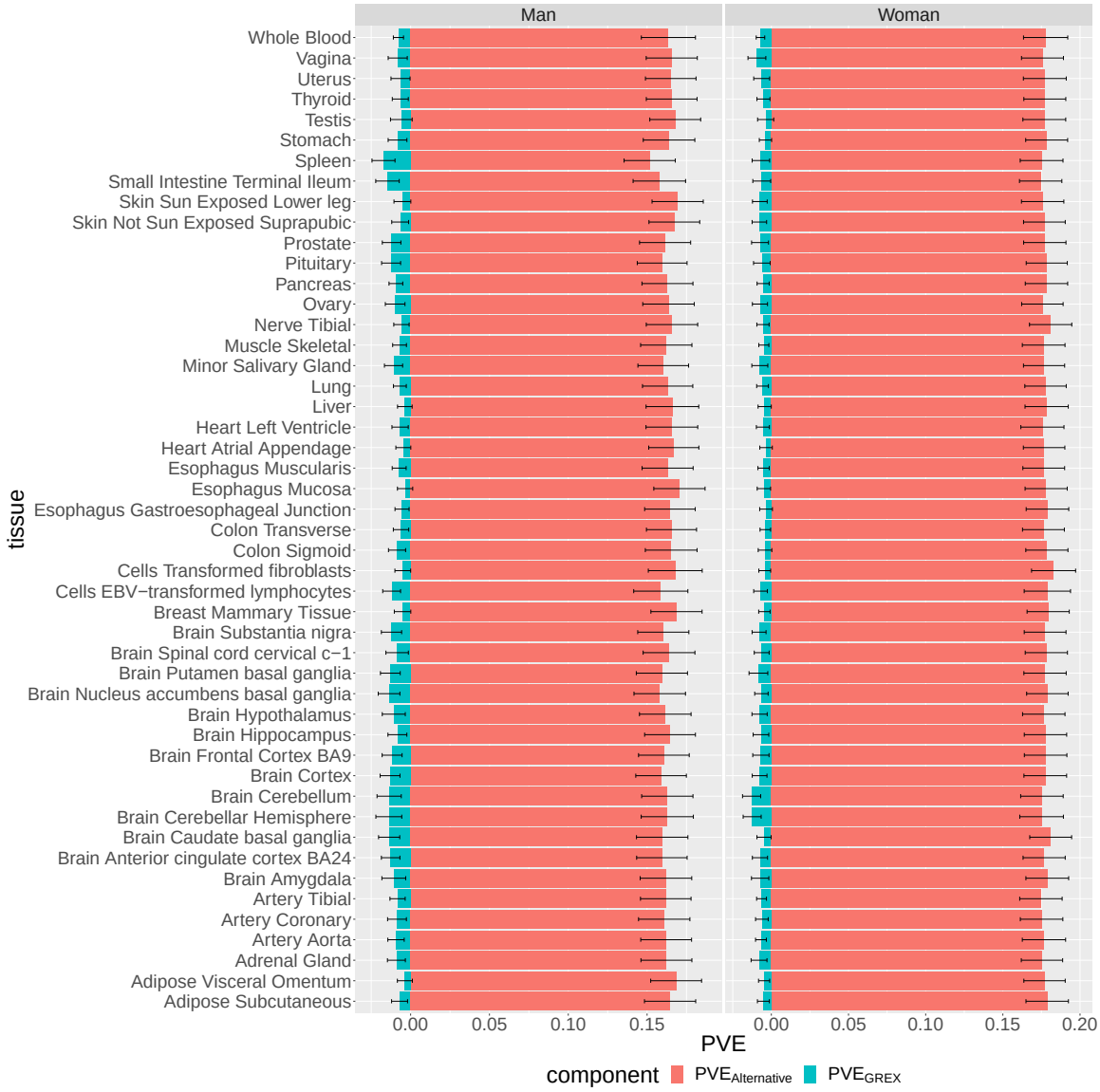
Supplementary Figure S26: Sensitivity analysis for number of sic-SNPs by varying the bandwidth. PVE_{GREX}'s are estimated for protein CD96 across 48 tissues. The results remain nearly the same when different choices bandwidth are applied.



Supplementary Figure S27: Sensitivity analysis for genotype imputation. PVE_{GREX}'s are estimated from the CD96 protein across 48 tissues. The x-axis is the $\widehat{\text{PVE}}_{\text{GREX}}$ using the original 900 thousand SNPs; the y-axis is the $\widehat{\text{PVE}}_{\text{GREX}}$ using the imputed 3 million SNPs. The fitted line is not significantly different from the diagonal line, indicating that the IGREX analysis is robust in the presence of imputation and highly correlated SNPs.



Supplementary Figure S28: Estimated $\text{PVE}_{\text{GREX}}/h_t^2$ for four independent schizophrenia datasets.



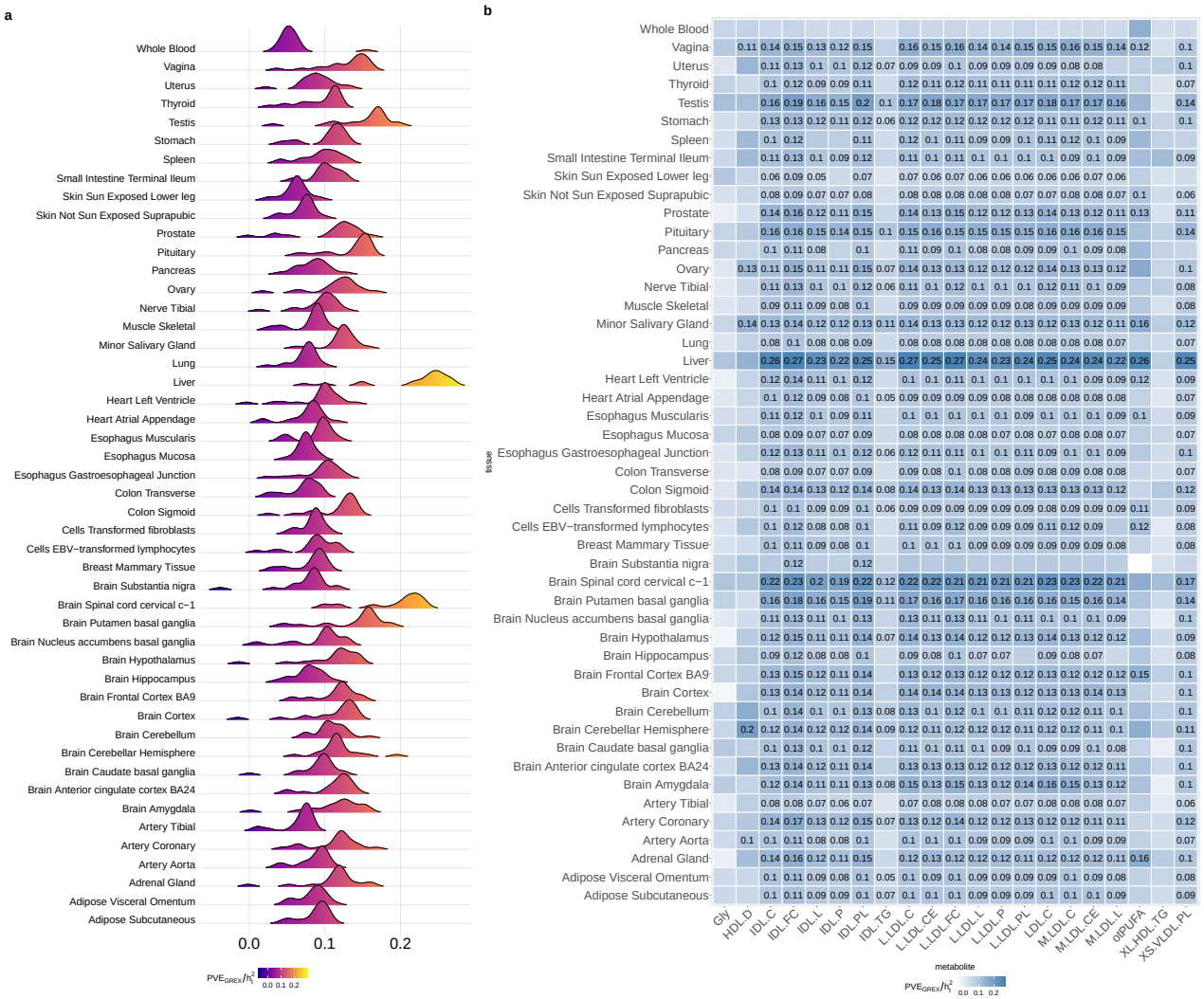
Supplementary Figure S29: Estimated PVE_{GREX} and $\text{PVE}_{\text{Alternative}}$ for BMI datasets.

2 Supplementary Note

2.1 IGREX model in the general form

IGREX can account for fixed effects in both eQTL and GWAS data. Let $\mathbf{W}_{r,g}$ be the set of c_g covariates to be adjusted (e.g. age, sex, and the first few PCs) in the eQTL study of g-th gene and $\mathbf{W}$ be the set of c covariates in the GWAS study. Then model (1) in the main text can be extended to incorporate the covariates:

$$\mathbf{y}_g = \mathbf{W}_{r,g}\mathbf{u}_g + \mathbf{X}_{r,g}\boldsymbol{\beta}_g + \mathbf{e}_{r,g}, \tag{S1}$$



Supplementary Figure S30: $\widehat{\text{PVE}}_{\text{GREX}}/h_t^2$ for 21 circulating metabolites. (**a**) The distributions of estimated $\text{PVE}_{\text{GREX}}/h_t^2$ in different tissues. (**b**) The heat map of estimated $\text{PVE}_{\text{GREX}}/h_t^2$. Entries that are significant at nominal level (0.05) are labeled with their estimated values.

where $\mathbf{u}_g \in \mathbb{R}^{c_g}$ is the fixed effect of $\mathbf{W}_{r,g} \in \mathbb{R}^{n_r \times c_g}$ and $\mathbf{e}_{r,g} \sim \mathcal{N}(0, \sigma_{r,g}^2 \mathbf{I}_{n_r})$ is a vector of independent noise. Similarly, model (2) in main text becomes:

$$\mathbf{t} = \mathbf{W}\mathbf{v} + \sum_{g=1}^{G} \alpha_g \mathbf{X}_g \boldsymbol{\beta}_g + \mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\epsilon}, \tag{S2}$$

where $\mathbf{v} \in \mathbb{R}^c$ is the fixed effect of $\mathbf{W} \in \mathbb{R}^{n \times c}$ and $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma_\epsilon^2 \mathbf{I}_n)$ is a vector of independent noise. The prior distributions of effects $\boldsymbol{\beta}_g$, α_g and $\boldsymbol{\gamma}$ are given as:

$$\boldsymbol{\beta}_g \sim \mathcal{N}(\mathbf{0}, \sigma_{\beta_g}^2 \mathbf{I}_{M_g}),\ \alpha_g \sim \mathcal{N}(\mathbf{0}, \sigma_\alpha^2),\ \boldsymbol{\gamma} \sim \mathcal{N}(\mathbf{0}, \sigma_\gamma^2 \mathbf{I}_M).$$

The two-stage IGREX procedure first evaluates the posterior of $\boldsymbol{\beta}_g$ from the eQTL study based on model (S1) and then estimates PVE_{GREX} from GWAS data by treating the obtained

posterior as prior of $\boldsymbol{\beta}_g$ in model (S2).

2.2 Stage-one: evaluate the posterior of $\boldsymbol{\beta}_g$ with PX-EM algorithm

At stage one, we evaluate the posterior $\boldsymbol{\beta}_g|\mathbf{y}_g, \mathbf{X}_{r,g}, \mathbf{W}_{r,g} \sim \mathcal{N}(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$ from the eQTL data set $\mathcal{D}_{r,g} = \{\mathbf{y}_g, \mathbf{X}_{r,g}, \mathbf{W}_{r,g}\}$ corresponding to the g-th gene. This requires estimating the collection of parameters $\boldsymbol{\theta}_g = \{\mathbf{u}_g, \sigma^2_{\beta_g}, \sigma^2_{r,g}\}$ in model (S1). Here we derive an efficient PX-EM algorithm [2] to solve this linear mixed model. We first consider the parameter expanded version of (S1):

$$\mathbf{y}_g = \mathbf{W}_{r,g}\mathbf{u}_g + \delta\mathbf{X}_{r,g}\boldsymbol{\beta}_g + \mathbf{e}_{r,g},$$

where $\delta \in \mathbb{R}$ is the expanded parameter. The complete-data log-likelihood is given as

$$\begin{aligned} &\log \Pr(\mathbf{y}_g, \boldsymbol{\beta}_g|\boldsymbol{\theta}_g; \mathbf{W}_{r,g}, \mathbf{X}_{r,g}) \\ =& -\frac{n_r}{2}\log(2\pi\sigma^2_{r,g}) - \frac{1}{2\sigma^2_{r,g}}||\mathbf{y}_g - \mathbf{W}_{r,g}\mathbf{u}_g - \delta\mathbf{X}_{r,g}\boldsymbol{\beta}_g||^2 \\ &-\frac{M_g}{2}\log(2\pi\sigma^2_{\beta_g}) - \frac{1}{2\sigma^2_{\beta_g}}||\boldsymbol{\beta}_g||^2, \end{aligned} \tag{S3}$$

from which we can easily recognize that the the terms involving $\boldsymbol{\beta}_g$ are of a quadratic form:

$$\boldsymbol{\beta}_g^T\left(-\frac{\delta^2}{2\sigma^2_{r,g}}\mathbf{X}^T_{r,g}\mathbf{X}_{r,g} - \frac{1}{2\sigma^2_{\beta_g}}\mathbf{I}_{M_g}\right)\boldsymbol{\beta}_g + \frac{\delta}{\sigma^2_{r,g}}(\mathbf{y}_g - \mathbf{W}_{r,g}\mathbf{u}_g)^T\mathbf{X}_{r,g}\boldsymbol{\beta}_g + \text{Constant}.$$

Therefore, the posterior distribution of $\boldsymbol{\beta}_g$ is Gaussian $\mathcal{N}(\boldsymbol{\beta}_g|\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$, where

$$\begin{aligned} \boldsymbol{\Sigma}_g^{-1} &= \frac{\delta^2}{\sigma^2_{r,g}}\mathbf{X}^T_{r,g}\mathbf{X}_{r,g} + \frac{1}{\sigma^2_{\beta_g}}\mathbf{I}_{M_g}, \\ \boldsymbol{\mu}_g &= \left(\frac{\delta^2}{\sigma^2_{r,g}}\mathbf{X}^T_{r,g}\mathbf{X}_{r,g} + \frac{1}{\sigma^2_{\beta_g}}\mathbf{I}_{M_g}\right)^{-1}\frac{\delta}{\sigma^2_{r,g}}\mathbf{X}^T_{r,g}(\mathbf{y}_g - \mathbf{W}_{r,g}\mathbf{u}_g). \end{aligned}$$

Now in the E-step, we evaluate the $\mathcal{Q}$-function by taking the expectation of the complete-data log-likelihood (S3) with respect to the posterior $\mathcal{N}(\boldsymbol{\beta}_g|\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$. Specifically, the quadratic terms involving $\boldsymbol{\beta}_g$ in (S3) are evaluated as following:

$$\begin{aligned} \mathbb{E}[||\tilde{\mathbf{y}}_g - \delta\mathbf{X}_{r,g}\boldsymbol{\beta}_g||^2] &= \mathbb{E}[\tilde{\mathbf{y}}_g^T\tilde{\mathbf{y}}_g - 2\delta\tilde{\mathbf{y}}_g^T\mathbf{X}_{r,g}\boldsymbol{\beta}_g + \delta^2\boldsymbol{\beta}_g^T\mathbf{X}^T_{r,g}\mathbf{X}_{r,g}\boldsymbol{\beta}_g] \\ &= \tilde{\mathbf{y}}_g^T\tilde{\mathbf{y}}_g - 2\delta\tilde{\mathbf{y}}_g^T\mathbf{X}_{r,g}\boldsymbol{\mu}_g + \delta^2\text{tr}(\mathbf{X}^T_{r,g}\mathbf{X}_{r,g}\boldsymbol{\Sigma}_g), \\ \mathbb{E}[||\boldsymbol{\beta}_g||^2] &= \boldsymbol{\mu}_g^T\boldsymbol{\mu}_g + \text{tr}(\boldsymbol{\Sigma}_g), \end{aligned}$$

where $\tilde{\mathbf{y}}_g = \mathbf{y}_g - \mathbf{W}_{r,g}\mathbf{u}_g$. Then the $\mathcal{Q}$-function given the current parameter estimates $\boldsymbol{\theta}_{g,old}$ is obtained as:

$$\begin{aligned}\mathcal{Q}(\boldsymbol{\theta}_g|\boldsymbol{\theta}_{g,old}) = &-\frac{n_r}{2}\log(2\pi\sigma^2_{r,g}) - \frac{M_g}{2}\log(2\pi\sigma^2_{\beta_g}) \\ &- \frac{1}{2\sigma^2_{r,g}}||\mathbf{y}_g - \mathbf{W}_{r,g}\mathbf{u}_g - \delta\mathbf{X}_{r,g}\boldsymbol{\mu}_g||^2 - \frac{1}{2\sigma^2_{\beta_g}}||\boldsymbol{\mu}_g||^2 \\ &- \frac{1}{2\sigma^2_{\beta_g}}\boldsymbol{\mu}_g^T\boldsymbol{\mu}_g - \mathrm{tr}\left(\left(\frac{\delta^2}{2\sigma^2_{r,g}}\mathbf{X}^T_{r,g}\mathbf{X}_{r,g} + \frac{1}{2\sigma^2_{\beta_g}}\mathbf{I}_{M_g}\right)\boldsymbol{\Sigma}_g\right).\end{aligned}$$

It the M-step, the new estimates of parameter $\boldsymbol{\theta}_g$ is obtained by setting the derivative of $\mathcal{Q}$-function to be zero. The resulting updates are given as follows:

$$\begin{aligned}\delta &= \frac{(\mathbf{y}_g - \mathbf{W}_{r,g}\mathbf{u}_g)^T\mathbf{X}_{r,g}\boldsymbol{\mu}_g}{\boldsymbol{\mu}_g^T\mathbf{X}^T_{r,g}\mathbf{X}_{r,g}\boldsymbol{\mu}_g + \mathrm{tr}(\mathbf{X}^T_{r,g}\mathbf{X}_{r,g}\boldsymbol{\Sigma}_g)}, \\ \mathbf{u}_g &= (\mathbf{X}^T_{r,g}\mathbf{X}_{r,g})^{-1}\mathbf{X}^T_{r,g}(\mathbf{y}_g - \delta\mathbf{X}_{r,g}\boldsymbol{\mu}_g), \\ \sigma^2_{r,g} &= \frac{1}{n_r}[||\mathbf{y}_g - \mathbf{W}_{r,g}\mathbf{u}_g - \delta\mathbf{X}_{r,g}\boldsymbol{\mu}_g||^2 + \delta^2\mathrm{tr}(\mathbf{X}^T_{r,g}\mathbf{X}_{r,g}\boldsymbol{\Sigma}_g)], \\ \sigma^2_{\beta_g} &= \frac{1}{M_g}[\boldsymbol{\mu}_g^T\boldsymbol{\mu}_g + \mathrm{tr}(\boldsymbol{\Sigma}_g)].\end{aligned}$$

This PX-EM algorithm is summarized in Algorithm 1. After convergence, the posterior mean and variance of $\mathcal{N}(\boldsymbol{\beta}_g|\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$ can be evaluated given the obtained parameter estimates $\hat{\boldsymbol{\theta}}_g = \{\hat{\mathbf{u}}_g, \hat{\sigma}^2_{\beta_g}, \hat{\sigma}^2_{r,g}\}$:

$$\begin{aligned}\boldsymbol{\Sigma}_g^{-1} &= \frac{1}{\hat{\sigma}^2_{r,g}}\mathbf{X}^T_{r,g}\mathbf{X}_{r,g} + \frac{1}{\hat{\sigma}^2_{\beta_g}}\mathbf{I}_{M_g}, \\ \boldsymbol{\mu}_g &= \left(\frac{1}{\hat{\sigma}^2_{r,g}}\mathbf{X}^T_{r,g}\mathbf{X}_{r,g} + \frac{1}{\hat{\sigma}^2_{\beta_g}}\mathbf{I}_{M_g}\right)^{-1}\frac{1}{\hat{\sigma}^2_{r,g}}\mathbf{X}^T_{r,g}(\mathbf{y}_g - \mathbf{W}_{r,g}\hat{\mathbf{u}}_g).\end{aligned} \tag{S4}$$

2.3 Stage-two: estimate $\mathrm{PVE}_{\mathrm{GREX}}$

Given the parameters in (S4), we then estimate $\mathrm{PVE}_{\mathrm{GREX}} = \frac{\mathrm{Var}(\sum_{g=1}^G \alpha_g \mathbf{x}_g^T\boldsymbol{\beta}_g)}{\mathrm{Var}(t)}$ by treating the posterior distribution $\mathcal{N}(\boldsymbol{\beta}_g|\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$ as prior distribution of $\boldsymbol{\beta}_g$ in model (S2). To account for the fixed effects $\mathbf{W}$, we first multiply the projection matrix $\mathbf{M} = \mathbf{I}_n - \mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T$ on both sides of model (S2). Then, we have $\mathbb{E}(\mathbf{Mt}|\boldsymbol{\alpha}) = \sum_{g=1}^G \alpha_g\mathbf{M}\mathbf{X}_g\boldsymbol{\mu}_g$ and $\mathrm{Cov}(\mathbf{t}|\boldsymbol{\alpha}) = \sum_{g=1}^G \alpha_g^2\mathbf{M}\mathbf{X}_g\boldsymbol{\Sigma}_g(\mathbf{M}\mathbf{X}_g)^T + \sigma^2_\gamma\mathbf{M}\mathbf{X}(\mathbf{M}\mathbf{X})^T + \sigma^2_\epsilon\mathbf{M}$. By applying the law of total expectation and total variance, we obtain $\mathbb{E}(\mathbf{Mt}) = \mathbb{E}(\mathbb{E}(\mathbf{Mt}|\boldsymbol{\alpha})) = \mathbf{0}$ and

$$\mathrm{Cov}(\mathbf{Mt}) = \mathrm{Cov}(\mathbb{E}(\mathbf{Mt}|\boldsymbol{\alpha})) + \mathbb{E}(\mathrm{Cov}(\mathbf{Mt}|\boldsymbol{\alpha})) = \sigma^2_\alpha\mathbf{M}\mathbf{K}_\alpha\mathbf{M} + \sigma^2_\gamma\mathbf{M}\mathbf{K}_\gamma\mathbf{M} + \sigma^2_\epsilon\mathbf{M} =: \boldsymbol{\Omega},$$

Algorithm 1 PX-EM algorithm for model (S1)

Initialization: Parameters are initialized by setting $\mathbf{u}_g = (\mathbf{W}_{r,g}^T\mathbf{W}_{r,g})^{-1}\mathbf{W}_{r,g}^T\mathbf{y}_g$, $\sigma_{r,g}^2 = \sigma_{\beta_g}^2 = \mathrm{Var}(y - \mathbf{W}_{r,g}\mathbf{u}_g)/2$.

repeat

E-step: At the t-th iteration, evaluate the posterior $\mathcal{N}(\boldsymbol{\beta}_g|\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$ given the current parameter estimates $\boldsymbol{\theta}_g^{(t)} = \{\mathbf{u}_g^{(t)}, (\sigma_{r,g}^{(t)})^2, (\sigma_{\beta_g}^{(t)})^2\}$ and $\delta^{(t)} = 1$:

$$\boldsymbol{\Sigma}_g^{-1} = \frac{(\delta^{(t)})^2}{(\sigma_{r,g}^{(t)})^2}\mathbf{X}_{r,g}^T\mathbf{X}_{r,g} + \frac{1}{(\sigma_{\beta_g}^{(t)})^2}\mathbf{I}_{M_g},$$

$$\boldsymbol{\mu}_g = \left(\frac{(\delta^{(t)})^2}{(\sigma_{r,g}^{(t)})^2}\mathbf{X}_{r,g}^T\mathbf{X}_{r,g} + \frac{1}{(\sigma_{\beta_g}^{(t)})^2}\mathbf{I}_{M_g}\right)^{-1}\frac{(\delta^{(t)})^2}{(\sigma_{r,g}^{(t)})^2}\mathbf{X}_{r,g}^T(\mathbf{y}_g - \mathbf{W}_{r,g}\mathbf{u}_g^{(t)})$$

M-step: Update the model parameters $\boldsymbol{\theta}_g$ by

$$\delta^{(t+1)} = \frac{(\mathbf{y}_g - \mathbf{W}_{r,g}\mathbf{u}_g^{(t)})^T\mathbf{X}_{r,g}\boldsymbol{\mu}_g}{\boldsymbol{\mu}_g^T\mathbf{X}_{r,g}^T\mathbf{X}_{r,g}\boldsymbol{\mu}_g + \mathrm{tr}(\mathbf{X}_{r,g}^T\mathbf{X}_{r,g}\boldsymbol{\Sigma}_g)},$$

$$\mathbf{u}_g^{(t+1)} = (\mathbf{X}_{r,g}^T\mathbf{X}_{r,g})^{-1}\mathbf{X}_{r,g}^T(\mathbf{y}_g - \delta\mathbf{X}_{r,g}\boldsymbol{\mu}_g),$$

$$(\sigma_{r,g}^{(t+1)})^2 = \frac{1}{n_r}[||\mathbf{y}_g - \mathbf{W}_{r,g}\mathbf{u}_g^{(t)} - \delta\mathbf{X}_{r,g}\boldsymbol{\mu}_g||^2 + \delta^2\mathrm{tr}(\mathbf{X}_{r,g}^T\mathbf{X}_{r,g}\boldsymbol{\Sigma}_g)],$$

$$(\sigma_{\beta_g}^{(t+1)})^2 = \frac{1}{M_g}[\boldsymbol{\mu}_g^T\boldsymbol{\mu}_g + \mathrm{tr}(\boldsymbol{\Sigma}_g)].$$

Reduction-step: Rescale $(\sigma_{\beta_g}^{(t+1)})^2 = (\delta^{(t+1)})^2(\sigma_{\beta_g}^{(t+1)})^2$ and reset $\delta^{(t+1)} = 1$.

until the incomplete-data log-likelihood stops increasing or maximum iteration is reached

respectively, where $\mathbf{K}_\alpha = \sum_{g=1}^{G} \mathbf{X}_g(\boldsymbol{\mu}_g\boldsymbol{\mu}_g^T + \boldsymbol{\Sigma}_g)\mathbf{X}_g^T$ and $\mathbf{K}_\gamma = \mathbf{X}\mathbf{X}^T$. Clearly, the i-th diagonal element of $\sigma_\alpha^2\mathbf{M}\mathbf{K}_\alpha\mathbf{M}$ and $\sigma_\gamma^2\mathbf{M}\mathbf{K}_\gamma\mathbf{M}$ represents the variance explained by GREX and alternative genetic effects, respectively. Hence, the PVE_{GREX}, $\text{PVE}_{\text{Alternative}}$, h_t^2 and $\text{PVE}_{\text{GREX}}/h_t^2$ are estimated by

$$\begin{aligned}
\widehat{\text{PVE}}_{\text{GREX}} &= \frac{\text{tr}(\hat{\sigma}_\alpha^2\mathbf{M}\mathbf{K}_\alpha\mathbf{M})}{\text{tr}(\hat{\sigma}_\alpha^2\mathbf{M}\mathbf{K}_\alpha\mathbf{M} + \hat{\sigma}_\gamma^2\mathbf{M}\mathbf{K}_\gamma\mathbf{M} + \hat{\sigma}_\epsilon^2\mathbf{M})}, \\
\widehat{\text{PVE}}_{\text{Alternative}} &= \frac{\text{tr}(\hat{\sigma}_\gamma^2\mathbf{M}\mathbf{K}_\gamma\mathbf{M})}{\text{tr}(\hat{\sigma}_\alpha^2\mathbf{M}\mathbf{K}_\alpha\mathbf{M} + \hat{\sigma}_\gamma^2\mathbf{M}\mathbf{K}_\gamma\mathbf{M} + \hat{\sigma}_\epsilon^2\mathbf{M})}, \\
\hat{h}_t^2 &= \frac{\text{tr}(\hat{\sigma}_\alpha^2\mathbf{M}\mathbf{K}_\alpha\mathbf{M} + \hat{\sigma}_\gamma^2\mathbf{M}\mathbf{K}_\gamma\mathbf{M})}{\text{tr}(\hat{\sigma}_\alpha^2\mathbf{M}\mathbf{K}_\alpha\mathbf{M} + \hat{\sigma}_\gamma^2\mathbf{M}\mathbf{K}_\gamma\mathbf{M} + \hat{\sigma}_\epsilon^2\mathbf{M})}, \\
\widehat{\text{PVE}}_{\text{GREX}}/\hat{h}_t^2 &= \frac{\text{tr}(\hat{\sigma}_\alpha^2\mathbf{M}\mathbf{K}_\alpha\mathbf{M})}{\text{tr}(\hat{\sigma}_\alpha^2\mathbf{M}\mathbf{K}_\alpha\mathbf{M} + \hat{\sigma}_\gamma^2\mathbf{M}\mathbf{K}_\gamma\mathbf{M})},
\end{aligned} \tag{S5}$$

where $\hat{\sigma}_\alpha^2$, $\hat{\sigma}_\gamma^2$ and $\hat{\sigma}_\epsilon^2$ are the estimated values of σ_α^2, σ_γ^2 and σ_ϵ^2, respectively. Letting $\boldsymbol{\theta} = \{\sigma_\alpha^2, \sigma_\gamma^2, \sigma_\epsilon^2\}$ be the collection of parameters to be estimated, its corresponding estimate $\hat{\boldsymbol{\theta}} = \{\hat{\sigma}_\alpha^2, \hat{\sigma}_\gamma^2, \hat{\sigma}_\epsilon^2\}$ can be obtained by REML or MoM when the individual-level GWAS data $\mathcal{D}_i = \{\mathbf{t}, \mathbf{X}, \mathbf{W}\}$ is available. Both approaches are provided in the IGREX framework (IGREX-i): REML is statistically more efficient by assuming normality of $\mathbf{t}$ while MoM is more robust and computationally efficient. Besides, we developed IGREX-s based on MoM to handle GWAS summary statistics when the individual-level GWAS data is inaccessible.

2.3.1 IGREX-i

The MoM estimate of $\boldsymbol{\psi}$ is obtained by minimizing the distance between the second moment of $\mathbf{t}$ at the population level and that at the sample level:

$$\min_{\boldsymbol{\psi}} \; f(\boldsymbol{\psi}) = ||(\mathbf{M}\mathbf{t})(\mathbf{M}\mathbf{t})^T - (\sigma_\alpha^2\mathbf{M}\mathbf{K}_\alpha\mathbf{M} + \sigma_\gamma^2\mathbf{M}\mathbf{K}_\gamma\mathbf{M} + \sigma_\epsilon^2\mathbf{M})||^2.$$

The solution of this optimization problem is obtained by setting $\frac{\partial f(\boldsymbol{\psi})}{\partial \sigma_\alpha^2} = \frac{\partial f(\boldsymbol{\psi})}{\partial \sigma_\gamma^2} = \frac{\partial f(\boldsymbol{\psi})}{\partial \sigma_\epsilon^2} = 0$, which gives the estimating equation:

$$\mathbf{S}\boldsymbol{\psi} = \mathbf{q}, \tag{S6}$$

$$\text{with } \mathbf{S} = \begin{bmatrix} \text{tr}((\mathbf{M}\mathbf{K}_\alpha)^2) & \text{tr}(\mathbf{M}\mathbf{K}_\alpha\mathbf{M}\mathbf{K}_\gamma) & \text{tr}(\mathbf{M}\mathbf{K}_\alpha) \\ \text{tr}(\mathbf{M}\mathbf{K}_\alpha\mathbf{M}\mathbf{K}_\gamma) & \text{tr}((\mathbf{M}\mathbf{K}_\gamma)^2) & \text{tr}(\mathbf{M}\mathbf{K}_\gamma) \\ \text{tr}(\mathbf{M}\mathbf{K}_\alpha) & \text{tr}(\mathbf{M}\mathbf{K}_\gamma) & n-c \end{bmatrix}, \; \boldsymbol{\psi} = \begin{bmatrix} \sigma_\alpha^2 \\ \sigma_\gamma^2 \\ \sigma_\epsilon^2 \end{bmatrix}, \; \mathbf{q} = \begin{bmatrix} \mathbf{t}^T\mathbf{M}\mathbf{K}_\alpha\mathbf{M}\mathbf{t} \\ \mathbf{t}^T\mathbf{M}\mathbf{K}_\gamma\mathbf{M}\mathbf{t} \\ \mathbf{t}^T\mathbf{M}\mathbf{t} \end{bmatrix}.$$

Thus, $\hat{\boldsymbol{\psi}}$ is obtained by calculating $\hat{\boldsymbol{\psi}} = \mathbf{S}^{-1}\mathbf{q}$. To estimate the standard error of $\widehat{\text{PVE}}_{\text{GREX}}$, we first calculate the standard errors of $\hat{\boldsymbol{\psi}}$ using sandwich estimator:

$$\text{Cov}(\hat{\boldsymbol{\psi}}) = \mathbf{S}^{-1}\text{Cov}(\mathbf{q})\mathbf{S}^{-1},$$

where

$$\mathrm{Cov}(\mathbf{q}) = \begin{bmatrix} \mathrm{Var}(\mathbf{t}^T\mathbf{M}\mathbf{K}_\alpha\mathbf{M}\mathbf{t}) & . & . \\ \mathrm{Cov}(\mathbf{t}^T\mathbf{M}\mathbf{K}_\alpha\mathbf{M}\mathbf{t}, \mathbf{t}^T\mathbf{M}\mathbf{K}_\gamma\mathbf{M}\mathbf{t}) & \mathrm{Var}(\mathbf{t}^T\mathbf{M}\mathbf{K}_\gamma\mathbf{M}\mathbf{t}) & . \\ \mathrm{Cov}(\mathbf{t}^T\mathbf{M}\mathbf{K}_\alpha\mathbf{M}\mathbf{t}, \mathbf{t}^T\mathbf{t}) & \mathrm{Cov}(\mathbf{t}^T\mathbf{M}\mathbf{K}_\gamma\mathbf{M}\mathbf{t}, \mathbf{t}^T\mathbf{M}\mathbf{t}) & \mathrm{Var}(\mathbf{t}^T\mathbf{M}\mathbf{t}) \end{bmatrix}$$
$$= \begin{bmatrix} 2\mathrm{tr}((\mathbf{M}\mathbf{K}_\alpha\mathbf{M}\boldsymbol{\Omega})^2) & . & . \\ 2\mathrm{tr}(\mathbf{M}\mathbf{K}_\alpha\mathbf{M}\mathbf{K}_\gamma(\mathbf{M}\boldsymbol{\Omega})^2) & 2\mathrm{tr}((\mathbf{M}\mathbf{K}_\gamma\mathbf{M}\boldsymbol{\Omega})^2) & . \\ 2\mathrm{tr}(\mathbf{M}\mathbf{K}_\alpha(\mathbf{M}\boldsymbol{\Omega})^2) & 2\mathrm{tr}(\mathbf{M}\mathbf{K}_\gamma(\mathbf{M}\boldsymbol{\Omega})^2) & 2\mathrm{tr}((\mathbf{M}\boldsymbol{\Omega})^2) \end{bmatrix}$$

is the symmetric covariance matrix of $\mathbf{q}$. Then, the variances of the PVE estimators defined in (S5) are obtained by the delta method:

$$\begin{aligned} \mathrm{Var}(\widehat{\mathrm{PVE}}_{\mathrm{GREX}}) &= \nabla_G^T \mathrm{Cov}(\hat{\boldsymbol{\psi}})\nabla_G \\ \mathrm{Var}(\widehat{\mathrm{PVE}}_{\mathrm{Alternative}}) &= \nabla_A^T \mathrm{Cov}(\hat{\boldsymbol{\psi}})\nabla_A \\ \mathrm{Var}(\hat{h}_t^2) &= \nabla_h^T \mathrm{Cov}(\hat{\boldsymbol{\psi}})\nabla_h \\ \mathrm{Var}(\widehat{\mathrm{PVE}}_{\mathrm{GREX}}/\hat{h}_t^2) &= \nabla_{prop}^T \mathrm{Cov}(\hat{\boldsymbol{\psi}})\nabla_{prop}, \end{aligned}$$

where

$$\nabla_G = \begin{bmatrix} \frac{\hat{\sigma}_\gamma^2\mathrm{tr}(\mathbf{M}\mathbf{K}_\alpha)\mathrm{tr}(\mathbf{M}\mathbf{K}_\gamma)+(n-c)\hat{\sigma}_\epsilon^2\mathrm{tr}(\mathbf{M}\mathbf{K}_\alpha)}{(\hat{\sigma}_\alpha^2\mathrm{tr}(\mathbf{M}\mathbf{K}_\alpha)+\hat{\sigma}_\gamma^2\mathrm{tr}(\mathbf{M}\mathbf{K}_\gamma)+\hat{\sigma}_\epsilon^2(n-c))^2} \\ -\frac{\hat{\sigma}_\alpha^2\mathrm{tr}(\mathbf{M}\mathbf{K}_\alpha)\mathrm{tr}(\mathbf{M}\mathbf{K}_\gamma)}{(\hat{\sigma}_\alpha^2\mathrm{tr}(\mathbf{M}\mathbf{K}_\alpha)+\hat{\sigma}_\gamma^2\mathrm{tr}(\mathbf{M}\mathbf{K}_\gamma)+\hat{\sigma}_\epsilon^2(n-c))^2} \\ -\frac{(n-c)\hat{\sigma}_\alpha^2\mathrm{tr}(\mathbf{M}\mathbf{K}_\alpha)}{(\hat{\sigma}_\alpha^2\mathrm{tr}(\mathbf{M}\mathbf{K}_\alpha)+\hat{\sigma}_\gamma^2\mathrm{tr}(\mathbf{M}\mathbf{K}_\gamma)+\hat{\sigma}_\epsilon^2(n-c))^2} \end{bmatrix},$$

$$\nabla_A = \begin{bmatrix} -\frac{\hat{\sigma}_\gamma^2\mathrm{tr}(\mathbf{M}\mathbf{K}_\alpha)\mathrm{tr}(\mathbf{M}\mathbf{K}_\gamma)}{(\hat{\sigma}_\alpha^2\mathrm{tr}(\mathbf{M}\mathbf{K}_\alpha)+\hat{\sigma}_\gamma^2\mathrm{tr}(\mathbf{M}\mathbf{K}_\gamma)+\hat{\sigma}_\epsilon^2(n-c))^2} \\ \frac{\hat{\sigma}_\alpha^2\mathrm{tr}(\mathbf{M}\mathbf{K}_\alpha)\mathrm{tr}(\mathbf{M}\mathbf{K}_\gamma)+(n-c)\hat{\sigma}_\epsilon^2\mathrm{tr}(\mathbf{M}\mathbf{K}_\gamma)}{(\hat{\sigma}_\alpha^2\mathrm{tr}(\mathbf{M}\mathbf{K}_\alpha)+\hat{\sigma}_\gamma^2\mathrm{tr}(\mathbf{M}\mathbf{K}_\gamma)+\hat{\sigma}_\epsilon^2(n-c))^2} \\ -\frac{(n-c)\hat{\sigma}_\gamma^2\mathrm{tr}(\mathbf{M}\mathbf{K}_\gamma)}{(\hat{\sigma}_\alpha^2\mathrm{tr}(\mathbf{M}\mathbf{K}_\alpha)+\hat{\sigma}_\gamma^2\mathrm{tr}(\mathbf{M}\mathbf{K}_\gamma)+\hat{\sigma}_\epsilon^2(n-c))^2} \end{bmatrix},$$

$$\nabla_h = \begin{bmatrix} \frac{(n-c)\hat{\sigma}_\epsilon^2\mathrm{tr}(\mathbf{M}\mathbf{K}_\alpha)}{(\hat{\sigma}_\alpha^2\mathrm{tr}(\mathbf{M}\mathbf{K}_\alpha)+\hat{\sigma}_\gamma^2\mathrm{tr}(\mathbf{M}\mathbf{K}_\gamma)+\hat{\sigma}_\epsilon^2(n-c))^2} \\ \frac{(n-c)\hat{\sigma}_\epsilon^2\mathrm{tr}(\mathbf{M}\mathbf{K}_\gamma)}{(\hat{\sigma}_\alpha^2\mathrm{tr}(\mathbf{M}\mathbf{K}_\alpha)+\hat{\sigma}_\gamma^2\mathrm{tr}(\mathbf{M}\mathbf{K}_\gamma)+\hat{\sigma}_\epsilon^2(n-c))^2} \\ -\frac{(n-c)\hat{\sigma}_\alpha^2\mathrm{tr}(\mathbf{M}\mathbf{K}_\alpha)+(n-c)\hat{\sigma}_\gamma^2\mathrm{tr}(\mathbf{M}\mathbf{K}_\gamma)}{(\hat{\sigma}_\alpha^2\mathrm{tr}(\mathbf{M}\mathbf{K}_\alpha)+\hat{\sigma}_\gamma^2\mathrm{tr}(\mathbf{M}\mathbf{K}_\gamma)+\hat{\sigma}_\epsilon^2(n-c))^2} \end{bmatrix},$$

$$\nabla_{prop} = \begin{bmatrix} \frac{\hat{\sigma}_\gamma^2\mathrm{tr}(\mathbf{M}\mathbf{K}_\alpha)\mathrm{tr}(\mathbf{M}\mathbf{K}_\gamma)}{(\hat{\sigma}_\alpha^2\mathrm{tr}(\mathbf{M}\mathbf{K}_\alpha)+\hat{\sigma}_\gamma^2\mathrm{tr}(\mathbf{M}\mathbf{K}_\gamma)} \\ -\frac{\hat{\sigma}_\alpha^2\mathrm{tr}(\mathbf{M}\mathbf{K}_\alpha)\mathrm{tr}(\mathbf{M}\mathbf{K}_\gamma)}{(\hat{\sigma}_\alpha^2\mathrm{tr}(\mathbf{M}\mathbf{K}_\alpha)+\hat{\sigma}_\gamma^2\mathrm{tr}(\mathbf{M}\mathbf{K}_\gamma)} \\ 0 \end{bmatrix}.$$

Alternatively, we can apply MM algorithm described in [3] to obtain REML estimate of $\boldsymbol{\theta}$ by further assuming the normality $\mathbf{t} \sim \mathcal{N}(\mathbf{t}|\mathbf{W}\mathbf{v}, \sigma_\alpha^2\mathbf{K}_\alpha + \sigma_\gamma^2\mathbf{K}_\gamma + \sigma_\epsilon^2\mathbf{I}_n)$.

2.3.2 IGREX-s

Next, we derive general form of IGREX-s for incorporating covariates for the LD reference $\tilde{\mathbf{X}}$. To obtain $\widehat{\mathrm{PVE}}_{\mathrm{GREX}}$ given by Equation (10) in the main text, we first reorganize (S6) by

eliminating the σ_ϵ^2 and dividing both sides by $(n-c)^2$:

$$\begin{bmatrix} \frac{\mathrm{tr}((\mathbf{MK}_\alpha)^2)-\frac{\mathrm{tr}^2(\mathbf{MK}_\alpha)}{(n-c)}}{(n-c)^2} & \frac{\mathrm{tr}(\mathbf{MK}_\alpha\mathbf{MK}_\gamma)-\frac{\mathrm{tr}(\mathbf{MK}_\alpha)\mathrm{tr}(\mathbf{MK}_\gamma)}{(n-c)}}{(n-c)^2} \\ \frac{\mathrm{tr}(\mathbf{MK}_\alpha\mathbf{MK}_\gamma)-\frac{\mathrm{tr}(\mathbf{MK}_\alpha)\mathrm{tr}(\mathbf{MK}_\gamma)}{(n-c)}}{(n-c)^2} & \frac{\mathrm{tr}((\mathbf{MK}_\gamma)^2)-\frac{\mathrm{tr}^2(\mathbf{MK}_\gamma)}{(n-c)}}{(n-c)^2} \end{bmatrix} \begin{bmatrix} \sigma_\alpha^2 \\ \sigma_\gamma^2 \end{bmatrix} = \begin{bmatrix} \frac{\mathbf{t}^T\mathbf{MK}_\alpha\mathbf{Mt}}{(n-c)^2} - \frac{\mathrm{tr}(\mathbf{MK}_\alpha)}{(n-c)^3}\mathbf{t}^T\mathbf{Mt} \\ \frac{\mathbf{t}^T\mathbf{MK}_\gamma\mathbf{Mt}}{(n-c)^2} - \frac{\mathrm{tr}(\mathbf{MK}_\gamma)}{(n-c)^3}\mathbf{t}^T\mathbf{Mt} \end{bmatrix}.$$

The terms on the left hand side does not involve $\mathbf{t}$ and thus can be approximated using an LD reference $\tilde{\mathbf{X}}$ [1]. For example, $\frac{\mathrm{tr}(\mathbf{MK}_\alpha^2)-\frac{\mathrm{tr}^2(\mathbf{MK}_\alpha)}{(n-c)}}{(n-c)^2}$ can be well approximated by $\frac{\mathrm{tr}(\tilde{\mathbf{M}}\tilde{\mathbf{K}}_\alpha^2)-\frac{\mathrm{tr}^2(\tilde{\mathbf{M}}\tilde{\mathbf{K}}_\alpha)}{(m-c)}}{(m-c)^2}$, where $\tilde{\mathbf{K}}_\alpha = \sum_{g=1}^G \tilde{\mathbf{X}}_g(\boldsymbol{\mu}_g\boldsymbol{\mu}_g^T + \boldsymbol{\Sigma}_g)\tilde{\mathbf{X}}_g^T$ and $\tilde{\mathbf{M}}$ is the projection matrix corresponding to the covariates associated with the LD reference matrix $\tilde{\mathbf{X}}$. Other terms on the left hand side can be approximated in the same way. Usually, the confounding covariates have been adjusted when calculating the z-scores. Thus, the terms in the right hand side can be approximated as follows:

$$\begin{aligned}
&\frac{\mathbf{t}^T\mathbf{MK}_\alpha\mathbf{Mt}}{(n-c)^2} \approx \frac{\mathbf{t}^T\mathbf{K}_\alpha\mathbf{t}}{n^2} \approx \frac{1}{n}\hat{\sigma}_t^2\sum_g \mathbf{z}_g^T(\boldsymbol{\mu}_g\boldsymbol{\mu}_g^T + \boldsymbol{\Sigma}_g)\mathbf{z}_g, \\
&\frac{\mathrm{tr}(\mathbf{MK}_\alpha)}{(n-c)^3}\mathbf{t}^T\mathbf{Mt} \approx \frac{\mathrm{tr}(\mathbf{K}_\alpha)}{n^3}\mathbf{t}^T\mathbf{t} \approx \frac{1}{n}\hat{\sigma}_t^2\mathrm{tr}(\sum_g(\boldsymbol{\mu}_g\boldsymbol{\mu}_g^T + \boldsymbol{\Sigma}_g)\hat{\mathbf{R}}_g), \\
&\frac{\mathbf{t}^T\mathbf{MK}_\gamma\mathbf{Mt}}{(n-c)^2} \approx \frac{\mathbf{t}^T\mathbf{K}_\gamma\mathbf{t}}{n^2} \approx \frac{1}{n}\hat{\sigma}_t^2\sum_{j=1}^M z_j^2, \\
&\frac{\mathrm{tr}(\mathbf{MK}_\gamma)}{(n-c)^3}\mathbf{t}^T\mathbf{Mt} \approx \frac{\mathrm{tr}(\mathbf{K}_\gamma)}{n^3}\mathbf{t}^T\mathbf{t} \approx \frac{1}{n}\hat{\sigma}_t^2.
\end{aligned}$$

Based on these approximations, the estimating equation becomes

$$\begin{aligned}
&\begin{bmatrix} \frac{\mathrm{tr}((\tilde{\mathbf{M}}\tilde{\mathbf{K}}_\alpha)^2)-\frac{\mathrm{tr}^2(\tilde{\mathbf{M}}\tilde{\mathbf{K}}_\alpha)}{(m-c)}}{(m-c)^2} & \frac{\mathrm{tr}(\tilde{\mathbf{M}}\mathbf{K}_\alpha\tilde{\mathbf{M}}\tilde{\mathbf{K}}_\gamma)-\frac{\mathrm{tr}(\tilde{\mathbf{M}}\tilde{\mathbf{K}}_\alpha)\mathrm{tr}(\tilde{\mathbf{M}}\tilde{\mathbf{K}}_\gamma)}{(m-c)}}{(m-c)^2} \\ \frac{\mathrm{tr}(\tilde{\mathbf{M}}\tilde{\mathbf{K}}_\alpha\tilde{\mathbf{M}}\tilde{\mathbf{K}}_\gamma)-\frac{\mathrm{tr}(\tilde{\mathbf{M}}\tilde{\mathbf{K}}_\alpha)\mathrm{tr}(\tilde{\mathbf{M}}\tilde{\mathbf{K}}_\gamma)}{(m-c)}}{(m-c)^2} & \frac{\mathrm{tr}((\tilde{\mathbf{M}}\tilde{\mathbf{K}}_\gamma)^2)-\frac{\mathrm{tr}^2(\tilde{\mathbf{M}}\tilde{\mathbf{K}}_\gamma)}{(m-c)}}{(m-c)^2} \end{bmatrix} \begin{bmatrix} \frac{\hat{\sigma}_\alpha^2}{\hat{\sigma}_t^2} \\ \frac{\hat{\sigma}_\gamma^2}{\hat{\sigma}_t^2} \end{bmatrix} \\
&= \begin{bmatrix} \frac{\sum_g \mathbf{z}_g^T(\boldsymbol{\mu}_g\boldsymbol{\mu}_g^T+\boldsymbol{\Sigma}_g)\mathbf{z}_g - \mathrm{tr}(\sum_g(\boldsymbol{\mu}_g\boldsymbol{\mu}_g^T+\boldsymbol{\Sigma}_g)\hat{\mathbf{R}}_g)}{n} \\ \sum_{j=1}^M \frac{z_j^2-1}{n} \end{bmatrix}.
\end{aligned} \tag{S7}$$

Solving this equation and plugging the obtained $\frac{\hat{\sigma}_\alpha^2}{\hat{\sigma}_t^2}$ into Equation (10) give the estimate of $\mathrm{PVE}_{\mathrm{GREX}}$. When the summary data is from meta-analysis of multiple GWASs, the sample size for each SNP may be different. To handle this heterogeneity in sample size, the right hand side of (S7) is further substituted by

$$\left[\sum_g \tilde{\mathbf{z}}_g^T(\boldsymbol{\mu}_g\boldsymbol{\mu}_g^T + \boldsymbol{\Sigma}_g)\tilde{\mathbf{z}}_g - \sum_g\left[\mathrm{tr}((\boldsymbol{\mu}_g\boldsymbol{\mu}_g^T + \boldsymbol{\Sigma}_g)\hat{\mathbf{R}}_g)(\tfrac{1}{M_g}\textstyle\sum_j^{M_g}\frac{1}{n_j})\right],\ \ \sum_{j=1}^M \frac{z_j^2-1}{n_j}\right]^T,$$

where $\tilde{\mathbf{z}}_g = [\mathbf{z}_{g,1}/\sqrt{n_1}, ..., \mathbf{z}_{g,j}/\sqrt{n_j}, ..., \mathbf{z}_{g,M_g}/\sqrt{n_{M_g}}]^T$. The standard errors of estimated PVE are calculated by block-wise jackknife where the z-scores are divided into $1,704$ approximately

independent blocks according to the LD scores of corresponding SNPs [4]. To guarantee the independence between blocks, only the genes whose local SNPs all lie inside an LD block are included for calculation. When there are large overlaps between gene coverage and LD block cut-off, may genes can be excluded from the calculation. In this case, the jackknife approach tends to slightly over-estimate the standard error (See Supplementary Figure S14).

2.4 Tissue-wise hypothesis testing

In addition to the point estimate of PVE_{GREX}, it is of interest to test whether PVE_{GREX} is significantly different from zero, i.e., testing the hypothesis $H_0 : \text{PVE}_{\text{GREX}} = 0$ or simply $H_0 : \sigma_\alpha^2 = 0$. Because $\hat{\sigma}_\alpha^2$ is on the boundary of the parameter space under the null hypothesis, it follows a mixture of χ^2 distribution: $\hat{\sigma}_\alpha^2 \sim \sum_{i=1}^n \lambda_i \chi_1^2$, where χ_1^2 is the chi-squared random variable with degree of freedom one and λ_i is the i-th eigenvalue of the matrix

$$(\hat{\sigma}_{\gamma,H_0}^2 \mathbf{M}\mathbf{K}_\gamma\mathbf{M} + \hat{\sigma}_{\epsilon,H_0}^2 \mathbf{M})^{1/2}\mathbf{H}(\hat{\sigma}_{\gamma,H_0}^2 \mathbf{M}\mathbf{K}_\gamma\mathbf{M} + \hat{\sigma}_{\epsilon,H_0}^2 \mathbf{M})^{1/2},$$

where $\hat{\sigma}_{\gamma,H_0}^2$ and $\hat{\sigma}_{\epsilon,H_0}^2$ are the estimates of σ_γ^2 and σ_ϵ^2 under H_0, respectively, and $\mathbf{H} = (\mathbf{S}^{-1})_{1,1}\mathbf{M}\mathbf{K}_\alpha\mathbf{M} + (\mathbf{S}^{-1})_{1,2}\mathbf{M}\mathbf{K}_\gamma\mathbf{M} + (\mathbf{S}^{-1})_{1,2}\mathbf{M}$. The exact p-values are usually obtained by applying the Davies method [5].

While the Davies method provides well calibrated p-values, it is computationally challenging when the sample size is large. Therefore, we consider an efficient approximation of this test using the point estimate of PVE_{GREX} and its standard error. Specifically, we first compute the test statistic of $\widehat{\text{PVE}}_{\text{GREX}}$ by $z = \frac{\widehat{\text{PVE}}_{\text{GREX}}}{se(\widehat{\text{PVE}}_{\text{GREX}})}$. Then we calculate p-value$= 2 \times \Phi(-|z|)$, assuming that z follows the standard normal distribution, where Φ is the cumulative distribution function of the standard normal distribtion. Indeed, we adopted the same approach as LDSC [6] and GNOVA [7] to calculate p-values, and then evaluated statistical significance. From the QQ-plots shown in the Supplementary Figure S7 (with details given in the Supplementary Section 2.5), we can see that the approximation is reasonably good in most cases.

2.5 Additional simulation analysis

In addition to the simulations using the NFBC genotypes in the main text, we further investigated the performance of IGREX under various parameter settings with simulated genotypes. For all the simulated data, we fixed $n = 4,000$, $G = 200$, $M = 20,000$ (i.e., 100 cis SNPs for each gene). The total phenotypic heritability was set as $h_t^2 = 0.5$. To simulate the

genotype data, we first sampled the minor allele frequencies (MAF) from uniform distribution $\mathcal{U}(0.05, 0.5)$ and data matrices from normal distribution $\mathcal{N}(\mathbf{0}, \Sigma(\rho))$, where $\Sigma_{jj'} = \rho^{|j-j'|}$ characterizes the LD patterns between SNPs. Then, the genotype matrices $\mathbf{X}_r$ and $\mathbf{X}$ were obtained by categorizing the entries of generated data matrices into $0, 1, 2$ according to MAF. Given the genotype matrices, $\boldsymbol{\beta}_g$ and α_g, the gene expression $\mathbf{y}_g$ and phenotype $\mathbf{t}$ were simulated following models (1) and (2) in the main text.

Supplementary Figure S1 shows the results when n_r is varied at $\{800, 1000, 2000\}$ and PVE_y is varied at $\{0.1, 0.2, 0.3\}$, where the pattern is similar to the one of the Figure 1a in the main text. Supplementary Figure S2 shows the results when n_r is varied at $\{800, 1000, 2000\}$ and PVE_{GREX} is varied at $\{0.1, 0.2, 0.3, 0.4\}$, indicating that the PVE_{GREX} does not influence the estimation accuracy when PVE_t is fixed. Supplementary Figure S3-S4 illustrate the model performance by assuming under sparse effects of $\boldsymbol{\beta}_g$ and $\boldsymbol{\alpha}$, respectively. The results are consistent with those given in Figure 1c-d in the main text.

To analyze the approximation performance of the normal test to the Davies method, we additionaly compared their p-values under the null hypothesis $H_0 : \sigma_\alpha^2 = 0$ (i.e., $\text{PVE}_{\text{GREX}} = 0$) using the simulated genotypes with $n_r = 800$, $n \in \{800, 2000, 4000\}$ and $\text{PVE}_y = 0.3$. As we can observe in Supplementary Figure S7, the approximations are reasonably good in most cases.

2.6 Sensitivity analysis

To facilitate real data analysis using IGREX, we investigated the sensitivity of IGREX using the 9 proteins with significant GREX components in at least one tissue type (same as those analyzed in the main text). We first analyzed the influence of SNP mismatch between the GTEx reference and the GWAS dataset. Supplementary Figure S24 shows the $\widehat{\text{PVE}}_{\text{GREX}}$ obtained by randomly removing 10%, 30%, 50% and 70% of SNPs in the GTEx reference panel. The $\widehat{\text{PVE}}_{\text{GREX}}$ remains accurate when the missing rate $\leq 50\%$, indicating that IGREX is robust to the mismatch of SNPs between eQTL and GWAS data. This is because information of the randomly removed SNPs has been largely captured by remaining SNPs due to the existence of LD.

We also inspected the sensitivity of IGREX estimate to the number of genes by randomly removed 10%, 20% and 30% genes in the GTEx reference panel. We used the protein CD96 to demonstrate the results. As shown in Supplementary Figure S25, the $\widehat{\text{PVE}}_{\text{GREX}}$ remains

accurate when the missing rate $\leq 20\%$, indicating that IGREX is also quite robust to mismatch of genes.

Next, we evaluated the influence of varying number of cis-SNPs using the pQTL summary statistics of CD96 protein. By default, the cis-SNPs used in IGREX are within 500Kb of the transcription start and end of each protein coding genes. We chose alternative bandwidths, i.e., 400Kb and 600Kb, and compared the obtained $\widehat{\text{PVE}}_{\text{GREX}}$'s of proteins with those obtained with 500Kb. The results are shown in Figure S26. As we can observe, the choice of bandwidth has little influence on the IGREX estimation.

To analyze the influence of imputation, we first phased the GTEx genotypes using SHAPEIT software with about $900,000$ SNPs of high quality. Then, we used the IMPUTE2 software to impute the GTEx SNPs. By removing the SNPs with low INFO scores (≤ 0.5), we obtained 17,181,884 SNPs of satisfactory imputation quality. After stringent QC, about $3,000,000$ SNPs are remained for analysis. We then compared the $\widehat{\text{PVE}}_{\text{GREX}}$'s of protein CD96 across 48 tissues obtained from the original 900K SNPs with those obtained from the 3M imputed SNPs. As shown in Figure S27, the results are nearly the same. However, we do not recommend to use more than 5 million imputed SNPs to estimate the GREX component. There are two reasons: 1. the sample size of the eQTL reference panel is too small, e.g, the average sample size of the GTEx project is only about 200. 2. Too many highly correlated SNPs will be involved in the prediction of gene expression (stage one of IGREX). Given limited sample size, high correlation can produce less accurate prediction of gene expression and further lead to unerestimation of the GREX component.

2.7 Application to metabolite traits

Metabolic phenotypes serve as important intermediate traits in high level biological processes. To understand the role of gene regulation in the genetics of such traits, we applied IGREX-s to a summary level data set of circulating metabolites [8], which was comprised of meta-analysis of 123 metabolites. We focused our analysis on the 21 metabolites that were highly heritable (estimated $h_t^2 > 10\%$) including glycine, various features of HDL, LDL, very low-density lipoprotein (VLDL) and intermediate-density lipoprotein (IDL) and other polyunsaturated fatty acids (otPUFA). The distributions of $\text{PVE}_{\text{GREX}}/h_t^2$ estimates in different tissues are given in Supplementary Fig. S30a. The median values of percentage estimates are higher than 10% in 6 out of the 48 tissues and only higher than 15% in liver and spinal cord (cervical c-1).

According to the estimated values shown in the heat map of Supplementary Fig. S30b, we can see that the features associated with IDL, LDL and VLDL have estimated $\text{PVE}_{\text{GREX}}/h_t^2$ around 20% in liver and 16% in spinal cord, suggesting that they are more related to the GREX effects in these two tissues. On the other hand, there is no signal of GREX components detected under the nominal level 0.05 in any GTEx tissue for HDL associated features or glycine. We note that the estimated values for LDL are not significantly different from the ones observed in NFBC analysis, implying the consistency between the two studies.

2.8 Details of trans-eQTLs and alternative splicing analysis

The trans-eQTLs used in analysis are reported by another eQTL study, the eQTLGen consortium. After matching with the SNPs in both the GWAS (or the LD reference panel) and the eQTL reference panel, around 20,000 trans-eQTLs comprised of 4,000 genes and 1600 SNPs were incorporated in the model for protein traits and the two schizophrenia datasets. In analyzing HDL, 8,072 trans-associations from 2,817 genes and 663 SNPs were included. In diseases from WTCCC dataset, around 5,700 trans-associations from 2,300 genes and 350 SNPs were considered.

In the analysis of sQTL data, we used the alternative splicing data from GTEx as reference and applied IGREX-MoM to four representative trait-tissue pairs: LDL-liver, TC-liver, SCZ-amygdala and SCZ-cerebellar hemisphere. The LDL and TC datasets were from the NFBC study. SCZ2, the dataset with the largest sample size, was used for SCZ. After matching with the GWAS, 58,272 splicing events were analyzed for LDL and TC while 57,509 splicing events were analyzed for SCZ.

References

[1] Xiang Zhou. A unified framework for variance component estimation with summary statistics in genome-wide association studies. *The annals of applied statistics*, 11(4):2027, 2017.

[2] Chuanhai Liu, Donald B Rubin, and Ying Nian Wu. Parameter expansion to accelerate EM: the PX-EM algorithm. *Biometrika*, 85(4):755–770, 1998.

[3] Hua Zhou, Liuyi Hu, Jin Zhou, and Kenneth Lange. MM algorithms for variance components models. *Journal of Computational and Graphical Statistics*, pages 1–12, 2019.

[4] Tomaz Berisa and Joseph K Pickrell. Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics*, 32(2):283, 2016.

[5] Robert B Davies. The distribution of a linear combination of $\chi 2$ random variables. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 29(3):323–333, 1980.

[6] Brendan K Bulik-Sullivan, Po-Ru Loh, Hilary K Finucane, Stephan Ripke, Jian Yang, Nick Patterson, Mark J Daly, Alkes L Price, Benjamin M Neale, Schizophrenia Working Group of the Psychiatric Genomics Consortium, et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature genetics*, 47(3):291, 2015.

[7] Qiongshi Lu, Boyang Li, Derek Ou, Margret Erlendsdottir, Ryan L Powles, Tony Jiang, Yiming Hu, David Chang, Chentian Jin, Wei Dai, Qidu He, Zefeng Liu, Shubhabrata Mukherjee, Paul K Crane, and Hongyu Zhao. A powerful approach to estimating annotation-stratified genetic covariance via GWAS summary statistics. *The American Journal of Human Genetics*, 101(6):939–964, 2017.

[8] Johannes Kettunen, Ayşe Demirkan, Peter Würtz, Harmen HM Draisma, Toomas Haller, Rajesh Rawal, Anika Vaarhorst, Antti J Kangas, Leo-Pekka Lyytikäinen, Matti Pirinen, et al. Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nature communications*, 7:11122, 2016.