Supporting Information for "Bayesian Shrinkage Estimation of High Dimensional Causal Mediation Effects in Omics Studies"

by

Yanyi Song[1], **Xiang Zhou**[1,*], **Min Zhang**[1], **Wei Zhao**[2], **Yongmei Liu**[3], **Sharon L. R. Kardia**[2], **Ana V. Diez Roux**[4], **Belinda L. Needham**[2], **Jennifer A. Smith**[2], **and Bhramar Mukherjee**[1,**]

[1]Department of Biostatistics, University of Michigan, Ann Arbor, MI, U.S.A.

[2]Department of Epidemiology, University of Michigan, Ann Arbor, MI, U.S.A.

[3]Division of Cardiology, Department of Medicine, Duke University School of Medicine, Durham, NC, U.S.A.

[4]Department of Epidemiology and Biostatistics, Drexel University, Philadelphia, PA, U.S.A.

[*] *email*: xzhousph@umich.edu

[**] *email*: bhramar@umich.edu

1 Detailed Proofs

1.1 Proof of deducting average NDE and the average NIE from observed data Given the assumptions in Section 2, we can express $E[Y_i(a, \boldsymbol{M_i}(a^\star))|\boldsymbol{C_i}]$ as below,

$$
\begin{aligned}
& E[Y_i(a, \boldsymbol{M_i}(a^\star))|\boldsymbol{C_i}] \\
= & \int_{\boldsymbol{m}} E[Y_i(a, \boldsymbol{m})|\boldsymbol{C_i}, \boldsymbol{M_i}(a^\star) = \boldsymbol{m}] P(\boldsymbol{M_i}(a^\star) = \boldsymbol{m}|\boldsymbol{C_i}) \mathrm{d}\boldsymbol{m} \\
= & \int_{\boldsymbol{m}} E[Y_i(a, \boldsymbol{m})|\boldsymbol{C_i}] P(\boldsymbol{M_i}(a^\star) = \boldsymbol{m}|\boldsymbol{C_i}, A_i = a^\star) \mathrm{d}\boldsymbol{m} \\
& \text{(by assumption (4) \& (3))} \\
= & \int_{\boldsymbol{m}} E[Y_i(a, \boldsymbol{m})|A_i = a, \boldsymbol{C_i}] P(\boldsymbol{M_i}(a^\star) = \boldsymbol{m}|\boldsymbol{C_i}, A_i = a^\star) \mathrm{d}\boldsymbol{m} \\
& \text{(by assumption (1))} \\
= & \int_{\boldsymbol{m}} E[Y_i(a, \boldsymbol{m})|A_i = a, \boldsymbol{M_i} = \boldsymbol{m}, \boldsymbol{C_i}] P(\boldsymbol{M_i} = \boldsymbol{m}|\boldsymbol{C_i}, A_i = a^\star) \mathrm{d}\boldsymbol{m} \\
& \text{(by assumption (2) and consistency)} \\
= & \int_{\boldsymbol{m}} E(Y_i|a, \boldsymbol{m}, \boldsymbol{C_i}) P(\boldsymbol{M_i} = \boldsymbol{m}|\boldsymbol{C_i}, a^\star) \mathrm{d}\boldsymbol{m} \\
& \text{(by consistency)}
\end{aligned}
$$

If we replace a with $a^\star$ in $E[Y_i(a, \boldsymbol{M_i}(a^\star))|\boldsymbol{C_i}]$, then we get $E[Y_i(a^\star, \boldsymbol{M_i}(a^\star))|\boldsymbol{C_i}] = \int_{\boldsymbol{m}} E(Y_i|a^\star, \boldsymbol{m}, \boldsymbol{C_i}) \times P(\boldsymbol{M_i} = \boldsymbol{m}|\boldsymbol{C_i}, a^\star)\mathrm{d}\boldsymbol{m}$. Therefore, we can express the average natural direct effect conditional on C as,

$$\begin{aligned} & E[Y_i(a, \boldsymbol{M_i}(a^\star)) - Y_i(a^\star, \boldsymbol{M_i}(a^\star))|\boldsymbol{C_i}] \\ = & \int_{\boldsymbol{m}} \{E(Y_i|a, \boldsymbol{m}, \boldsymbol{C_i}) - E(Y_i|a^\star, \boldsymbol{m}, \boldsymbol{C_i})\} P(\boldsymbol{M_i} = \boldsymbol{m}|\boldsymbol{C_i}, a^\star)\mathrm{d}\boldsymbol{m}. \end{aligned} \tag{1}$$

If we replace $a^\star$ with a in $E[Y_i(a, \boldsymbol{M_i}(a^\star))|\boldsymbol{C_i}]$, then we get $E[Y_i(a, \boldsymbol{M_i}(a))|\boldsymbol{C_i}] = \int_{\boldsymbol{m}} E(Y_i|a, \boldsymbol{m}, \boldsymbol{C_i}) \times P(\boldsymbol{M_i} = \boldsymbol{m}|\boldsymbol{C_i}, a)\mathrm{d}\boldsymbol{m}$, and thus the average indirect effect conditional on C is given by,

$$\begin{aligned} & E[Y_i(a, \boldsymbol{M_i}(a)) - Y_i(a, \boldsymbol{M_i}(a^\star))|\boldsymbol{C_i}] \\ = & \int_{\boldsymbol{m}} E(Y_i|a, \boldsymbol{m}, \boldsymbol{C_i}) \{P(\boldsymbol{M_i} = \boldsymbol{m}|\boldsymbol{C_i}, a) - P(\boldsymbol{M_i} = \boldsymbol{m}|\boldsymbol{C_i}, a^\star)\}\mathrm{d}\boldsymbol{m}. \end{aligned} \tag{2}$$

Finally, one can get the average NDE and NIE by taking expectation over C of the two conditional effects defined in (1) and (2). Importantly, Equation (1) and (2) show that if the assumptions in Section 2 hold, the average NDE and the average NIE can then be identified by modeling $Y_i|A_i, \boldsymbol{M_i}, \boldsymbol{C_i}$ and $\boldsymbol{M_i}|A_i, \boldsymbol{C_i}$ using observed data.

1.2 Proof of Equation (3), (4)

$$
\begin{aligned}
\text{NDE:} \quad & E[Y_i(a, \boldsymbol{M_i}(a^\star)) - Y_i(a^\star, \boldsymbol{M_i}(a^\star))|\boldsymbol{C_i}] \\
= \; & \int_{\boldsymbol{m}} \{E(Y_i|a, \boldsymbol{m}, \boldsymbol{C_i}) - E(Y_i|a^\star, \boldsymbol{m}, \boldsymbol{C_i})\} P(\boldsymbol{M_i} = \boldsymbol{m}|\boldsymbol{C_i}, a^\star) \mathrm{d}\boldsymbol{m} \\
= \; & \int_{\boldsymbol{m}} (a\beta_a - a^\star\beta_a) P(\boldsymbol{M_i} = \boldsymbol{m}|\boldsymbol{C_i}, a^\star) \mathrm{d}\boldsymbol{m} \\
= \; & \beta_a(a - a^\star) \\
\text{NIE:} \quad & E[Y_i(a, \boldsymbol{M_i}(a)) - Y_i(a, \boldsymbol{M_i}(a^\star))|\boldsymbol{C_i}] \\
= \; & \int_{\boldsymbol{m}} E(Y_i|a, \boldsymbol{m}, \boldsymbol{C_i}) \{P(\boldsymbol{M_i} = \boldsymbol{m}|\boldsymbol{C_i}, a) - P(\boldsymbol{M_i} = \boldsymbol{m}|\boldsymbol{C_i}, a^\star)\} \mathrm{d}\boldsymbol{m} \\
= \; & \int_{\boldsymbol{m}} (\boldsymbol{m}^T\boldsymbol{\beta_m} + a\beta_a + \boldsymbol{C_i}^T\boldsymbol{\beta_c}) \{P(\boldsymbol{M_i} = \boldsymbol{m}|\boldsymbol{C_i}, a) \\
& - P(\boldsymbol{M_i} = \boldsymbol{m}|\boldsymbol{C_i}, a^\star)\} \mathrm{d}\boldsymbol{m} \\
= \; & \{E(\boldsymbol{M_i}|\boldsymbol{C_i}, a) - E(\boldsymbol{M_i}|\boldsymbol{C_i}, a^\star)\}^T \boldsymbol{\beta_m} \\
= \; & \boldsymbol{\alpha_a}^T \boldsymbol{\beta_m} (a - a^\star) \\
= \; & (a - a^\star) \sum_{j=1}^{p} (\boldsymbol{\alpha_a})_j (\boldsymbol{\beta_m})_j
\end{aligned}
$$

2 Posterior Sampling Algorithm Details for Bayesian Mediation Analysis

Let $\boldsymbol{\theta_1} = (\boldsymbol{\beta_m}, \beta_a, \pi_m, \boldsymbol{r_m}, \sigma^2_{m1}, \sigma^2_{m0}, \sigma^2_a, \sigma^2_e)$ denote all the unknown parameters in the outcome model, and $\boldsymbol{\theta_2} = (\boldsymbol{\alpha_a}, \pi_a, \boldsymbol{r_a}, \sigma^2_{ma1}, \sigma^2_{ma0}, \sigma^2_g)$ for the mediator model. The joint log posterior distribution is,

$$
\begin{aligned}
& \log P(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}|(Y_i, \boldsymbol{M_i}, A_i)_{i=1}^n) \\
\propto \; & \sum_{i=1}^{n} \log P(Y_i|\boldsymbol{\theta_1}, A_i, \boldsymbol{M_i}) + \sum_{i=1}^{n} \log P(\boldsymbol{M_i}|\boldsymbol{\theta_2}, A_i) + \log P(\boldsymbol{\theta_1}) + \log P(\boldsymbol{\theta_2})
\end{aligned}
$$

Sampling β_{mj} and r_{mj}

$$
\log p(\beta_{mj}|r_{mj} = 1, .) = -\frac{\beta^2_{mj}}{2\sigma^2_{m1}} - \sum_{i=1}^{n} \{\frac{(M_i^{(j)}\beta_{mj})^2}{2\sigma^2_e} + \sigma_e^{-2} M_i^{(j)} (Y_i - A_i\beta_a - \sum_{s \neq j} M_i^{(s)}\beta_{ms} - \boldsymbol{C_i}^T\boldsymbol{\beta_c})\beta_{mj}\}
$$

$$
p(\beta_{mj}|r_{mj} = z, .) = N(\mu_{mjz}, s^2_{mjz}), z = 0, 1
$$

$$\mu_{mjz} = \frac{\sum_{i=1}^{n} M_i^{(j)}(Y_i - A_i\beta_a - \sum_{s\neq j} M_i^{(s)}\beta_{ms} - \boldsymbol{C_i}^T\boldsymbol{\beta_c})}{\sigma_e^2/\sigma_{mz}^2 + \sum_{i=1}^{n}(M_i^{(j)})^2}, s_{mjz}^2 = \frac{1}{1/\sigma_{mz}^2 + \sum_{i=1}^{n}(M_i^{(j)})^2/\sigma_e^2}$$

$$p(r_{mj} = z|.) \propto \exp(\mu_{mjz}^2/2s_{mjz}^2 + \log(s_{mjz}) - \log(\sigma_{mz}) + \log(p(r_{mj} = z)))$$

Sampling β_a

$$\log p(\beta_a|.) = -\frac{\beta_a^2}{2\sigma_a^2} - \sum_{i=1}^{n}\{\frac{(A_i\beta_a)^2}{2\sigma_e^2} + \sigma_e^{-2}A_i(Y_i - \boldsymbol{M_i}^T\boldsymbol{\beta_m} - \boldsymbol{C_i}^T\boldsymbol{\beta_c})\beta_a\}$$

$$p(\beta_a|.) = N(\mu_a, s_a^2)$$

$$\mu_a = \frac{\sum_{i=1}^{n} A_i(Y_i - \boldsymbol{M_i}^T\boldsymbol{\beta_m} - \boldsymbol{C_i}^T\boldsymbol{\beta_c})}{\sigma_e^2/\sigma_a^2 + \sum_{i=1}^{n} A_i^2}, s_a^2 = \frac{1}{1/\sigma_a^2 + \sum_{i=1}^{n} A_i^2/\sigma_e^2}$$

Sampling α_{aj} *and* r_{aj}

$$\log p(\alpha_{aj}|r_{aj} = 1, .) = -\frac{\alpha_{aj}^2}{2\sigma_{ma1}^2} - \sum_{i=1}^{n}\{\frac{(A_i\alpha_{aj})^2}{2\sigma_g^2} + \sigma_g^{-2}A_i(M_i^{(j)} - (\boldsymbol{\alpha_c}\boldsymbol{C_i})_j)\alpha_{aj}\}$$

$$p(\alpha_{aj}|r_{aj} = z, .) = N(\frac{\sum_{i=1}^{n} A_i(M_i^{(j)} - (\boldsymbol{\alpha_c}\boldsymbol{C_i})_j)}{\sigma_g^2/\sigma_{maz}^2 + \sum_{i=1}^{n} A_i^2}, \frac{1}{1/\sigma_{maz}^2 + \sum_{i=1}^{n} A_i^2/\sigma_g^2})$$

$$p(r_{aj} = z|.) \propto \exp(\mu_{\alpha jz}^2/2s_{\alpha jz}^2 + \log(s_{\alpha jz}) - \log(\sigma_{maz}) + \log(p(r_{aj} = z))), z = 0, 1$$

Sampling $\sigma_{m1}^2, \sigma_{m0}^2$

$$\log p(\sigma_{m1}^2|.) = -(\frac{\sum_{j=1}^{q} r_{mj}}{2} + k_m + 1)\log(\sigma_{m1}^2) - (\frac{\sum_{j=1}^{q} r_{mj}\beta_{mj}^2}{2} + l_m)\sigma_{m1}^{-2}$$

$$\log p(\sigma_{m0}^2|.) = -(\frac{\sum_{j=1}^{q}(1 - r_{mj})}{2} + k_m + 1)\log(\sigma_{m0}^2) - (\frac{\sum_{j=1}^{q}(1 - r_{mj})\beta_{mj}^2}{2} + l_m)\sigma_{m0}^{-2}$$

$$p(\sigma_{m1}^2|.) \sim \text{inverse-gamma}(\frac{\sum_{j=1}^{q} r_{mj}}{2} + k_m, \frac{\sum_{j=1}^{q} r_{mj}\beta_{mj}^2}{2} + l_m)$$

$$p(\sigma_{m0}^2|.) \sim \text{inverse-gamma}(\frac{\sum_{j=1}^{q}(1 - r_{mj})}{2} + k_m, \frac{\sum_{j=1}^{q}(1 - r_{mj})\beta_{mj}^2}{2} + l_m)$$

Sampling σ_a^2

$$\log p(\sigma_a^2|.) = -(\frac{1}{2} + k_a + 1)\log(\sigma_a^2) - (\frac{\beta_a^2}{2} + l_a)\sigma_a^{-2}$$

$$p(\sigma_a^2|.) \sim \text{inverse-gamma}(\frac{1}{2} + k_a, \frac{\beta_a^2}{2} + l_a)$$

Sampling $\sigma_{ma1}^2, \sigma_{ma0}^2$

$$\log p(\sigma_{ma1}^2|.) = -(\frac{\sum_j r_{aj}}{2} + k_{ma} + 1)\log(\sigma_{ma1}^2) - (\frac{\sum_j r_{aj}\alpha_{aj}^2}{2} + l_{ma})\sigma_{ma1}^{-2}$$

$$\log p(\sigma_{ma0}^2|.) = -(\frac{\sum_j (1 - r_{aj})}{2} + k_{ma} + 1)\log(\sigma_{ma0}^2) - (\frac{\sum_j (1 - r_{aj})\alpha_{aj}^2}{2} + l_{ma})\sigma_{ma0}^{-2}$$

$$p(\sigma_{ma1}^2|.) \sim \text{inverse-gamma}(\frac{\sum_j r_{aj}}{2} + k_{ma}, \frac{\sum_j r_{aj}\alpha_{aj}^2}{2} + l_{ma})$$

$$p(\sigma_{ma0}^2|.) \sim \text{inverse-gamma}(\frac{\sum_j (1 - r_{aj})}{2} + k_{ma}, \frac{\sum_j (1 - r_{aj})\alpha_{aj}^2}{2} + l_{ma})$$

Sampling σ_e^2

$$\log p(\sigma_e^2|.) = -(\frac{n}{2} + k_e + 1)\log(\sigma_e^2) - (\frac{\sum_{i=1}^n (Y_i - \boldsymbol{M_i}^T \boldsymbol{\beta_m} - A_i\beta_a - \boldsymbol{C_i}^T \boldsymbol{\beta_c})^2}{2} + l_e)\sigma_e^{-2}$$

$$p(\sigma_e^2|.) \sim \text{inverse-gamma}(\frac{n}{2} + k_e, \frac{\sum_{i=1}^n (Y_i - \boldsymbol{M_i}^T \boldsymbol{\beta_m} - A_i\beta_a - \boldsymbol{C_i}^T \boldsymbol{\beta_c})^2}{2} + l_e)$$

Sampling σ_g^2

$$\log p(\sigma_g^2|.) = -(\frac{qn}{2} + k_g + 1)\log(\sigma_g^2) - (\frac{\sum_{i=1}^n \sum_q (M_i^{(q)} - A_i\alpha_{aq} - (\boldsymbol{\alpha_c}\boldsymbol{C_i})_q)^2}{2} + l_g)\sigma_g^{-2}$$

$$p(\sigma_g^2|.) \sim \text{inverse-gamma}(\frac{qn}{2} + k_g, (\frac{\sum_{i=1}^n \sum_q (M_i^{(q)} - A_i\alpha_{aq} - (\boldsymbol{\alpha_c}\boldsymbol{C_i})_q)^2}{2} + l_g))$$

Sampling β_{cw}

$$\log p(\beta_{cw}|.) = -\sum_{i=1}^n \{\frac{(C_{iw}\beta_{cw})^2}{2\sigma_e^2} + \sigma_e^{-2} C_{iw}(Y_i - \boldsymbol{M_i}^T \boldsymbol{\beta_m} - A_i\beta_a - \sum_{s \neq w} C_{iw}\beta_{cw})\beta_{cw}\}$$

$$p(\beta_{cw}|.) = N(\frac{\sum_{i=1}^n C_{iw}(Y_i - A_i\beta_a - \boldsymbol{M_i}^T \boldsymbol{\beta_m} - \sum_{s \neq w} C_{iw}\beta_{cw})}{\sum_{i=1}^n C_{iw}^2}, \frac{\sigma_e^2}{\sum_{i=1}^n C_{iw}^2})$$

Sampling $(\boldsymbol{\alpha_{cw}})_j$

$$\log p((\boldsymbol{\alpha_{cw}})_j|.) = -\sum_{i=1}^{n}\{\frac{(C_{iw}(\boldsymbol{\alpha_{cw}})_j)^2}{2\sigma_g^2} + \sigma_g^{-2}C_{iw}(M_i^{(j)} - A_i\alpha_{aj} - \sum_{s\neq w} C_{is}(\boldsymbol{\alpha_{cs}})_j)(\boldsymbol{\alpha_{cw}})_j\}$$

$$p((\boldsymbol{\alpha_{cw}})_j|.) = N(\frac{\sum_{i=1}^{n} C_{iw}(M_i^{(j)} - A_i\alpha_{aj} - \sum_{s\neq w} C_{is}(\boldsymbol{\alpha_{cs}})_j)}{\sum_{i=1}^{n} C_{iw}^2}, \frac{\sigma_g^2}{\sum_{i=1}^{n} C_{iw}^2})$$

Sampling π_m, π_a

For π_m, π_a, their conditional distributions don't appear to be of any known form, so we use a random-walk standard Metropolis-Hastings algorithm to draw posterior samples of them.

As for the proposal distribution, we update the parameters by adding a random variable from U(-0.07, 0.07) to the current value. New values that lie outside the boundary [0,1] are reflected back.

3 Convergence Diagnosis

We used the potential scale reduction factor (PSRF) [1] to quantify the mixing property of the proposed MCMC algorithm. With multiple MCMC chains, PSRF for a parameter is essentially the ratio between the overall-chain variance and the average within-chain variance. A PSRF value in the range of (0.9, 1.2) suggests that the MCMC algorithm has good mixing property and the posterior samples converge well. As an example, in Figure 1, we present the PSRFs for the PIPs of 60 top significant mediators identified from univariate analysis in the baseline simulation setting with the number of mediators $p = 2,000$. We find that all the PSRFs from our MCMC algorithm fall within (0.9, 1.2), which indicates the good mixing property of our algorithm.

4 Power Comparison with Spike-and-slab Priors and Horseshoe Priors

Both horseshoe priors and spike-and-slab priors [2, 3] are widely used methods for Bayesian shrinkage, and it is natural to apply them to the two regression models in high-dimensional mediation analysis. The horseshoe prior can be represented as a scale mixture of normals, with the mixing distribution being a standard half-Cauchy distribution. The horseshoe prior

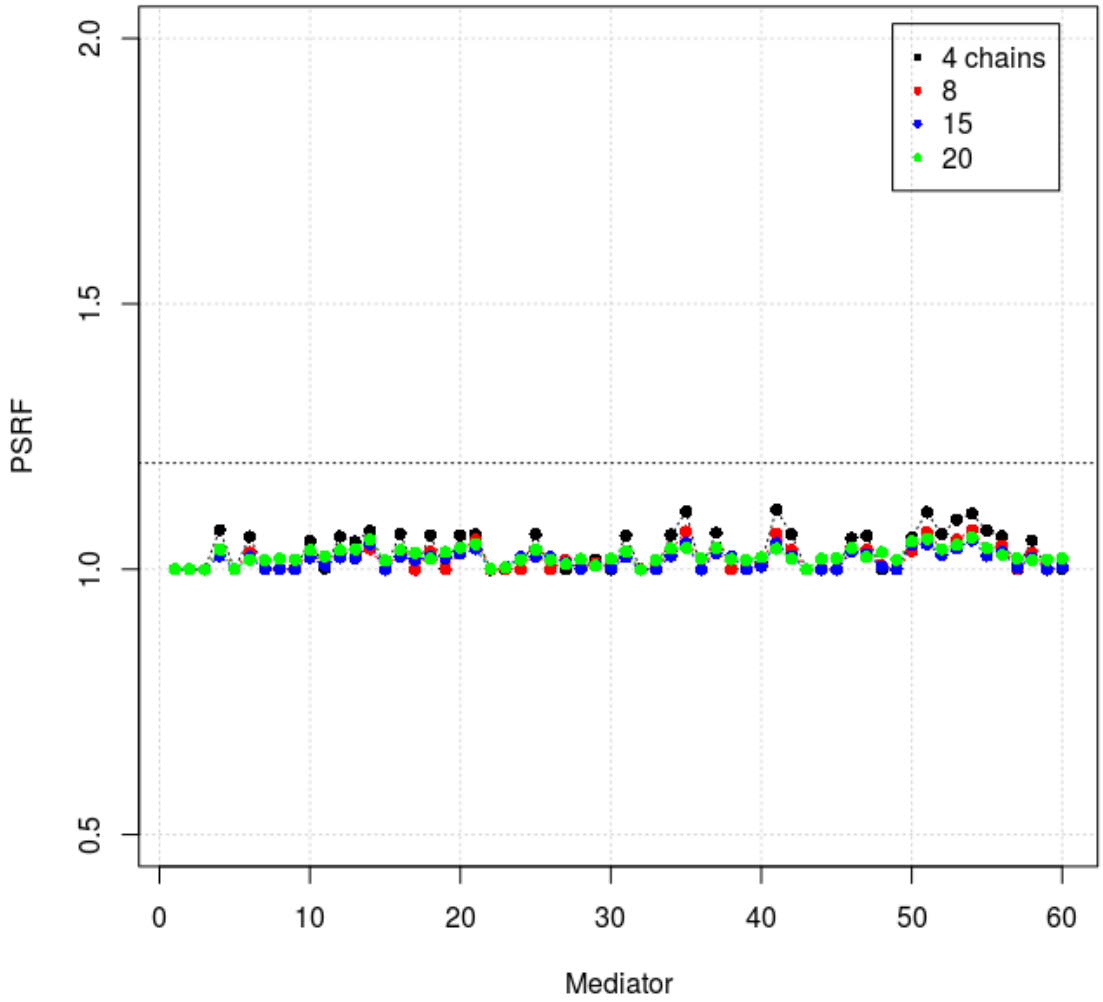

Figure 1: Potential scale reduction factors (PSRF) of the Bayesian posterior inclusion probabilities of 60 top marginally significant mediators with 3, 8, 15, and 20 MCMC chains, where PSRF within (0.9, 1.2) suggests good mixing property.

is not a discrete mixture prior and therefore it does not directly categorize mediators into one actively mediating group and three inactive (null) groups. To achieve categorization and selection of mediators, one can use the shrinkage factors [3] for the coefficients and develop a thresholding rule on the continuous values to determine inclusion or not. For the spike-and-slab prior, we can directly use the posterior inclusion probability to perform mediator selection.

We implemented both horseshoe priors and spike-and-slab priors for high-dimensional mediation analysis and compared them with our method in simulations. We first simulated $(\boldsymbol{\beta_m})_j$ $(j = 1, ..., p)$ from a mixture of two normals: $\pi_m N(0, 1) + (1 - \pi_m)N(0, 0.001)$, and $(\boldsymbol{\alpha_a})_j$ $(j = 1, ..., p)$ from $\pi_a N(0, 1) + (1 - \pi_a)N(0, 0.001)$. The other configurations are same as the baseline setting for $p = 2,000$ in the main paper. We find that our Bayesian method with normal-normal priors outperforms the other two methods. For example, when $PVE_{IE} = 0.8$, at 0.01 FPR, our method achieves a power of 0.528, while the methods with point-normal priors and horseshoe priors have a power of 0.484 and 0.467, respectively. The full results are shown in Table 5.

p	Setting	Normal-normal priors	Point-normal priors	Horseshoe priors
2,000	$PVE_A = 0.3$	0.509	0.461	0.437
	$PVE_A = 0.5$	0.474	0.424	0.461
	$PVE_A = 0.8$	0.512	0.413	0.479
	$PVE_{IE} = 0.2$	0.473	0.415	0.453
	$PVE_{IE} = 0.4$	0.474	0.424	0.461
	$PVE_{IE} = 0.8$	0.528	0.484	0.467
	$\pi_a = 0.03$	0.474	0.424	0.461
	$\pi_a = 0.1$	0.146	0.131	0.092
	$\pi_a = 0.25$	0.072	0.062	0.042
	$\pi_m = 0.02$	0.474	0.424	0.461
	$\pi_m = 0.1$	0.471	0.420	0.454
	$\pi_m = 0.25$	0.462	0.401	0.440

Table 1: Power comparison among our normal-normal priors, the point-normal priors and the horseshoe priors when $p = 2,000, n = 1,000$ and the effect sizes are sampled from a mixture of two normals. In each setting, we change one parameter at a time from the baseline setting. We calculate the true positive rate (TPR) for the power comparison. The average TPR at FPR = 0.01 is calculated across 200 replicates for each simulation scenario.

In addition, we also performed simulations in which $(\boldsymbol{\beta_m})_j$ $(j = 1, ..., p)$ is from a point-normal prior: $\pi_m N(0,1) + (1-\pi_m)\delta_0$, and $(\boldsymbol{\alpha_a})_j$ $(j = 1, ..., p)$ from $\pi_a N(0,1) + (1-\pi_a)\delta_0$, where δ_0 is a point mass at zero. The other configurations are same as the baseline setting for $p = 2,000$ in the main paper. Now the effect size distribution is not a mixture of two normals, and favors the spike-and-slab priors, as one might expect. The results (in Table 6) show that all the three methods have similar performance in most scenarios, and our method appears to have decent power when the effects are not polygenic and the signals are relatively sparse.

p	Setting	Normal-normal priors	Point-normal priors	Horseshoe priors
2,000	$PVE_A = 0.3$	0.525	0.526	0.465
	$PVE_A = 0.5$	0.483	0.503	0.490
	$PVE_A = 0.8$	0.470	0.513	0.493
	$PVE_{IE} = 0.2$	0.456	0.488	0.476
	$PVE_{IE} = 0.4$	0.483	0.503	0.490
	$PVE_{IE} = 0.8$	0.510	0.543	0.491
	$\pi_a = 0.03$	0.483	0.503	0.490
	$\pi_a = 0.1$	0.135	0.145	0.106
	$\pi_a = 0.25$	0.047	0.094	0.052
	$\pi_m = 0.02$	0.483	0.503	0.490
	$\pi_m = 0.1$	0.468	0.488	0.486
	$\pi_m = 0.25$	0.450	0.465	0.470

Table 2: Power comparison among our normal-normal priors, the point-normal priors and the horseshoe priors when $p = 2,000, n = 1,000$ and the effect sizes are sampled from point-normal priors. In each setting, we change one parameter at a time from the baseline setting. We calculate the true positive rate (TPR) for the power comparison. The average TPR at FPR = 0.01 is calculated across 200 replicates for each simulation scenario.

Therefore, our Bayesian mediation method with mixture of normals prior performs well in identifying active mediators in a wide range of scenarios, and is relatively robust to the effect size distribution, appearing to have decent power when the effects are not polygenic. Our method can distinguish the large effects from the small effects in the polygenic model, and is also more practically appealing than the horseshoe prior since it can directly categorize mediators into four possible groups without the need of specifying a thresholding rule.

5 Estimation of the Global Mediation Effects τ and the Four Proportions of Mediators

In this section, we examine the ability of our Bayesian mediation method to estimate the overall mediation effects and the proportion of mediators in the four different categories. We use π_{g1}, π_{g2}, π_{g3}, π_{g4} to represent the proportion of mediators in Group 1, Group 2, Group 3 and Group 4, respectively. We examine eight different simulation scenarios based on different combinations of π_{g1}, π_{g2}, π_{g3} and π_{g4}, which include four null scenarios with $\pi_{g1} = 0$ and four alternative scenarios with $\pi_{g1} \neq 0$. In these simulations, we set PVEs to be the same as in the baseline setting ($PVE_A = 0.5, PVE_{IE} = 0.4, PVE_{DE} = 0.1$; except when $\pi_{g4} = 1$ where PVE_A and PVE_{IE} are zero). We provide the estimated global mediation effects (τ) and proportion of active mediators (π_{g1}), as well as their 95% credible intervals in Table 3. We find that our method provides decent estimates for π_{g1} and τ across different scenarios, especially when $p = 100$. Note that our estimates for π_{g1} are slightly conservative due to the fact that our model does not have full power to detect all the mediators. The 95% credible intervals of τ also shows that the posterior distribution of τ is asymmetric and depends on the composition of the four groups.

We examine the posterior distribution of τ, which is bounded at zero and not symmetric. The distribution also depends on the composition of the four groups. In below we show a distribution graph (Figure 2) based on the posterior samples of τ in four different scenarios with $n = 1,000, p = 100$ as in Table 2 in the main manuscript.

6 The Runtime Comparison with HDMM

We performed simulations on a single core of Intel(R) Xeon(R) Platinum 8176 CPU @ 2.10GHz, and the runtime comparison of the proposed method relative to HDMM is shown in Table 4. The runtime of our method is comparable to that of HDMM. In the settings of $p = 2,000$, the effective dimension of the orthogonalized meditators in HDMM is 250, and the reduced dimension helps reduce the runtime compared to our method that uses the full dimension. While the runtime of our method is comparable with existing methods, we still

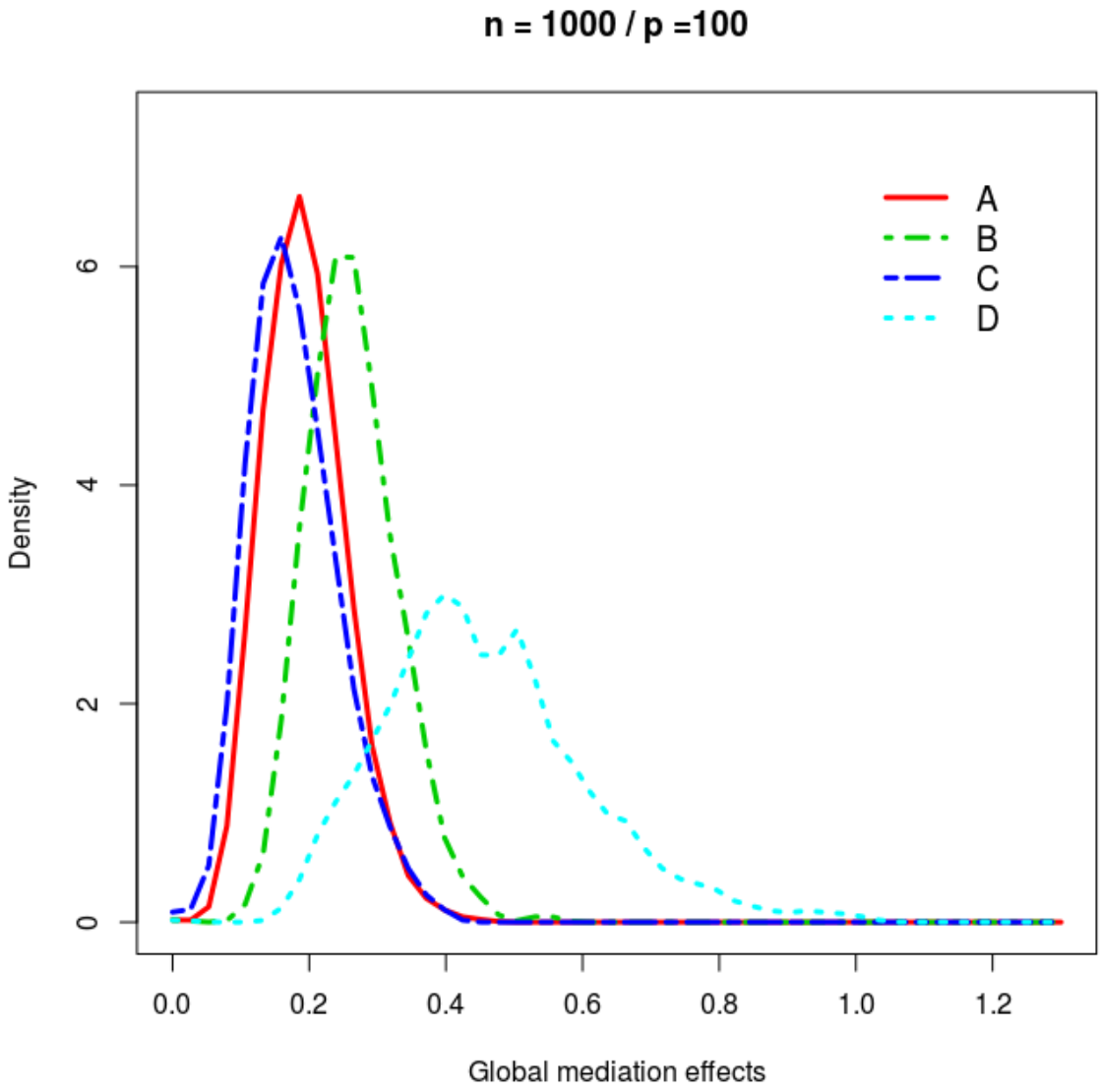

Figure 2: The distribution from the posterior samples of τ in four different scenarios with $n = 1000, p = 100$. We denote π_{g1}, π_{g2}, π_{g3} and π_{g4} to represent the proportion of mediators in Group 1, Group 2, Group 3 and Group 4 as defined in Table 1 in the main paper. The four settings are: A: $\pi_{g1} = 0.1, \pi_{g2} = 0.2, \pi_{g3} = 0.1, \pi_{g4} = 0.6$; B: $\pi_{g1} = 0.1, \pi_{g2} = 0.1, \pi_{g3} = 0.2, \pi_{g4} = 0.6$; C: $\pi_{g1} = 0.1, \pi_{g2} = 0.1, \pi_{g3} = 0.1, \pi_{g4} = 0.7$; D: $\pi_{g1} = 0.1, \pi_{g2} = 0, \pi_{g3} = 0, \pi_{g4} = 0.9$;

Table 3: Estimation of the global mediation effects τ under different compositions. We report the posterior mean ($\widehat{\tau}$) of τ and its 95% credible intervals when $p = 100/2,000$. We denote π_{g1}, π_{g2}, π_{g3} and π_{g4} to represent the proportion of mediators in Group 1, Group 2, Group 3 and Group 4 as defined in Table 1 of the main paper, and $\hat{\pi}_{g1}$ is the estimated proportion of active mediators from our Bayesian method. We also provide the 95% credible intervals (CI) for $\widehat{\tau}$ and $\hat{\pi}_{g1}$.

p	π_{g1}	π_{g2}	π_{g3}	π_{g4}	$\widehat{\pi_{g1}}$	(95% CI)	τ	$\widehat{\tau}$	(95% CI)
100	0	0.2	0.1	0.7	0.003	(0.000, 0.010)	0	0.025	(0.006, 0.066)
	0	0.1	0.2	0.7	0.003	(0.000, 0.010)	0	0.035	(0.006, 0.117)
	0	0.1	0.1	0.8	0.001	(0.000, 0.012)	0	0.009	(0.000, 0.012)
	0	0	0	1	0.001	(0.000, 0.008)	0	0.000	(0.000, 0.000)
100	0.1	0.2	0.1	0.6	0.064	(0.010, 0.092)	0.128	0.194	(0.100, 0.324)
	0.1	0.1	0.2	0.6	0.058	(0.030, 0.080)	0.249	0.263	(0.159, 0.400)
	0.1	0.1	0.1	0.7	0.078	(0.006, 0.110)	0.110	0.172	(0.014, 0.960)
	0.1	0	0	0.9	0.051	(0.040, 0.063)	0.961	0.458	(0.215, 0.792)
2,000	0	0.03	0.02	0.95	0.000	(0.000, 0.000)	0	0.230	(0.074, 0.565)
	0	0.1	0.02	0.88	0.000	(0.000, 0.000)	0	0.315	(0.149, 0.644)
	0	0.1	0.1	0.8	0.000	(0.000, 0.000)	0	0.211	(0.075, 0.509)
	0	0	0	1	0.000	(0.000, 0.000)	0	0.000	(0.000, 0.000)
2,000	0.01	0.02	0.01	0.96	0.001	(0.000, 0.003)	1.392	0.642	(0.408, 1.491)
	0.01	0.04	0.01	0.94	0.001	(0.000, 0.004)	1.273	0.436	(0.156, 0.981)
	0.01	0.09	0.09	0.81	0.001	(0.000, 0.003)	0.347	0.544	(0.217, 1.074)
	0.01	0	0	0.99	0.010	(0.002, 0.015)	0.103	0.113	(0.018, 0.334)

Table 4: The average runtime of the proposed method relative to HDMM for $n = 1,000$, $p = 100/2,000$ in the simulations. Comparison was carried out on a single core of Intel(R) Xeon(R) Platinum 8176 CPU @ 2.10GHz. For the proposed method, we ran 200,000 iterations for $p = 100$ and 500,000 iterations for $p = 2,000$.

Method	$n = 1,000$, $p = 100$	$n = 1,000$, $p = 2,000$
Bayesian continuous shrinkage	1.3min	98min
HDMM	4.5min	75min

acknowledge that future development of new algorithms and/or new methods will likely be required to scale our method to handle thousands of individuals and millions of mediators.

7 Simulations for $n = 100$

To enrich the simulation settings, we now examine how the method performs with a smaller sample size, i.e., $n = 100$. We included the results in Table 5 (for $n = 100$ and $p = 200$) and Table 6 (for $n = 100$ and $p = 50$). The results are consistent with simulations with large samples presented previously in the manuscript. In particular, we find that in most cases

our Bayesian method outperforms the other existing methods. The performance difference becomes more apparent in the presence of large PVE_{IE}, small PVE_A, small π_a and/or π_m. However, we do note that the small sample size here makes it difficult for all the methods to identify active mediators when $p > n$, resulting in small power and small power differences. This is to be expected. To achieve a decent performance with $n = 100, p = 200$ may require either a much higher PVE_{IE} (e.g. $PVE_{IE} = 0.99$) or a much sparser model (e.g. $\pi_a = 0.025$) as shown in Table 5. Our example at hand, the MESA has approximately 1,200 participants in the analytic dataset and thus we discuss this case in more details in the main text.

Table 5: Power comparison among our Bayesian mediation method, single mediation and HDMM when $p = 200, n = 100$ and the effect sizes are sampled from a mixture of two normals. The baseline setting is $PVE_A = 0.5, PVE_{IE} = 0.4, \pi_a = 0.075, \pi_m = 0.05$, and we simulate five truly active mediators. Within each block in the table, we change one parameter at a time from the baseline setting. We calculate the true positive rate (TPR) for the power comparison. The average TPR at FDR = 0.10 is calculated across 200 replicates for each simulation scenario.

Setting	Bayesian Method	Single Mediation	HDMM
Baseline	0.279	0.255	0.200
$PVE_A = 0.2$	0.275	0.254	0.200
$PVE_A = 0.8$	0.308	0.258	0.200
$PVE_{IE} = 0.2$	0.259	0.260	0.200
$PVE_{IE} = 0.8$	0.325	0.280	0.210
$PVE_{IE} = 0.99$	0.439	0.297	0.200
$\pi_a = 0.025$	0.567	0.470	0.240
$\pi_a = 0.125$	0.241	0.243	0.200
$\pi_m = 0.025$	0.290	0.283	0.200
$\pi_m = 0.125$	0.287	0.262	0.200

8 Detailed Description of MESA Data

MESA is a population-based longitudinal study designed to identify risk factors for the progression of subclinical cardiovascular disease (CVD) [4]. A total of 6,814 non-Hispanic white, African-American, Hispanic, and Chinese-American women and men aged 45−84 without clinically apparent CVD were recruited between July 2000 and August 2002 from the following 6 regions in the US: Forsyth County, NC; Northern Manhattan and the Bronx, NY; Baltimore City and Baltimore County, MD; St. Paul, MN; Chicago, IL; and Los Angeles

Table 6: Power comparison among our Bayesian mediation method, multivariate mediation, single mediation and HDMM when $p = 50, n = 100$ and the effect sizes are sampled from a mixture of two normals. The baseline setting is $PVE_A = 0.5, PVE_{IE} = 0.4, \pi_a = 0.3, \pi_m = 0.2$, and we simulate five truly active mediators. Within each block in the table, we change one parameter at a time from the baseline setting. We calculate the true positive rate (TPR) for the power comparison. The average TPR at FDR = 0.10 is calculated across 200 replicates for each simulation scenario.

Setting	Bayesian Method	Multivariate Mediation	Single Mediation	HDMM
Baseline	0.302	0.288	0.278	0.200
$PVE_A = 0.2$	0.293	0.294	0.278	0.200
$PVE_A = 0.8$	0.296	0.296	0.301	0.214
$PVE_{IE} = 0.2$	0.273	0.292	0.259	0.200
$PVE_{IE} = 0.8$	0.429	0.380	0.279	0.231
$\pi_a = 0.1$	0.522	0.458	0.465	0.267
$\pi_a = 0.5$	0.278	0.280	0.250	0.200
$\pi_m = 0.1$	0.341	0.327	0.296	0.216
$\pi_m = 0.5$	0.275	0.300	0.266	0.230

County, CA. Each field center recruited from locally available sources, which included lists of residents, lists of dwellings, and telephone exchanges. At Exam 1, respondents reported the highest level of education they completed. We created a dichotomous measure of respondent's educational attainment (less than college, low adult SES = 1; college degree or more = 0). The descriptive statistics for the exposure and outcome can be found in Table 8.

In the MESA data, between April 2010 and February 2012 (corresponding to MESA Exam 5), DNAm were assessed on a random subsample of 1,264 non-Hispanic white, African-American, and Hispanic MESA participants aged 55−94 from the Baltimore, Forsyth County, New York, and St. Paul field centers. After excluding respondents with missing data on one or more variables, we had phenotype and DNAm data from purified monocytes on a total of 1,231 individuals and we focused on this set of individuals for analysis. The detailed description of DNAm data collection, quantitation and data processing procedures can be found in Liu et al [5]. Briefly, the Illumina HumanMethylation450 BeadChip was used to measure DNAm, and bead-level data were summarized in GenomeStudio. Quantile normalization was performed using the *lumi* package with default settings [6]. Quality control (QC) measures included checks for sex and race/ethnicity mismatches and outlier identification by multidimensional scaling plots. Further probe filtering criteria included:

"detected" DNAm levels in <90% of MESA samples (detection p-value cut-off = 0.05), existence of a SNP within 10 base pairs of the target CpG site, overlap with a non-unique region, and suggestions by DMRcate [7] (mostly cross-reactive probes). Those procedures leave us 403,713 autosomal probes for analysis.

For computational reasons, we first selected a subset of CpG sites to be used in the final multivariate mediation analysis model. In particular, for each single CpG site in turn, we fit the following linear mixed model to test the marginal association between the CpG site and the exposure variable:

$$M_i = A_i\psi_a + \boldsymbol{C_{1i}}^T\boldsymbol{\psi_c} + \boldsymbol{Z_i}^T\boldsymbol{\psi_u} + \epsilon_i, i = 1, ..., n \tag{3}$$

where A_i represents adult SES value for the i'th individual and ψ_a is its coefficient; $\boldsymbol{C_{1i}}$ is a vector of covariates that include age, gender, race/ethnicity and enrichment scores for each of 4 major blood cell types (neutrophils, B cells, T cells and natural killer cells) to account for potential contamination by non-monocyte cell types; $\boldsymbol{Z_i}^T\boldsymbol{\psi_u}$ represent methylation chip and position random effects and are used to control for possible batch effects. The error term $\epsilon_i \sim MVN(0, \sigma^2 I_n)$ and is independent of the random effects. We obtained p-values for testing the null hypothesis $\psi_a = 0$ from the above model. We further applied the R/Bioconductor package BACON [8] to these p-values to further adjust for possible inflation using an empirical null distribution. Based on these marginal p-values, we selected top 2,000 CpG sites with the smallest p-values for our Bayesian multivariate analysis.

We implemented our proposed methods as well as methods with different prior specifications and HDMM on the MESA data. The current HDMM cannot handle covariates, so we apply it to the residuals after regressing the Y and $\boldsymbol{M}$ on the covariates. There may exist certain numerical stability issue with the HDMM on the MESA data, and the resulting weights of the first direction of mediation do not suggest obvious signal or pattern. The estimated first direction of mediation across the selected 2,000 sites is presented in Figure 3.

We also listed the top 2 sites and nearby genes identified by the four methods in the Table 7,

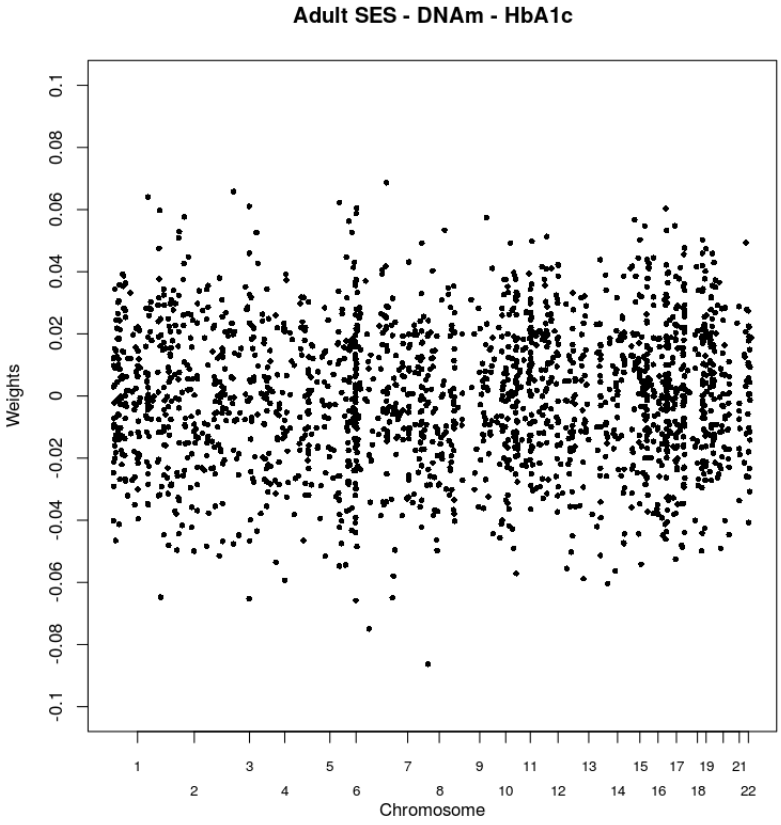

Figure 3: Consider the mediation trio: Adult SES → DNAm → HbA1c. The black dots are the weights for the first direction of mediation from HDMM for the selected 2,000 CpG sites across the genome.

Method	Top 2 sites	Nearby genes
Our method	cg19582614	CCND2
	cg04514392	CCDC54
Spike-and-Slab Priors	cg19582614	CCND2
	cg26610247	RP11-10J21.3
Horseshoe Priors	cg15531249	C16orf74
	cg15149205	TRIB1
HDMM	cg13488078	CLU
	cg12880602	MAP3K7

Table 7: The top 2 sites and their nearby genes identified by our proposed method as well as methods with different prior specifications and HDMM for mediation analysis on Adult SES $\rightarrow$ DNAm $\rightarrow$ HbA1c.

We do not know the truth in the real data so it is hard to evaluate the effectiveness of different methods here. In addition to the genes *CCDC54* and *CCND2*, which are associated with the outcome HbA1c as discussed in the main paper, *CLU* is associated with diabetes, probably through an increase in insulin resistance [9]. There is a lack of biological evidence to support a mediating role of the other genes.

After fitting the Bayesian mediation models, we then empirically check whether the functional forms of covariates in models 7 and 8 as linear terms in the main manuscript are reasonable. We perform the posterior predictive checks on the outcome model, and create the following graphical displays comparing the observed outcome to the replications drawn from the posterior predictive distribution.

In Figure 4, we compare the distribution of the observed outcomes HbA1c (y) and the kernel density estimates of replications of the outcome, y_{rep} from the posterior predictive distribution. The distributions of the outcomes randomly generated from the posterior predictive distribution largely resemble the true distribution of HbA1c in the data. The Bayesian predictive p-values [10] are 0.5 and 0.45 for the sample mean and variance, respectively, also suggesting adequate fit of the outcome model in terms of moments of the posterior predictive distribution.

We have discussed the identifiability assumptions required for causal inference in high-dimensional mediation analysis in Section 2 of the main paper. For Assumption 1 (no

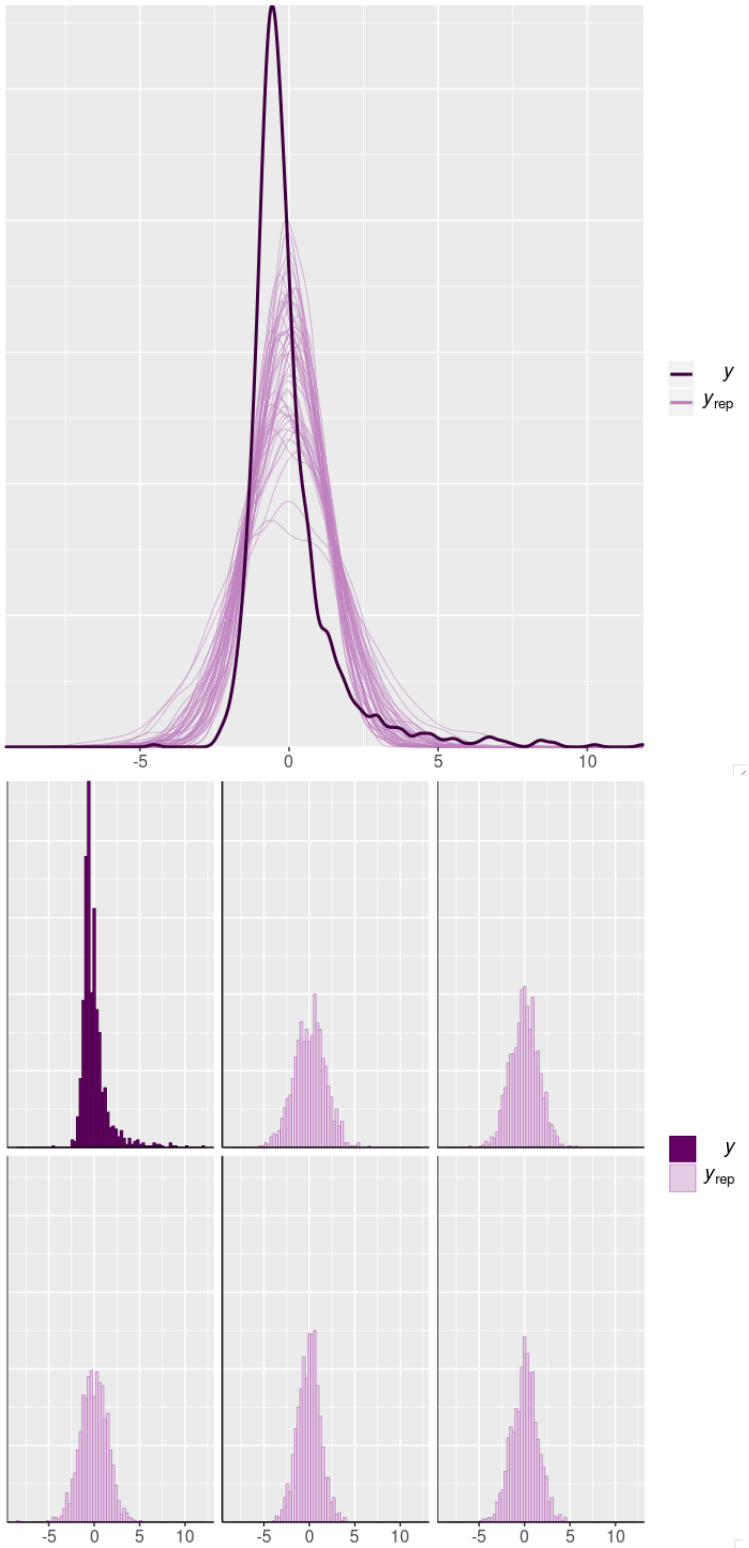

Figure 4: The top panel shows the distribution of the observed outcomes HbA1c (y, the dark line) and each of the 100 lighter lines is the kernel density estimate of a set of random draws of y_{rep} from the posterior predictive distribution of y. The bottom panel displays the separate histograms of y and of y_{rep} for five simulated datasets.

unmeasured exposure-outcome confounding), we believe that the available covariates of age, gender and race/ethnicity in model 7 are natural to control for confounders that are associated both to adult SES and HbA1c [11]. For Assumption 2 (no unmeasured mediator-outcome confounding given exposure), within each of the two exposure groups (low and high adult SES), age, gender and race/ethnicity are also natural to control for mediator-outcome confounding based on existing domain-specific literature [12]. For Assumption 3 (no unmeasured exposure-mediator confounding), we included all the important potential exposure-mediator confounders in model 8 as in [13]. Assumption 4 (no mediator-outcome confounding affected by the exposure) is usually a challenging condition to justify. However, we note that Assumption 4, as well as Assumption 3, is only required for identifying natural direct/indirect effects and is not the main focus here as we focus on estimating controlled direct/indirect effects. The influence of violating the above identifiability assumptions can be assessed using sensitivity analysis, which has been well-developed for the single mediator setting [14], and additional work is required to extend that approach to the sparse high-dimensional setting. Regarding the temporal assumptions, in MESA, adult SES (exposure) was collected in Exam 1 between July 2000 and August 2002, and DNAm (mediators) and HbA1c (outcome) were assessed in Exam 5, and all of them are one-time measurements. While it is hard to disentangle the temporality between DNAm and HbA1c measurements, our conceptual model supports the statistical model. A low level of education may lead to fewer financial resources, which could produce stress, resulting in changes in DNAm. That could then contribute to the development of metabolic dysregulation.

	Full Sample (n, %)	Low Adult SES %	HbA1c Mean (SD)
Full sample	1231 (100)	67	5.99 (0.92)
Age			
55−65 years	466 (38)	64	5.92 (0.97)
66−75 years	398 (32)	68	6.08 (0.98)
76−85 years	301 (24)	67	5.98 (0.80)
86−95 years	66 (5)	74	5.95 (0.72)
Race			
Non-Hispanic white	582 (47)	51	5.76 (0.65)
African-American	263 (22)	72	6.23 (1.03)
Hispanic	386 (31)	86	6.16 (1.11)
Gender			
Female	633 (51)	73	5.99 (0.88)
Male	598 (49)	60	5.99 (0.97)

Table 8: Descriptive statistics for adult SES measures and HbA1c. n: number of subjects. %: proportion in the corresponding category. SD: standard deviation.

References

[1] Andrew Gelman and Donald B Rubin. "Inference from iterative simulation using multiple sequences". In: *Statistical science* (1992), pp. 457–472.

[2] Toby J Mitchell and John J Beauchamp. "Bayesian variable selection in linear regression". In: *Journal of the American Statistical Association* 83.404 (1988), pp. 1023–1032.

[3] Carlos M Carvalho, Nicholas G Polson, and James G Scott. "The horseshoe estimator for sparse signals". In: *Biometrika* 97.2 (2010), pp. 465–480.

[4] Diane E Bild et al. "Multi-ethnic study of atherosclerosis: objectives and design". In: *American journal of epidemiology* 156.9 (2002), pp. 871–881.

[5] Yongmei Liu et al. "Methylomics of gene expression in human monocytes". In: *Human molecular genetics* 22.24 (2013), pp. 5065–5074.

[6] Pan Du, Warren A Kibbe, and Simon M Lin. "lumi: a pipeline for processing Illumina microarray". In: *Bioinformatics* 24.13 (2008), pp. 1547–1548.

[7] Yi-an Chen et al. “Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray”. In: *Epigenetics* 8.2 (2013), pp. 203–209.

[8] Maarten van Iterson, Erik W van Zwet, and Bastiaan T Heijmans. “Controlling bias and inflation in epigenome-and transcriptome-wide association studies using the empirical null distribution”. In: *Genome biology* 18.1 (2017), p. 19.

[9] Makoto Daimon et al. “Association of the clusterin gene polymorphisms with type 2 diabetes mellitus”. In: *Metabolism* 60.6 (2011), pp. 815–822.

[10] Brian H Neelon, A James O′Malley, and Sharon-Lise T Normand. “A Bayesian model for repeated measures zero-inflated count data with application to outpatient psychiatric service use”. In: *Statistical modelling* 10.4 (2010), pp. 421–439.

[11] Gregory L Burke et al. “The impact of obesity on cardiovascular disease risk factors and subclinical vascular disease: the Multi-Ethnic Study of Atherosclerosis”. In: *Archives of internal medicine* 168.9 (2008), pp. 928–935.

[12] Arleen F Brown et al. “Socioeconomic position and health among persons with diabetes mellitus: a conceptual framework and review of the literature”. In: *Epidemiologic reviews* 26.1 (2004), pp. 63–77.

[13] Belinda L Needham et al. “Life course socioeconomic status and DNA methylation in genes related to stress reactivity and inflammation: the multi-ethnic study of atherosclerosis”. In: *Epigenetics* 10.10 (2015), pp. 958–969.

[14] Kosuke Imai, Luke Keele, Teppei Yamamoto, et al. “Identification, inference and sensitivity analysis for causal mediation effects”. In: *Statistical science* 25.1 (2010), pp. 51–71.