Received: 29 October 2018 | Revised: 30 October 2019 | Accepted: 4 November 2019

DOI: 10.1111/biom.13189

BIOMETRIC METHODOLOGY

Biometrics A JOURNAL OF THE INTERNATIONAL BIOMETRIC SOCIETY WILEY

Bayesian shrinkage estimation of high dimensional causal mediation effects in omics studies

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] | **Bhramar Mukherjee**[1]

[1]Department of Biostatistics, University of Michigan, Ann Arbor, Michigan

[2]Department of Epidemiology, University of Michigan, Ann Arbor, Michigan

[3]Division of Cardiology, Department of Medicine, Duke University School of Medicine, Durham, North Carolina

[4]Department of Epidemiology and Biostatistics, Drexel University, Philadelphia, Pennsylvania

Correspondence
Xiang Zhou and Bhramar Mukherjee, Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109.
Email: xzhousph@umich.edu (X. Z.) and bhramar@umich.edu (B. M.)

Funding information
Division of Mathematical Sciences, Grant/Award Number: 1712933; National Human Genome Research Institute, Grant/Award Number: 009124; National Institute on Minority Health and Health Disparities, Grant/Award Number: 011724; National Heart, Lung, and Blood Institute, Grant/Award Number: 141292

Abstract

Causal mediation analysis aims to examine the role of a mediator or a group of mediators that lie in the pathway between an exposure and an outcome. Recent biomedical studies often involve a large number of potential mediators based on high-throughput technologies. Most of the current analytic methods focus on settings with one or a moderate number of potential mediators. With the expanding growth of -omics data, joint analysis of molecular-level genomics data with epidemiological data through mediation analysis is becoming more common. However, such joint analysis requires methods that can simultaneously accommodate high-dimensional mediators and that are currently lacking. To address this problem, we develop a Bayesian inference method using continuous shrinkage priors to extend previous causal mediation analysis techniques to a high-dimensional setting. Simulations demonstrate that our method improves the power of global mediation analysis compared to simpler alternatives and has decent performance to identify true nonnull contributions to the mediation effects of the pathway. The Bayesian method also helps us to understand the structure of the composite null cases for inactive mediators in the pathway. We applied our method to Multi-Ethnic Study of Atherosclerosis and identified DNA methylation regions that may actively mediate the effect of socioeconomic status on cardiometabolic outcomes.

KEYWORDS

Bayesian sparse models, continuous shrinkage, epigenetics, high-dimensional mediators

1 | INTRODUCTION

Causal mediation analysis has been of significant interest across many disciplines (Ten Have and Joffe, 2012; VanderWeele, 2016). It investigates how an intermediate variable, referred to as mediator, explains the mechanism through which the exposure variable affects the outcome. Under certain regularity conditions, mediation analysis allows us to disentangle the exposure's effect into two parts: effect that acts through the mediator of interest (indirect/mediation effect) and effect that is unexplained by the mediator (direct effect). The state-of-the-art causal mediation analysis (Ten Have and Joffe, 2012), which builds upon the counterfactual framework (Robins and Greenland, 1992), establishes rigorous assumptions regarding the exposure-outcome, exposure-mediator, and mediator-outcome relationships to justify appropriate use of the classical formulas from Baron and Kenny in the linear regression setting (Baron and Kenny, 1986; MacKinnon, 2008) and creates a framework for other general extensions. Many of the existing methods focus on univariate mediation analysis that analyzes one mediator at a time in the causal inference framework and are applicable to both continuous

(Imai *et al.*, 2010) and binary outcomes (VanderWeele and Vansteelandt, 2010). Several studies have recently extended mediation analysis models to jointly account for multiple mediators. However, most of the literature considered settings with two or three mediators, where each mediator is ordered along a priori known mediation pathways and the path-specific effects, are estimated (Daniel *et al.*, 2015). In the presence of multiple unordered mediators, one often has to rely on an *ad hoc* approach to fit a series of univariate mediation models (Taguri *et al.*, 2018; Huang and Pan, 2016) and then summarize the mediation effects across all the mediators. Such approach ignores correlation among mediators, and the estimated mediation effect does not necessarily have a causal interpretation, particularly when the dimension of the potential mediators is truly large.

In this article, using the potential outcome framework for causal inference, we develop a Bayesian mediation analysis method to characterize the indirect effect through an entire set of high-dimensional mediators. Note that Bayesian methods for mediation have also been proposed in a principal stratification framework (Elliott *et al.*, 2010), though there are subsequent discussions on whether the principal stratification framework is a plausible framework to estimate indirect effects (VanderWeele, 2011). In addition, for estimating natural direct and indirect effects, recent work applied Bayesian nonparametric models, especially Dirichlet process mixture models (Kim *et al.*, 2017, 2019) in both univariate and multiple mediators analysis. In contrast, here, we rely on Bayesian variable selection methods to simultaneously analyze a relatively large number of mediators in a pathway with potentially a small number being truly active. With sparsity inducing priors on active coefficients, we assume only a small proportion of mediators in the whole set may mediate the exposure effect on the outcome. This sparsity assumption allows us to extend previous univariate mediation analysis methods to a high-dimensional setting by framing the identification of active mediators in the whole set as a variable selection problem and applying Bayesian methods with continuous shrinkage priors. Unlike previous methods developed for multiple mediators, ours can jointly analyze much larger number of potential mediators without making any path-specific or causal ordering assumptions on mediators. Our method enables us to identify the joint indirect effects of all the mediators and the subset of active ones in the set and propagates uncertainty in inference in a principled way. Recently, there has been emerging interest in high-dimensional mediation analysis, and our method adds to the burgeoning literature for high-dimensional mediators (Chén *et al.*, 2017; Derkach *et al.*, 2019).

While our method is generally applicable to many settings, we examine its performance in the setting of genomics studies. Recent studies have proposed the molecular traits such as gene expression and DNA methylation (DNAm) may act as a mechanism through which various aspects of socioeconomic status (SES) and neighborhood disadvantages affect physical health. For example, childhood/adult SES and neighborhood crime rates have recently been shown to influence DNAm in several genes related to stress and inflammation (Needham *et al.*, 2015; Smith *et al.*, 2017). DNAm of inflammatory markers has also been associated with cardiovascular risk and disease (Zhong *et al.*, 2016). Here, we show through simulations and data analysis that our high-dimensional mediation analysis framework can increase power of a joint analysis and facilitate the identification of active mediators in the set.

2 | NOTATION, DEFINITIONS, AND ASSUMPTIONS

In this article, we focus on causal mediation analysis for the setting where there is a single exposure of interest but there exists a high-dimensional set of candidate mediators that may mediate the effect of exposure on an outcome. Suppose our analysis is based on a study of n subjects and for subject i, $i = 1, \dots, n$, we collect data on exposure A_i, p candidate mediators $\boldsymbol{M_i} = (M_i^{(1)}, M_i^{(2)}, \dots, M_i^{(p)})^T$, outcome Y_i, and q covariates $\boldsymbol{C_i} = (C_i^{(1)}, \dots, C_i^{(q)})^T$. In particular, we focus on the case where Y_i and $\boldsymbol{M_i}$ are all continuous variables.

We adopt the counterfactual (or potential outcomes) framework to formally define mediators and their causal effects. Let $M_i^{(j)}(a)$ denote the potential (or counterfactual) value of the jth mediator, $j = 1, \dots, p$, for subject i under exposure level at a. Suppose the exposure has K levels, then $K \times p$ potential counterfactual random variables for mediators are defined, i.e. $M^{(1)}(1), M^{(2)}(1), \dots, M^{(p)}(1), M^{(1)}(2), M^{(2)}(2), \dots, M^{(p)}(2), M^{(1)}(K), M^{(2)}(K), \dots, M^{(p)}(K)$. Let $Y_i(a, \boldsymbol{m}) = Y_i(a, m^{(1)}, \dots, m^{(p)})$ denote the ith subject's potential outcome if the subject's exposure were a and mediators were $\boldsymbol{m} = (m^{(1)}, \dots, m^{(p)})$. As this paper focuses on the joint effects of the whole set of mediators, for simplicity, we define $\boldsymbol{M_i}(a) = (M_i^{(1)}(a), M_i^{(2)}(a), \dots, M_i^{(p)}(a))$. These counterfactuals are hypothetical variables and may not be observed in real data. To connect potential variables to observed data, we make the Stable Unit Treatment Value Assumption (SUTVA) (Rubin, 1980), which is a commonly made assumption in causal inference. Specifically, the SUTVA assumes there is no interference between subjects and the consistency assumption, which states that the observed variables are the same as the potential variables corresponding to the actually observed treatment level, that is, $\boldsymbol{M_i} = \sum_a \boldsymbol{M_i}(a) I(A_i = a)$, and $Y_i = \sum_a \sum_{\boldsymbol{m}} Y_i(a, \boldsymbol{m}) I(A_i = a, \boldsymbol{M_i} = \boldsymbol{m})$, where $I(\cdot)$ is the indicator function. For simplicity, we define $Y_i(a) = Y_i(a, \boldsymbol{M_i}(a))$, the potential outcome had the exposure been a and the whole set of mediators been the value that would have been observed under exposure a. The defined

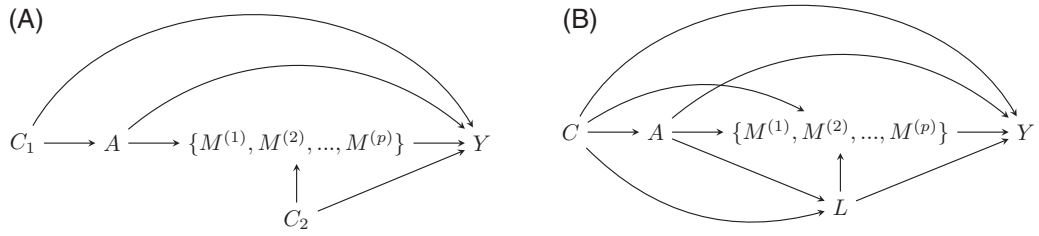

FIGURE 1 Left (A): High-dimensional mediators ($(M^{(1)}, M^{(2)}, \dots, M^{(p)})$) between exposure ($A$) and outcome ($Y$) with exposure-outcome confounders C_1 and mediator-outcome confounders C_2; right (B): An example of mediator-outcome confounder L that is affected by the exposure A

potential variables are hypothetical, and actually most of them are not observed in real data. For example, if $A_i \neq a$, then $Y_i(a)$ or $\boldsymbol{M}_i(a)$ are not observed.

We may decompose the effect of an exposure into its direct effect and effect mediated through the whole set of mediators (VanderWeele and Vansteelandt, 2014). The controlled direct effect (CDE) of the exposure on the outcome is defined as $Y_i(a, \boldsymbol{m}) - Y_i(a^\star, \boldsymbol{m})$, which is the effect of changing exposure from level $a^\star$ (the reference level) to a while hypothetically controlling mediators at level $\boldsymbol{m}$. The natural direct effect (NDE) is defined as $Y_i(a, \boldsymbol{M_i}(a^\star)) - Y_i(a^\star, \boldsymbol{M_i}(a^\star))$, which is the CDE when mediators are controlled at the level that would have naturally been had the exposure been $a^\star$. The natural indirect effect (NIE) is defined by $Y_i(a, \boldsymbol{M_i}(a)) - Y_i(a, \boldsymbol{M_i}(a^\star))$, capturing the effect mediated through the whole set of mediators, that is, the change in potential outcomes when mediators change from $\boldsymbol{M_i}(a^\star)$ to $\boldsymbol{M_i}(a)$ while fixing exposure at a. The total effect (TE), $Y_i(a) - Y_i(a^\star)$, can then be decomposed into natural direct and indirect effect, written as $Y_i(a) - Y_i(a^\star) = Y_i(a, \boldsymbol{M_i}(a)) - Y_i(a^\star, \boldsymbol{M_i}(a^\star)) = Y_i(a, \boldsymbol{M_i}(a)) - Y_i(a, \boldsymbol{M_i}(a^\star)) + Y_i(a, \boldsymbol{M_i}(a^\star)) - Y_i(a^\star, \boldsymbol{M_i}(a^\star))$ $= \text{NIE} + \text{NDE}$.

Causal effects are formally defined in terms of potential variables, which are not necessarily observed, but the identification of causal effects must be based on observed data. Therefore, similar to missing data problems, further assumptions regarding the confounders are required for the identification of causal effects in mediation analysis (VanderWeele and Vansteelandt, 2014). We will use $A \perp\!\!\!\perp B|C$ to denote that A is independent of B conditional on C. For estimating the average CDE, two assumptions on confounding are needed: (1) $Y_i(a, \boldsymbol{m}) \perp\!\!\!\perp A_i|\boldsymbol{C_i}$, namely, there is no unmeasured confounding for the exposure effect on the outcome; (2) $Y_i(a, \boldsymbol{m}) \perp\!\!\!\perp \boldsymbol{M_i}|\{\boldsymbol{C_i}, A_i\}$, namely, there is no unmeasured confounding for any of mediator-outcome relationship after controlling for the exposure. The two assumptions are illustrated in the left panel of Figure 1, and controlling for exposure-outcome and mediator-outcome confounding corresponds to controlling for C_1, C_2 in the figure. In practice, both sets of covariates C_1 and C_2 need not to be distinguished from one another and can simply be included in the overall set of C that we adjust for. The identification of the average NDE and NIE requires assumption (1) and (2), along with two additional assumptions: (3) $\boldsymbol{M_i}(a) \perp\!\!\!\perp A_i|\boldsymbol{C_i}$, namely, there is no unmeasured confounding for the exposure effect on all the mediators; (4) $Y_i(a, \boldsymbol{m}) \perp\!\!\!\perp \boldsymbol{M_i}(a^\star)|\boldsymbol{C_i}$, which can be interpreted as there is no downstream effect of the exposure that confounds the mediator-outcome relationship for any of the mediators. Graphically, assumption (4) implies that there should be no arrow going from exposure A to mediator-outcome confounder C_2 in Figure 1A. It is thus violated in Figure 1B since the mediator-outcome confounder L is itself affected by the exposure. The four assumptions are required to hold with respect to the whole set of mediators $\boldsymbol{M_i}(a)$. Finally, as in all mediation analysis, the temporal ordering assumption also needs to be satisfied, that is, the exposure precedes the mediators, which precede the outcome. With the above assumptions, the average NDE and NIE can 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. The full derivation can be found in the Supporting Information.

We note that as the main interest of this article lies in the joint effect of the whole set of mediators,thus the definition of NIE and NDE only involve the counterfactuals of the form $\boldsymbol{M_i}(a) = (M_i^{(1)}(a), M_i^{(2)}(a), \dots, M_i^{(p)}(a))$. If one is interested in estimating the effect of a specific mediator, then one needs to consider the K^p counterfactuals $(M_i^{(1)}(a_1), M_i^{(2)}(a_2), \dots, M_i^{(p)}(a_p))$, $a_1, a_2, \dots, a_p \in \{1, 2, \dots, K\}$. Characterizing mediator-specific NIE is a much more challenging task and requires stronger assumptions, in particular when the multiple mediators influence and interact with one another.

3 | MODELS AND ESTIMANDS

Since the effects of mediators (average NDE and NIE) defined in terms of potential outcomes can be deduced from two conditional models for $Y_i|A_i, \boldsymbol{M_i}, \boldsymbol{C_i}$ and $\boldsymbol{M_i}|A_i, \boldsymbol{C_i}$ using observed data, we propose two regression models and subsequently deduce the causal effects of mediators. For modeling $Y_i|A_i, \boldsymbol{M_i}, \boldsymbol{C_i}$, we assume for subject i ($i = 1, \dots, n$), a continuous outcome of interest Y_i is associated with exposure A_i,

p potential mediators $\boldsymbol{M_i} = (M_i^{(1)}, M_i^{(2)}, \ldots, M_i^{(p)})^T$ that may be on the pathway from A_i to Y_i, and q covariates $\boldsymbol{C_i}$ with the first element being the scalar 1 for the intercept:

$$Y_i = \boldsymbol{M_i}^T \boldsymbol{\beta_m} + A_i \beta_a + \boldsymbol{C_i}^T \boldsymbol{\beta_c} + \epsilon_{Yi}, \tag{1}$$

where $\boldsymbol{\beta_m} = ((\boldsymbol{\beta_m})_1, \ldots, (\boldsymbol{\beta_m})_p)^T$, $\boldsymbol{\beta_c} = (\beta_{c1}, \ldots, \beta_{cq})^T$, $\epsilon_{Yi} \sim N(0, \sigma_e^2)$. Here we assume there is no interaction between A_i and $\boldsymbol{M_i}$. Next for modeling $\boldsymbol{M_i}|A_i, \boldsymbol{C_i}$, we consider a multivariate regression model that jointly analyzes the p potential mediators:

$$\boldsymbol{M_i} = A_i \boldsymbol{\alpha_a} + \boldsymbol{\alpha_c} \boldsymbol{C_i} + \boldsymbol{\epsilon}_{Mi}, \tag{2}$$

where $\boldsymbol{\alpha_a} = ((\boldsymbol{\alpha_a})_1, \ldots, (\boldsymbol{\alpha_a})_p)^T$, $\boldsymbol{\alpha_c} = (\boldsymbol{\alpha_{c1}}^T, \ldots, \boldsymbol{\alpha_{cp}}^T)^T$, $\boldsymbol{\alpha_{c1}}, \ldots, \boldsymbol{\alpha_{cp}}$ are q-by-1 vectors, $\boldsymbol{\epsilon}_{Mi} \sim MVN(\boldsymbol{0}, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma}$ captures the correlation among the mediators. ϵ_{Yi} and $\boldsymbol{\epsilon}_{Mi}$ are assumed independent of A_i, $\boldsymbol{C_i}$ and each other.

With assumptions made in Section 2, we show in the Supporting Information that the average NDE, NIE, and TE can then be computed as presented below, and in the rest of the article, we refer to NDE as a direct effect and NIE as an indirect/mediation effect.

$$\text{NDE} = E[Y_i(a, \boldsymbol{M_i}(a^\star)) - Y_i(a^\star, \boldsymbol{M_i}(a^\star))|\boldsymbol{C_i}] = \beta_a(a - a^\star). \tag{3}$$

$$\begin{aligned}\text{NIE} &= E[Y_i(a, \boldsymbol{M_i}(a)) - Y_i(a, \boldsymbol{M_i}(a^\star))|\boldsymbol{C_i}] \\ &= (a - a^\star) \sum_{j=1}^{p} (\boldsymbol{\alpha_a})_j (\boldsymbol{\beta_m})_j. \end{aligned} \tag{4}$$

$$\text{TE} = E[Y_i(a) - Y_i(a^\star)|\boldsymbol{C_i}] = (\beta_a + \boldsymbol{\alpha_a}^T \boldsymbol{\beta_m})(a - a^\star). \tag{5}$$

As noted in Equation (4), under the assumptions of model (1) the NIE through the whole set of mediators turns out to be the sum of the product of $(\boldsymbol{\alpha_a})_j$ and $(\boldsymbol{\beta_m})_j$ over the entire set. Those individual product terms do not correspond to the NIE of a specific (say jth) mediator. We define active mediators as the ones with nonnull contribution to the global NIE, that is $(\boldsymbol{\alpha_a})_j (\boldsymbol{\beta_m})_j$ being nonzero. The proposed Bayesian shrinkage and selection methods are used to identify and estimate these active components. Any inactive mediator will naturally fall into one of the following three categories: $(\boldsymbol{\beta_m})_j$ is nonzero, while $(\boldsymbol{\alpha_a})_j$ is zero; $(\boldsymbol{\alpha_a})_j$ is nonzero, while $(\boldsymbol{\beta_m})_j$ is zero; both are zero. Such a refined partition for the high-dimensional set of mediators provides useful and insightful interpretations for the structure of the composite null. Regarding a summary global measure of the indirect effect, we note that the quantity of $\sum_{j=1}^{p} (\boldsymbol{\alpha_a})_j (\boldsymbol{\beta_m})_j$ is not a good summary of the global mediation effects when the terms have opposite directions. Considering this, we propose to use the L_2 norm of the p-by-1 vector of $(\boldsymbol{\alpha_a})_j (\boldsymbol{\beta_m})_j$ (Huang and Pan, 2016) as a global measure of mediation effects, that is, $\tau = ||((\boldsymbol{\alpha_a})_1 (\boldsymbol{\beta_m})_1, (\boldsymbol{\alpha_a})_1 (\boldsymbol{\beta_m})_2, \ldots, (\boldsymbol{\alpha_a})_p (\boldsymbol{\beta_m})_p)||^2 = \sum_{j=1}^{p} \{(\boldsymbol{\alpha_a})_j (\boldsymbol{\beta_m})_j\}^2$.

4 | BAYESIAN METHOD FOR ESTIMATION

4.1 | Prior specification

In order to conduct high-dimensional mediation analysis, we need to make certain model assumptions on the effect sizes. In genomewide association studies, Bayesian sparse regression models, such as Bayesian variable selection regression models (BVSR), have been proven to yield better power in detecting relevant covariates (Guan and Stephens, 2011). Here, we also make the reasonable sparsity assumption, which implies that only a small proportion of mediators mediate the exposure effects on the outcome. Linear mixed models (LMM), on the other hand, assume that every mediator transmits certain effects from exposure to outcome, with the effect sizes normally distributed. We first assume that all the potential mediators contribute small, nonzero effects in mediating the exposure-outcome relationship, which is aligned with the main idea of polygenic (Zhou *et al.*, 2013) and omnigenic (Boyle *et al.*, 2017) models. Besides these small effects, we also assume that there is a small proportion of mediators exhibiting additional/large effects. We refer to these mediators with additional effects as active mediators, which is consistent with the concept of core genes defined in the omnigenic model. Therefore, in this article, we use the Baysian sparse linear mixed model (BSLMM) priori, which imposes continuous shrinkage on the effects (Zhou *et al.*, 2013) and assumes the presence of small and additional effects, for high-dimensional mediation analysis. The BSLMM is capable of learning the underlying mediation architecture from the data, producing good performances across a wide range of scenarios. Our model assumptions are also akin to the notion of quasi-sparsity that has become popular with continuous shrinkage priors (Ge *et al.*, 2019). Specifically, we assume a mixture of two normal components a priori for the jth mediator, $j = 1, 2, \ldots p$,

$$\begin{aligned}(\boldsymbol{\beta_m})_j &\sim \pi_m N(0, \sigma_{m1}^2) + (1 - \pi_m) N(0, \sigma_{m0}^2) \\ (\boldsymbol{\alpha_a})_j &\sim \pi_a N(0, \sigma_{ma1}^2) + (1 - \pi_a) N(0, \sigma_{ma0}^2),\end{aligned}$$

where $\sigma_{m1}^2 > \sigma_{m0}^2$, $\sigma_{ma1}^2 > \sigma_{ma0}^2$, and π_m, π_a denote the proportion of coefficients that belong to the normal distribution with a larger variance.

For the other coefficients, we assume $\beta_a \sim N(0, \sigma_a^2)$ and $\boldsymbol{\beta_c}, \boldsymbol{\alpha_c} \sim MVN(\boldsymbol{0}, \sigma_c^2 \boldsymbol{I}), \sigma_c^2 \to \infty$. Here we use a limiting normal prior for $\boldsymbol{\beta_c}, \boldsymbol{\alpha_c}$ with its variance going to infinity, since we often have insufficient information from the data to overwhelm any prior assumptions. For the convenience of modeling, we set the correlation structure among mediators $\boldsymbol{\Sigma}$ as $\sigma_g^2 \boldsymbol{I}$.

TABLE 1 Mediators are categorized into four groups based on their relationships with exposure and outcome

$(\boldsymbol{\beta_m})_j$	Larger component	Smaller component
$(\boldsymbol{\alpha_a})_j$		
Larger component	$r_{mj} * r_{aj} = 1$ (Group 1)	$r_{mj} = 0, r_{aj} = 1$ (Group 2)
Smaller component	$r_{mj} = 1, r_{aj} = 0$ (Group 3)	$r_{mj} = r_{aj} = 0$ (Group 4)

Group 1: Both $(\boldsymbol{\beta_m})_j$ and $(\boldsymbol{\alpha_a})_j$ come from larger normal components; Group 2: $(\boldsymbol{\alpha_a})_j$ from larger normal component while $(\boldsymbol{\beta_m})_j$ from smaller normal component; Group 3: $(\boldsymbol{\beta_m})_j$ from larger normal component while $(\boldsymbol{\alpha_a})_j$ from smaller normal component; Group 4: both $(\boldsymbol{\beta_m})_j$ and $(\boldsymbol{\alpha_a})_j$ come from smaller normal components.

For the hyperparameters of variances in the model, we use the standard conjugate priors,

$$\begin{aligned}
\sigma^2_{\text{ms}} &\sim \text{inverse-gamma}(k_{\text{ms}}, l_{\text{ms}}), s = 0, 1 \\
\sigma^2_{\text{mas}} &\sim \text{inverse-gamma}(k_{\text{mas}}, l_{\text{mas}}), s = 0, 1 \\
\sigma^2_a &\sim \text{inverse-gamma}(k_a, l_a) \text{ and} \\
\sigma^2_e, \sigma^2_g &\sim \text{inverse-gamma}(k_e, l_e).
\end{aligned}$$

We set $k_{m0} = k_{m1} = k_a = k_{ma0} = k_{ma1} = k_e = 2.0$, and $l_{m0} = l_{ma0} = 10^{-4}, l_a = l_{m1} = l_{ma1} = l_e = 1.0$. The prior inclusion probabilities π_m, π_a encode the prior information about the sparsity of the coefficients. We place a uniform prior on $\log(\pi_m), \log(\pi_a)$,

$$\log(\pi_m), \log(\pi_a) \sim U(\log(1/p), \log(1)).$$

The priors were chosen so that π_m and π_a range from $1/p$ to 1, and the lower and upper bounds correspond to an expectation of 1 and p covariates in each model. A uniform prior on $\log(\pi_m)$ and $\log(\pi_a)$ reflects the fact that the uncertainty in π_m, π_a is large due to the sparsity of the models. We do not choose a uniform prior on π_m, π_a since that would put appreciable prior probability on large numbers of covariates (Guan and Stephens, 2011).

4.2 | Posterior sampling algorithm

We develop a Markov chain Monte Carlo (MCMC) sampling algorithm to obtain the posterior samples from our Bayesian method. To facilitate MCMC, we introduce indicator variables $\boldsymbol{r_m}, \boldsymbol{r_a} \in \{0, 1\}^p$ to indicate which normal component $(\boldsymbol{\beta_m})_j$ and $(\boldsymbol{\alpha_a})_j$ are from, and for the jth mediator, $r_{mj} = \text{I}((\boldsymbol{\beta_m})_j \sim N(0, \sigma^2_{m1}))$, $r_{aj} = \text{I}((\boldsymbol{\alpha_a})_j \sim N(0, \sigma^2_{ma1}))$, where $\text{I}(\cdot)$ represents an indicator function. We use a Hastings-within-Gibbs algorithm to obtain posterior samples, and full details of the algorithm appear in the Supporting Information.

For the jth mediator, we can estimate the posterior probability of both $(\boldsymbol{\beta_m})_j$ and $(\boldsymbol{\alpha_a})_j$ being in the normal components with larger variances as the posterior inclusion probability (PIP), defined as $P(r_{mj} = 1, r_{aj} = 1|\text{Data})$ in our model. Mediators with larger $(\boldsymbol{\beta_m})_j$ and $(\boldsymbol{\alpha_a})_j$ tend to be categorized into larger variance normals, and such tendency can be quantified by the mediator's PIP. PIP provides nonnull evidence for both $(\boldsymbol{\beta_m})_j$ and $(\boldsymbol{\alpha_a})_j$, and, therefore, we select mediators with the highest PIP as potentially active mediators.

4.3 | Mediator categorization

Under the above Bayesian mediation framework, active mediators are the ones whose $(\boldsymbol{\beta_m})_j$ and $(\boldsymbol{\alpha_a})_j$ both come from larger normal components. The three categories for the inactive mediators are $(\boldsymbol{\beta_m})_j$ from larger normal component while $(\boldsymbol{\alpha_a})_j$ from smaller normal component; $(\boldsymbol{\alpha_a})_j$ from larger normal component while $(\boldsymbol{\beta_m})_j$ from smaller normal component; both from smaller components. In practice, we have the indicator variables r_{mj} and r_{aj} to denote which component the coefficients $(\boldsymbol{\beta_m})_j$, $(\boldsymbol{\alpha_a})_j$ belong to and can easily obtain the posterior probabilities for each group. The four groups are illustrated in Table 1,

5 | SIMULATIONS

We evaluate the performance of the proposed Bayesian mediation method and compare it with the three existing mediation methods, which include single mediation analysis, multivariate mediation analysis, and high-dimensional multivariate mediation (HDMM) methodology of Chén *et al.* (2017). Single mediation analysis tests one mediator at a time for its mediation effect, and we use the R package **mediation** with the nonparametric bootstrap option for standard error estimation. Multivariate mediation analysis (VanderWeele and Vansteelandt, 2014), on the other hand, jointly analyzes all the mediators in both model (1) and (2) and tests the product term $(\boldsymbol{\beta_m})_j(\boldsymbol{\alpha_a})_j$ for each j at a time while controlling for all other variables. This method can only be fit when a multivariate ordinary least squares regression model can be fit for the outcome model (1). The HDMM is a novel method recently developed for high-dimensional mediation analysis and aims to identify active mediators through dimension reduction techniques. We use p-values for univariate and multivariate mediation analysis, estimated indirect effect for HDMM and PIP for our Bayesian method as measures of the evidence for mediation. We compare the power to identify active mediators based on either 5% or 10% false discovery rate (FDR).

We consider various simulation settings with $n = 1000$ and p mediators ($p = 100/2000$). We first examine the settings of $p = 100$ (ie, $p < n$) in order to include the multivariate mediation analysis for comparison. In each setting, we simulate the continuous exposure variables $\{A_i, i = 1, \ldots, 1000\}$ independently from a standard normal distribution. We then generate a p-vector of mediators for the ith individual from $\boldsymbol{M_i} = A_i\boldsymbol{\alpha_a} + \boldsymbol{\epsilon}_{Mi}$. Each element of $\boldsymbol{\alpha_a}$, $(\boldsymbol{\alpha_a})_j$ $(j = 1, \ldots, p)$ is simulated from a point-normal prior: $\pi_a N(0, 1) + (1 - \pi_a)\delta_0$, where δ_0 is a point mass at zero. The residual errors $\boldsymbol{\epsilon}_{Mi}$ are simulated from a multivariate normal distribution with mean zero and a covariance $\boldsymbol{\Sigma}$. $\boldsymbol{\Sigma}$ accounts for the correlation among mediators commonly seen in real data, and we use the sample covariance estimated from the Multi-Ethnic Study of Atherosclerosis (MESA) data to serve as $\boldsymbol{\Sigma}$. Since our Bayesian mediation model does not explicitly account for the correlation structure of mediators in the exposure-mediators model, the simulations with correlated mediators allow us to examine the robustness of our modeling assumption on independence. We scale the two terms $A_i\boldsymbol{\alpha_a}$ and $\boldsymbol{\epsilon}_{Mi}$ further so that the former explains a fixed proportion of variance: $PVE_A = \text{Var}(A_i\boldsymbol{\alpha_a})/\text{Var}(\boldsymbol{M_i})$, where Var denotes the sample variance.

Given the exposure and mediators, we then generate the outcome Y_i from the linear model: $Y_i = \boldsymbol{M_i}^T\boldsymbol{\beta_m} + A_i\beta_a + \epsilon_{Yi}$. Here, each element of $\boldsymbol{\beta_m}$, $(\boldsymbol{\beta_m})_j$ $(j = 1, \ldots, p)$, is simulated from $\pi_m N(0, 1) + (1 - \pi_m)\delta_0$. The residual error ϵ_{Yi} is simulated independently from $N(0, 1)$. We assume that only 10% of the mediators are truly active ones, whose $(\boldsymbol{\beta_m})_j$ and $(\boldsymbol{\alpha_a})_j$ are both sampled from the large variance normal distribution. After simulating $\boldsymbol{M_i}^T\boldsymbol{\beta_m}$, $A_i\beta_a$, and ϵ_{Yi}, we scale these three terms further to achieve two desirable PVEs: $PVE_{IE} = \text{Var}(\boldsymbol{\alpha_a}^T\boldsymbol{\beta_m}A_i)/\text{Var}(Y_i)$ and $PVE_{DE} = \text{Var}(A_i\beta_a)/\text{Var}(Y_i)$.

We consider a baseline scenario where we set $PVE_A = 0.5, PVE_{IE} = 0.4, PVE_{DE} = 0.1, \pi_a = 0.3, \pi_m = 0.2$. We then vary each of the four parameters (PVE_A, PVE_{IE}, π_a, π_m) one at a time to investigate their individual influences on the results. We perform 200 replicates for each scenario to do the power comparison.

For $p = 100$, we display the comparative results in Figure 2. The results show that our Bayesian multivariate mediation method outperforms the other three methods in all scenarios. For example, in the baseline setting, at 10% FDR, the Bayesian mediation method achieves a power of 0.725, while the univariate and multivariate methods and HDMM achieve a power of 0.527, 0.676, and 0.167, respectively. The power of the four approaches increases with increasing PVE_{IE}, which increases the effect sizes of $\boldsymbol{\beta_m}$. In addition, the power of most approaches reduces with increased π_a or π_m, which reduces the effect sizes of either $\boldsymbol{\alpha_a}$ or $\boldsymbol{\beta_m}$, respectively. As expected, the advantage of our Bayesian method over the univariate and multivariate methods is more apparent in sparse settings with smaller values of π_a and π_m. In terms of PVE_A, which determines the effect size of $\boldsymbol{\alpha_a}$, we found that the power of different methods first increases slightly when PVE_A changes from 0.3 to 0.5 and then decreases slightly as PVE_A changes further to 0.8. The later decrease in power in the setting of $PVE_A = 0.8$ is presumably due to the increased correlation between the exposure and mediators, which makes it difficult for all the methods to distinguish between direct and indirect effects in model (1). The performance of HDMM is relatively stable to PVE_A, PVE_{IE} and π_a and improves slightly with increased π_a. HDMM does not assume sparsity on mediation effects and thus does not fare well in relatively sparse situations. Between single and multivariate mediation methods, the latter yields better power than the former in all scenarios, as the multivariate one properly controls for the correlation among mediators.

Next, we examine the settings for $p = 2000$. Now we select 1% of the mediators to be active and set $\pi_m = 2\%$, $\pi_a = 3\%$ as the baseline setting with all other configurations being same as in the baseline setting of $p = 100$. We use a threshold of 1% false positive rate (FPR) instead of false discovery rate due to low power in the $p = 2000$ settings. The comparisons are shown in Figure 3. The Bayesian mediation method yields more power than the single mediator analysis and HDMM in all the scenarios. For example, in the baseline setting, at 1% FPR, the Bayesian mediation method achieves a power of 0.470, while the univariate method and HDMM have a power of 0.357 and 0.248, respectively. The power of our method and the univariate approach again increase with increasing PVE_{IE} and decrease with increasing π_a/π_m. Increasing PVE_A decreases the power of the Bayesian method, while tends to improve the performance of HDMM possibly due to the dimension reduction applied on the high-dimensional mediators. Comparing the settings with varied π_a, we note that the major power gain of our method lies in the joint analysis of mediators in the outcome model and appropriate shrinkage on $\boldsymbol{\beta_m}$. For the mediator model, we are essentially fitting a series of regression models for each mediator and exposure. Therefore, shrinkage on the vector of $\boldsymbol{\alpha_a}$ does not help much in mediator selection, especially if π_a is relatively large, for example, 0.1 or 0.25. In situations where the number of nonzero $\boldsymbol{\beta_m}$ gets closest to the reduced dimension of $\boldsymbol{M}$ in HDMM, its power is improved (eg, $\pi_m = 0.1$).

Finally, we examine the ability of our method to estimate the global NIE and the proportions of mediators in the four categories as shown in Table 1. The full results are provided in the Supporting Information. Overall, 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 the model does not have full power to detect all the active mediators. The 95% credible intervals of τ shows that its posterior distribution is asymmetric and depends on the composition of the four groups. We also show a distribution graph from the

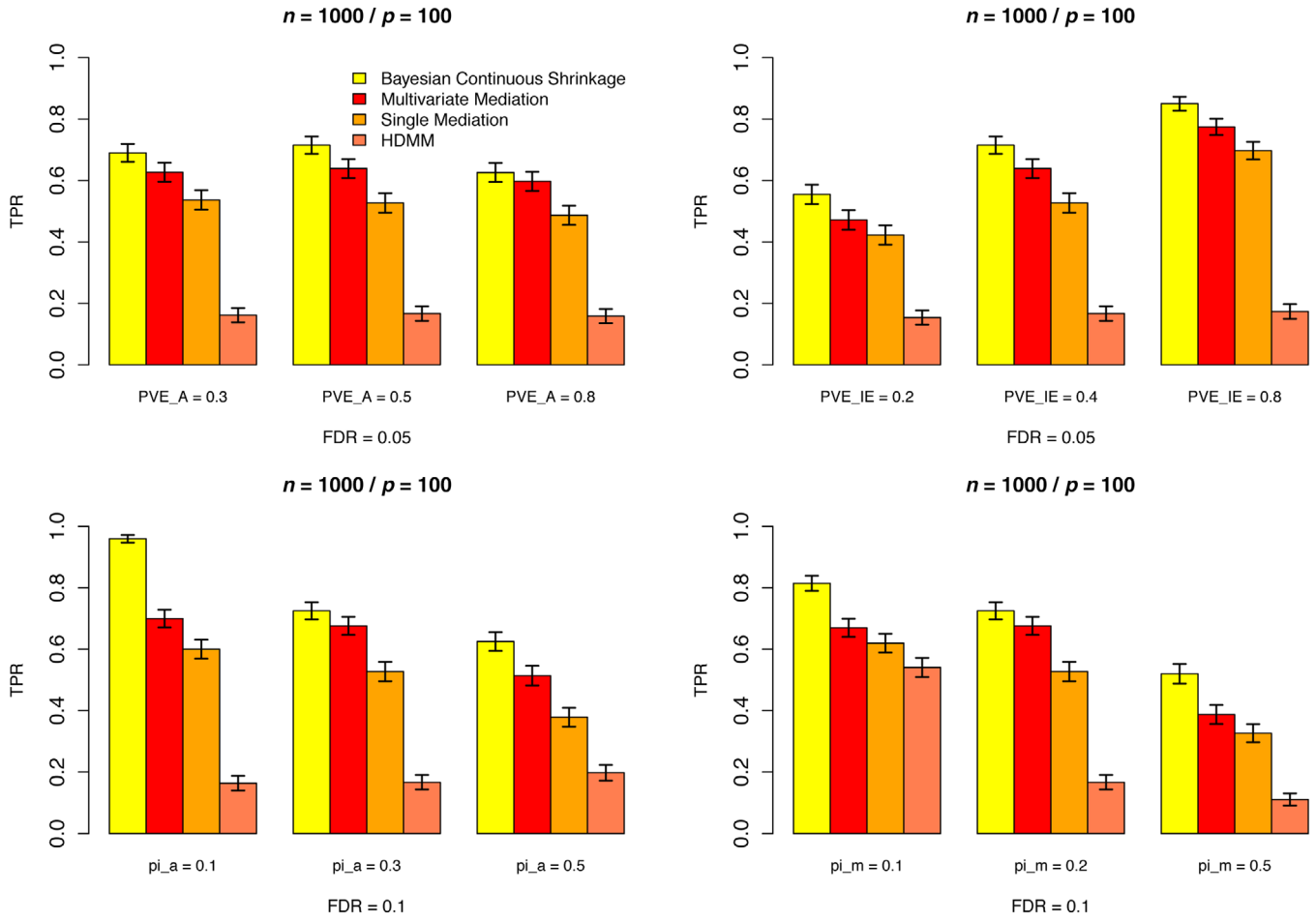

FIGURE 2 Power comparison among our Bayesian mediation method (yellow), multivariate mediation method (red), single mediation method (orange), and HDMM (coral) when the number of mediators is 100 and sample size 1000. The *x*-axis marks the one parameter we change at a time from the baseline setting. We calculate the true positive rate (TPR) for the power comparison. The average TPR at FDR = 0.05/0.1 and its error bar based on ±2 standard errors are calculated across 200 replicates. The standard error of the empirically estimated proportions is computed using variance of a binomial random variable

posterior samples of τ in four different scenarios with $n = 1000, p = 100$ in Figure 2 of Supporting Information.

6 | DATA ANALYSIS

We applied the proposed Bayesian method to investigate the mediation mechanism of DNAm in the pathway from adult SES to glycated hemoglobin (HbA1c) in the MESA (Bild *et al.*, 2002). The exposure, adult SES, is indicated by adult educational attainment and is an important risk factor for cardiovascular diseases. The outcome, HbA1c, is a surrogate measurement of average blood glucose levels and a critical variable for various diseases including type 2 diabetes (T2D) and cardiovascular disease (CVD) (Selvin *et al.*, 2010). We provide our detailed processing steps for MESA data in the Supporting Information. For computational reasons, we focused on a final set of 2000 cytosine-phosphate-guanine (CpG) sites that have the strongest marginal associations with adult SES for the following mediation analysis with 1231 individuals.

We applied both univariate mediation analysis and our Bayesian multivariate mediation analysis to analyze the selected 2000 CpG sites. For the multivariate analysis, we consider

$$Y_i = \boldsymbol{M_i}^T \boldsymbol{\beta_m} + A_i \beta_a + \boldsymbol{C_{2i}}^T \boldsymbol{\beta_c} + \epsilon_{Yi} \quad (6)$$

$$\boldsymbol{M_i} = A_i \boldsymbol{\alpha_a} + \boldsymbol{\alpha_c} \boldsymbol{C_{1i}} + \boldsymbol{\epsilon}_{Mi}, \quad (7)$$

where Y_i represents HbA1c levels, A_i represents adult SES values, and $\boldsymbol{M_i}$ represents the methylation level for 2000 CpG sites. In Equation (6), the model controls for age, gender, and race/ethnicity, and in Equation (7), we adjust for age, gender, race/ethnicity, and enrichment scores for four major blood cell types (neutrophils, B cells, T cells, and natural killer cells). All the continuous variables are standardized to have zero mean and unit variance. The univariate analysis is applied in a similar fashion except that it is used to analyze one site at a time.

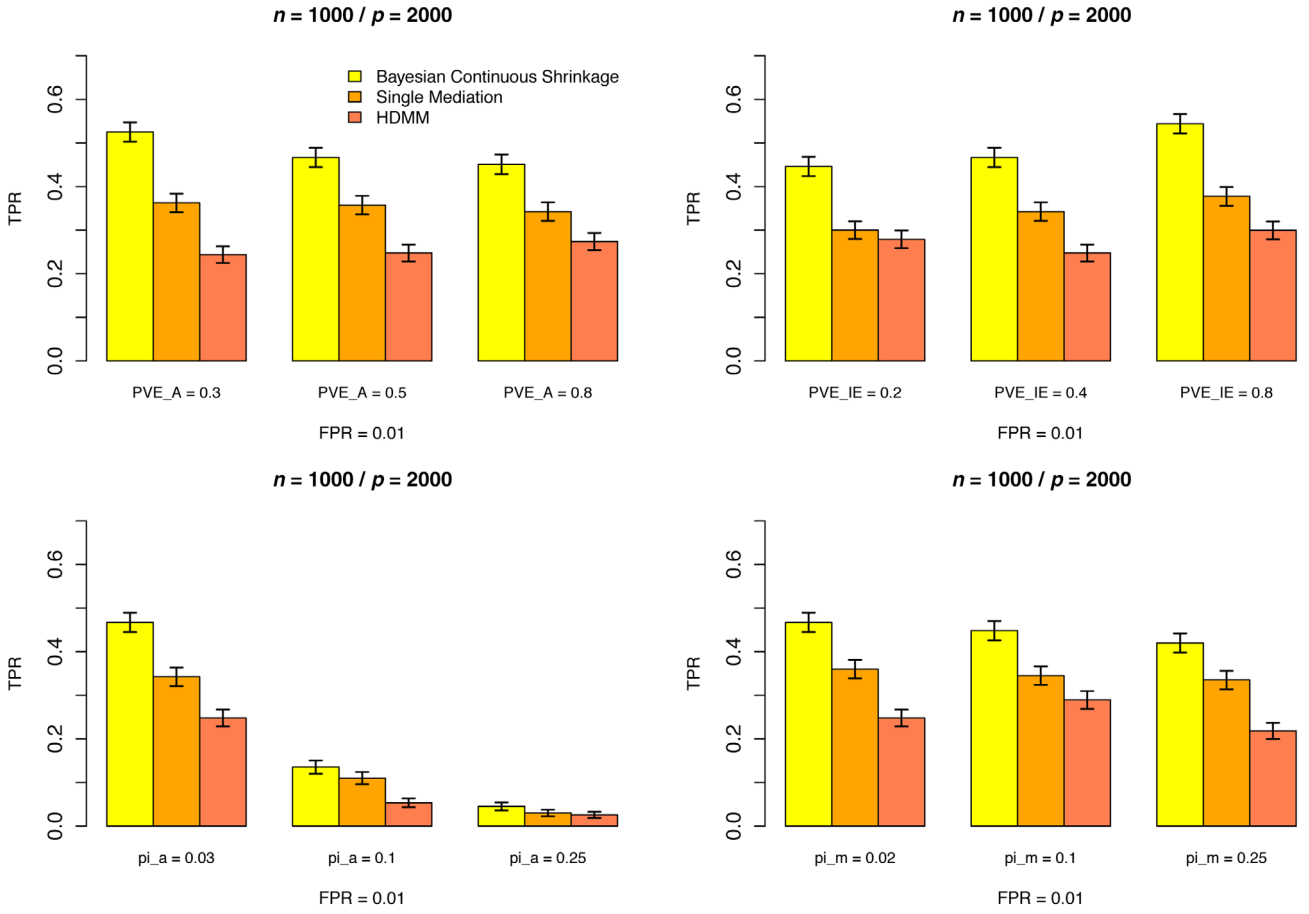

FIGURE 3 Power comparison among our Bayesian mediation method (yellow), single mediation method (orange), and HDMM (coral) when the number of mediators is 2000 and sample size 1000. The x-axis marks the one parameter we change at a time from the baseline setting. We calculate the TPR for the power comparison. The average TPR at FPR = 0.01 and its error bar based on ± 2 standard errors are calculated across 200 replicates. The standard error of the empirically estimated proportions is computed using variance of a binomial random variable

We display PIP values for each of the 2000 CpG sites from the Bayesian multivariate analysis in Figure 4. Two CpG sites were identified with strong evidence (PIP > 0.5) for mediating the adult SES effects on HbA1c. They are also among the top 10 sites with the smallest p-values obtained from univariate mediation analysis. In addition, these two CpG sites are close to genes *CCDC54* and *CCND2*, both of which are known candidates associated with HbA1c. Specifically, the expression of *CCND2* has been shown to be associated with risk of T2D and the related glycemic traits of glucose, HbA1c, and insulin (Yaghootkar *et al.*, 2015). The gene *CCDC54* interacts with valproic acid and acrylamide, both of which are associated with diabetes and blood insulin (Lin *et al.*, 2009). Therefore, strong evidence from our method suggests that adult SES may act through these two genes to affect HbA1c. We also apply the HDMM and Bayesian methods with spike-and-slab and horseshoe priors to the data. For HDMM, the weights for the first direction of mediation do not suggest obvious signal or pattern. We also note that there is a lack of biological evidence to support a mediating role of the genes picked out by the other methods, except for one gene (*CLU*).

In addition, we estimate the global mediation effects $\hat{\tau}$ as 0.0084 and its 95% credible interval from the posterior as (0.0063, 0.0115). The $\widehat{PVE}_{IE}$ is 0.096, indicating that approximately 10% of the outcome variance is indirectly explained by DNAm jointly after controlling for covariates. We also estimate the proportion of CpG sites in each of the four categories as defined in Section 4.3: $\hat{\pi}_{g1} = 0.002$, $\hat{\pi}_{g2} = 0.031$, $\hat{\pi}_{g3} = 0.001$, $\hat{\pi}_{g4} = 0.966$. We find that a small proportion of DNAm has large effects on the HbA1c level, and a small proportion of DNAm is notably associated with adult SES. The results also suggest that adult SES acts through certain important DNAm sites to influence HbA1c.

7 | DISCUSSION

In this article, we develop a Bayesian sparse linear mixed model for high-dimensional mediation analysis. The advantage of a Bayesian method is to propagate uncertainty for functions of parameters in a natural way instead of resorting to delta methods or two-step approaches. Our method can

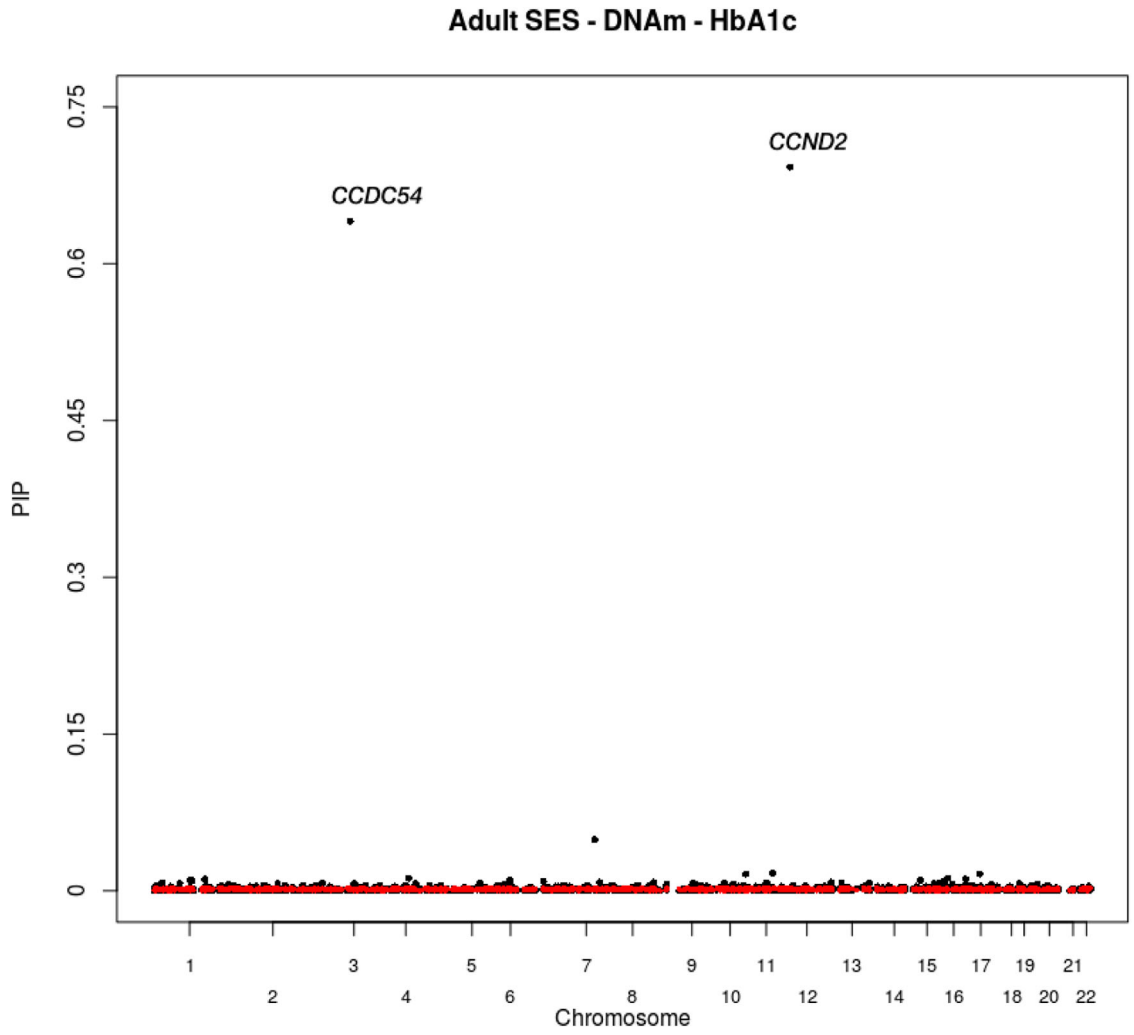

FIGURE 4 Consider the trio: Adult SES → DNAm → HbA1c. The black dots are the estimated PIP for each CpG site from the Bayesian mediation method, and the red dots are the estimated PIPs when we permute the outcome once and fit the Bayesian mediation method

jointly analyze a large number of unordered mediators and characterize their global mediation effect without making any assumptions on their joint distribution. By imposing continuous shrinkage priors on the key regression coefficients in mediation analysis, our method achieves up to 30% power gain in identifying true nonnull mediators compared with the univariate mediation method and approximately 10% power gain over the multivariate methods from simulations. The Bayesian method also provides better interpretations of the way in which a mediator links or does not link exposure to outcome and automatically categorize mediators into four groups based on exposure-mediator and mediator-outcome relationship. Implementing our method to MESA, we have identified two genes, *CCDC54* and *CCND2*, with strong evidence for actively mediating the adult SES effects on HbA1c.

Although our proposed method can simultaneously analyze high-dimensional mediators, like other posterior sampling-based methods, the computation speed is not fast due to the large number of sampling iterations required for reasonable convergence. Also, throughout the article, we focus on one continuous outcome of interest. For binary outcome, we can naively treat it as a quantitative trait, which is justified by recognizing the linear model as a first-order Taylor approximation to a generalized linear model (Zhou *et al.*, 2013). One may hope to adapt our method to directly model binary outcomes through nonlinear link functions, but such an approach will substantially increase the computational cost and may not bring much power gain, as is shown in Zhou *et al.* (2013). 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 in generalized regression models.

Recent literature proposes a convex penalty on the product term of indirect effect (Zhao and Luo, 2016), which improves power of pathway selection and reduces estimation bias in the indirect effects. In the Bayesian framework, direct shrinkage

on the product term may be a more appropriate choice, as it takes into account the correlation between the two models in mediation analysis and is more straightforward when the goal is to identify nonnull mediators. In addition, the biological annotations like pathways can be important predictors for the underlying mediation mechanism, and integrating them into high-dimensional mediation analysis would be promising to facilitate the selection of active mediators. Possible extensions include linking the functional annotation information for mediators to the mediator-specific group probabilities, for example, π_{mj}, π_{aj} for the jth mediator through a logistic regression model (Carbonetto and Stephens, 2013). We leave these interesting extensions for future work.

ACKNOWLEDGMENTS

This work was supported by NSF DMS1712933 (B.M., X.Z.), NIH R01HG009124 (X.Z.), NIH R01HL141292 (J.S.), and NIH R01MD011724 (B.N., D.R.). MESA and the MESA SHARe project are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts HHSN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420, UL1-TR-001881, and DK063491. The MESA Epigenomics & Transcriptomics Study was funded by NHLBI, NIA, and NIDDK grants: 1R01HL101250, R01 AG054474, and R01 DK101921. The authors thank the other investigators, the staff, and the participants of the MESA study for their valuable contributions. A full list of participating MESA investigators and institutions can be found at http://www.mesa-nhlbi.org.

ORCID

Yanyi Song https://orcid.org/0000-0002-7652-6864
Min Zhang https://orcid.org/0000-0003-3331-3583

REFERENCES

Baron, R.M. and Kenny, D.A. (1986) The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173.

Bild, D.E., Bluemke, D.A., Burke, G.L., Detrano, R.Diez Roux, A.V., Folsom, A.R., Greenland, P., Jacob, D.R., Jr, KronmalR. Liu, K., Nelson, J.C., O'Leary, D., Saad, M.F., Shea, S., Szklo, M. and Tracy, R.P.. (2002) Multi-ethnic study of atherosclerosis: objectives and design. *American Journal of Epidemiology*, 156, 871–881.

Boyle, E.A., Li, Y.I. and Pritchard, J.K. (2017) An expanded view of complex traits: from polygenic to omnigenic. *Cell*, 169, 1177–1186.

Carbonetto, P. and Stephens, M. (2013) Integrated enrichment analysis of variants and pathways in genome-wide association studies indicates central role for il-2 signaling genes in type 1 diabetes, and cytokine signaling genes in crohn's disease. *PLoS Genetics*, 9, e1003770.

Chén, O.Y., Crainiceanu, C., Ogburn, E.L., Caffo, B.S., Wager, T.D. and Lindquist, M.A. (2017) High-dimensional multivariate mediation with application to neuroimaging data. *Biostatistics*, 19, 121–136.

Daniel, R., De Stavola, B., Cousens, S. and Vansteelandt, S. (2015) Causal mediation analysis with multiple mediators. *Biometrics*, 71, 1–14.

Derkach, A., Pfeiffer, R.M., Chen, T.H. and Sampson, J.N. (2019) High dimensional mediation analysis with latent variables. *Biometrics*, 75, 745–756.

Elliott, M.R., Raghunathan, T.E. and Li, Y. (2010) Bayesian inference for causal mediation effects using principal stratification with dichotomous mediators and outcomes. *Biostatistics*, 11, 353–372.

Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C.A. and Smoller, J.W. (2019) Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature Communications*, 10, 1776.

Guan, Y. and Stephens, M. (2011) Bayesian variable selection regression for genome-wide association studies and other large-scale problems. *The Annals of Applied Statistics*, 1780–1815.

Huang, Y.-T. and Pan, W.-C. (2016) Hypothesis test of mediation effect in causal mediation model with high-dimensional continuous mediators. *Biometrics*, 72, 402–413.

Imai, K., Keele, L. and Yamamoto, T. (2010) Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science*, 25, 51–71.

Kim, C., Daniels, M., Hogan, J., Choirat, C. and Zigler, C. (2019) Bayesian methods for multiple mediators: relating principal stratification and causal mediation in the analysis of power plant emission controls. arXiv preprint arXiv:1902.06194.

Kim, C., Daniels, M.J., Marcus, B.H. and Roy, J.A. (2017) A framework for Bayesian nonparametric inference for causal effects of mediation. *Biometrics*, 73, 401–409.

Lin, C.-Y., Lin, Y.-C., Kuo, H.-K., Hwang, J.-J., Lin, J.-L., Chen, P.-C. and Lin, L.-Y. (2009) Association between acrylamide, blood insulin and insulin resistance in adults. *Diabetes Care*, 32 (12) , 2206–2211.

MacKinnon, D.P. (2008) *Introduction to Statistical Mediation Analysis*. London: Routledge.

Needham, B.L., Smith, J.A., Zhao, W., Wang, X., Mukherjee, B., Kardia, S.L., Shively, C.A., Seeman, T.E., Liu, Y. and Diez Roux, A.V. (2015) Life course socioeconomic status and DNA methylation in genes related to stress reactivity and inflammation: the multi-ethnic study of atherosclerosis. *Epigenetics*, 10, 958–969.

Robins, J.M. and Greenland, S. (1992) Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, 3 (2) , 143–155.

Rubin, D.B. (1980) Randomization analysis of experimental data: the fisher randomization test comment. *Journal of the American Statistical Association*, 75, 591–593.

Selvin, E., Steffes, M.W., Zhu, H., Matsushita, K., Wagenknecht, L., Pankow, J.Coresh ,J. and Brancati, F.L. (2010) Glycated hemoglobin, diabetes, and cardiovascular risk in nondiabetic adults. *New England Journal of Medicine*, 362, 800–811.

Smith, J.A., Zhao, W., Wang, X., Ratliff, S.M., Mukherjee, B., Kardia, S.L.R.Liu ,Y., Roux, A.V.D. and Needham, B.L. (2017) Neighborhood characteristics influence dna methylation of genes involved in stress response and inflammation: the multi-ethnic study of atherosclerosis. *Epigenetics*, 12, 662–673.

Taguri, M., Featherstone, J. and Cheng, J. (2018) Causal mediation analysis with multiple causally non-ordered mediators. *Statistical Methods in Medical Research*, 27(1), 3–19. https://doi.org/10.1177/0962280215615899.

Ten Have, T.R. and Joffe, M.M. (2012) A review of causal estimation of effects in mediation analyses. *Statistical Methods in Medical Research*, 21, 77–107.

VanderWeele, T.J. (2011) Principal stratification–uses and limitations. *The International Journal of Biostatistics*, 7, 1–14.

VanderWeele, T.J. (2016) Mediation analysis: a practitioner's guide. *Annual Review of Public Health*, 37, 17–32.

VanderWeele, T.J. and Vansteelandt, S. (2010) Odds ratios for mediation analysis for a dichotomous outcome. *American Journal of Epidemiology*, 172, 1339–1348.

VanderWeele, T. and Vansteelandt, S. (2014) Mediation analysis with multiple mediators. *Epidemiologic Methods*, 2, 95–115.

Yaghootkar, H., Stancáková, A., Freathy, R.M., Vangipurapu, J., Weedon, M.N., Xie, W., *et al.* (2015) Association analysis of 29,956 individuals confirms that a low frequency variant at CCND2 halves the risk of type 2 diabetes by enhancing insulin secretion. *Diabetes*, 64(6), 2279–2285. https://doi.org/10.2337/db14-1456.

Zhao, Y. and Luo, X. (2016) Pathway lasso: estimate and select sparse mediation pathways with high dimensional mediators. arXiv preprint, arXiv:1603.07749.

Zhong, J., Agha, G. and Baccarelli, A.A. (2016) The role of dna methylation in cardiovascular risk and disease. *Circulation Research*, 118, 119–131.

Zhou, X., Carbonetto, P. and Stephens, M. (2013) Polygenic modeling with Bayesian sparse linear mixed models. *PLoS Genetics*, 9, e1003264.

SUPPORTING INFORMATION

Web Appendices referenced in Sections 2, 3, 4, 5, and 6 are available with this article at the Biometrics website on Wiley Online Library. The software for implementing our method is available on github: https://github.com/umich-cphds/bama.

How to cite this article: Song Y, Zhou X, Zhang M, *et al*. Bayesian shrinkage estimation of high dimensional causal mediation effects in omics studies. *Biometrics*. 2020;76:700–710. https://doi.org/10.1111/biom.13189