Published in final edited form as:

Hum Hered. 2016 ; 81(1): 26–34. doi:10.1159/000447357.

Leveraging Methylome-Environment Interaction to Detect Genetic Determinants of Disease

Emily Slade^a and Peter Kraft^{a,b}

^aDepartment of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Mass., USA

^bDepartment of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Mass., USA

Abstract

Objective—The association between DNA methylation and a trait of interest may depend on an environmental exposure, and incorrectly accounting for this dependence can lead to a reduction in power of the standard tests used in epigenome-wide association studies. We present the M-ME test to jointly test for the main effect of DNA methylation and methylation-environment interaction.

Methods—Through simulation, we compare the power and type 1 error of the M-ME test to a standard marginal test (M test) and a standard interaction test (ME test) under 1,800 different underlying models. These models allow for methylation-environment correlation and measurement error in the exposure.

Results—In many true underlying models, either the M test or the ME test has very low power, but the M-ME test has optimal or nearly optimal power to detect a DNA methylation effect in all models considered, including those with methylation-environment dependence and measurement error in the exposure. Type 1 error inflation occurs in the tests when the exposure is measured with error and correlated with DNA methylation.

Conclusion—The M-ME test is an attractive choice for studies aiming to detect any DNA methylation association when little is known about the epigenetic associations a priori.

Keywords

DNA methylation; Epigenetics; Epigene-environment interaction; Power and sample size; Measurement error

Introduction

DNA methylation is an epigenetic modification that can affect the expression of a gene without altering the underlying genome. It has been shown to be associated with a wide range of environmental exposures such as smoking [1], age [2], and air pollution [3], as well as complex diseases such as asthma [4], cancer [5], and coronary artery disease [6]. The complex interplay between the genome, epigenome, environment, and disease leads to a variety of plausible causal pathways relating these variables. For instance, DNA methylation

has been observed to mediate the association between maternal smoking during pregnancy and birthweight of the offspring [7] . Alternatively, methylation may act as an effect modifier of an environment-outcome relationship. For example, it has been shown that the association between exposure to traffic-related air pollutants and biomarkers for cardiovascular disease is stronger among subjects with higher Alu methylation status [8] .

Epigenome-wide association studies (EWAS) aim to identify epigenetic variants that are associated with a trait of interest. This can be performed with a standard marginal test for the association between the trait and DNA methylation using a simple regression model at each CpG site. However, suppose there exists an environmental exposure such that among unexposed subjects, higher methylation is associated with higher trait levels, but among exposed subjects it is associated with lower trait levels, or vice versa. The marginal test will have low power to detect the association between methylation and trait levels in this scenario. Similarly, if the association between DNA methylation and the trait is restricted to exposed (or unexposed) individuals, then the marginal test can also have low power. In these scenarios, a standard interaction test for differences in the methylation effect between exposed and unexposed individuals can have greater power than the marginal test. However, the standard interaction test typically has low power except in the most extreme interaction scenarios, and it may miss regions where there is a methylation effect in both exposed and unexposed individuals, although the effects may differ slightly. When the goal of a study is to identify a region of the genome where DNA methylation affects the disease outcome in any way – for example, to find regions where methylation is associated with a trait in at least one exposure subgroup – a test is needed to detect either the marginal effect or methylation-environment interaction. The aim of this study is to design such a test which jointly assesses both of these effects.

Past work in gene-environment interaction considered a similar problem with a binary genotype, environmental exposure, and disease status. Kraft et al. [9] proposed a joint test for genetic main effect or gene-environment interaction that has optimal or nearly optimal power over the standard marginal and interaction tests across a wide range of true penetrance models. Lindström et al. [10] identified that this test has inflated type 1 error in the presence of gene-environment dependence and misclassification of the environmental exposure. This finding is worrisome when introducing a similar test in the epigenetic realm because it is very common for DNA methylation to be associated with an environmental exposure. The strength of this association can also be much greater than gene-environment correlation. Genes typically exhibit only weak correlation with environmental exposures, but it has been well established that many environmental exposures have a direct (and strong) effect on DNA methylation. The results presented by Lindström et al. [10] focus on a binary exposure, genotype, and trait, and their power calculations cannot accommodate continuous data. As such, we expand on their results by developing a test that can accommodate a continuous exposure, DNA methylation, and trait. We also explore the impact of various underlying distributions for the environmental exposure on the performance of this test.

We propose a joint test for methylation-environment interaction that is analogous to the joint test for gene-environment interaction of Kraft et al. [9] . Our joint (M-ME) test accommodates the analytical challenges of continuous epigenetic data, and we compare it to

the marginal DNA methylation (M) test as well as the standard interaction (ME) test. We present the power and type 1 error rates of these tests for a wide range of underlying models varying the effect sizes, exposure distribution, methylation-environment correlation, and measurement error of the environmental exposure.

Methods

We denote the outcome (a continuous trait or biomarker for disease) as Y , the continuous DNA methylation measurement as M , and the continuous environmental exposure as E . We assume that the true underlying model relating these variables has the form

$$E[Y|E, M] = \beta_0 + \beta_e E + \beta_m M + \beta_{me} ME. \quad (1)$$

We consider three tests for association between the environmental exposure, DNA methylation, and disease. We will denote the marginal test for the main effect of methylation as M, the test for methylation-environment interaction as ME, and the joint test for the main effect of methylation and the interaction between methylation and environment as M-ME.

We compare the power and type 1 error rates for the M, ME, and M-ME tests in a wide range of scenarios, considering the presence of methylation-environment correlation as well as measurement error in the exposure. To address measurement error in the exposure, we fit models with E^* instead of E , where E^* denotes the observed environmental exposure, which is different from the true exposure in the presence of measurement error. The M test, which is based on standard linear regression, is a likelihood ratio test comparing a model regressing Y on E^* and M to a model regressing Y on just E^* , which is a one degree of freedom test. The ME test, also based on standard linear regression, and the M-ME test compare a model regressing Y on E^* , M , and $E^* \times M$ to the models regressing Y on E^* and M and on just E^* , respectively, leading to a one degree of freedom likelihood ratio test for the ME test and a two degree of freedom likelihood ratio test for the M-ME test.

The test statistics for these three likelihood ratio tests are χ^2 distributed with non-centrality parameter

$$\delta = 2n \int_{Y, E, M} [\ell_1(\tilde{\beta}_1; Y, E^*, M) - \ell_0(\tilde{\beta}_0; Y, E^*, M)] \times f(Y, E^*, M) \quad (2)$$

where

$$f(Y, E^*, M) = f(Y|E, M)f(M|E)f(E^*|E)f(E).$$

ℓ_1 and ℓ_0 are the log-likelihoods of the observed data under the alternative and null hypotheses, respectively, $\tilde{\beta}_1$ and $\tilde{\beta}_0$ maximize the expected log-likelihoods under each hypothesis, and $f(Y, E^*, M)$ denotes the joint probability density function of outcome Y , observed environmental exposure E^* , and methylation measurement M . To perform these tests in practice, one would compare the test statistic to a null distribution (χ^2 with non-

centrality parameter 0), but to estimate the power or type 1 error of the tests, we must estimate the non-centrality parameter in equation 2.

Because of the continuous nature of Y , E^* , and M , an analytical solution of the integral in equation 2 is not obtainable, in general. If convenient choices of the true underlying model parameters are made, a simple closed-form solution of the non-centrality parameter can be written using the covariance matrix of the estimated coefficients. For example, a centered normal distribution is required for higher-order moments to cancel in the computation. We use this method to prove one simple result in the Appendix, but we instead implement Monte Carlo integration to estimate the non-centrality parameter in all simulations. The Monte Carlo integration method can be used for any selection of underlying model parameters or distributions. We perform this Monte Carlo integration to estimate the non-centrality parameter by simulating data for 1,000,000 observations of (Y, E^*, M) and approximating the integral in equation 2 as an average over these data realizations. The degrees of freedom for the χ^2 distribution of each test statistic is given by $\dim(\tilde{\beta}_1) - \dim(\tilde{\beta}_0)$.

We explore the power and type 1 error of the three tests over a wide range of true underlying models. We first consider the scenario when DNA methylation and the environmental exposure are independent and there is no measurement error in the exposure. Next, we consider the scenario when there is methylation-environment dependence, which is induced in our true underlying model through first simulating the true environmental exposure (E), then simulating the methylation measurement (M) using the relationship

$$E[M|E] = \alpha_0 + \alpha_e E, \quad (3)$$

where α_e controls the proportion of variation in DNA methylation explained by the environmental exposure. Then, E and M are used to simulate the outcome, Y , using equation 1. Lastly, we consider the potential for measurement error in the exposure, where now we use

$$E^* = E + N(0, \sigma_{\text{error}}^2) \quad (4)$$

to represent the observed environmental exposure, where E is the true unobserved exposure and σ_{error}^2 represents the amount of measurement error in the exposure. In the presence of measurement error in E , we consider both type 1 error and power. In summary, data were simulated as follows:

$$\begin{aligned} E &\sim N(0, 1) \\ E^* &= E + N(0, \sigma_{\text{error}}^2) \\ M &= \alpha_0 + \alpha_e E + \varepsilon_m, \end{aligned}$$

where $\varepsilon_m \sim N(0, \sigma_m^2)$

$$Y = \beta_0 + \beta_e E + \beta_m M + \beta_{me} ME + \varepsilon_y,$$

where $\varepsilon_y \sim N(0, \sigma_y^2)$.

For the results presented in the following section, we simulate the true environmental exposure (E) from a standard normal distribution. It is typical to assume E is centered, but there are some consequences of this choice that are explored in the online supplementary material (see www.karger.com/doi/10.1159/000447357 for all online suppl. material). We also provide R code (available at www.hsph.harvard.edu/peter-kraft/software/) that can estimate the power of the three tests for a wide range of exposure distributions, including a binary exposure.

In practice, DNA methylation is measured as the percentage of methylated molecules at a given CpG site, also called a beta-value. Because beta-values are bounded between 0 and 1, it has become increasingly common to transform their distribution to achieve normality [11–13]. One commonly used transformation produces M-values, defined as the logit of the beta-value using a base 2 logarithm instead of a natural logarithm [14]. The use of M-values is more statistically valid than using beta-values since the distribution of M-values is approximately homoscedastic, whereas beta-values demonstrate severe heteroscedasticity in the lower and upper methylation ranges [14]. As a result, we chose to simulate DNA methylation from a standard normal distribution, where these can be thought of as standardized M-values.

For all simulations presented in the main Results section, we set $\alpha_0 = \beta_0 = 0$ and choose σ_m and σ_y such that $\text{Var}(M) = \text{Var}(Y) = 1$. (Results from simulations assuming E is not centered are presented in the online suppl. material). We set $(\beta_0)^2 = 0.025$ to reflect 2.5% of the variation in the outcome (Y) coming from the main effect of the environmental exposure. This value reflects many practical scenarios – for example, the proportion of variation in body mass index coming from a (continuous) aggregate SNP score has been shown to be $\approx 2.7\%$ [15] – but the magnitude of β_e does not affect the relative ranking in power and type 1 error of the M, ME, and M-ME tests. We varied the remaining parameters as follows:

$$\sigma_{\text{error}} \in [0, 1], \alpha_e \in [0, \sqrt{0.5}], \beta_m \in [0, \sqrt{0.01}], \beta_{me} \in [0, \sqrt{0.01}].$$

For all simulations, we fixed the sample size to include 500 subjects. In total, we calculated type 1 error and power in 1,800 different scenarios.

Results

Figure 1 shows the power of the marginal test (M), interaction test (ME), and joint test (M-ME) to detect the association with the outcome when there is no methylation-environment correlation ($\alpha_e = 0$) and no measurement error in the exposure ($\sigma_{\text{error}} = 0$). When there is no main effect of methylation ($\beta_m = 0$) but there is an effect of the interaction between methylation and the environmental exposure ($\beta_{me} > 0$), the marginal test (M) has no power,

and the interaction test (ME) has the highest power. However, the difference in power between the interaction test and the joint test (M-ME) is small. In scenarios when there is little or no interaction effect (small β_{me}) but there is a main effect of methylation, the ME test has very low power, while the M test has the highest power. However, once again, the difference in the power between the most powerful test (M test, in this case) and the M-ME test is small. When there is both a main effect and an interaction in the true underlying model, the M-ME test can outperform both the M and ME tests. It takes only small main and interaction effects for this to occur, for example, $(\beta_m)^2 > 0.0025$ and $(\beta_{me})^2 > 0.0025$.

From previous work in the context of gene-environment testing with binary data, the power of the marginal genetic test did increase as the interaction effect increased [9]. However, our results show that the M test does not gain power as β_{me} increases. This occurs because our simulated exposure distribution is centered around 0. As shown in the online supplementary material, an uncentered exposure distribution does lead to the M test gaining some power with β_{me} , but this does not affect the relative ranking in power of the three tests.

Figure 2 shows how methylation-environment correlation (α_e) affects the power of the M, ME, and M-ME tests when there is still no measurement error in the environmental exposure ($\sigma_{\text{error}} = 0$). Figure 2 includes three panels: one which represents a model with an interaction but no main effect of methylation, one with a main effect of methylation but no interaction, and one where both the main effect of methylation and the interaction with the environment each explain 1% of the variation in the outcome. Across these three plots, we see that in general as the methylation-environment correlation increases, the power to detect the marginal effect of methylation decreases, and the power to detect the interaction increases. The decrease in power of the M test is due to the effect of the correlation between methylation and the environmental exposure on the non-centrality parameter. (See the Appendix for details.) Despite these changes in power across α_e , for a fixed α_e the relative ranking in power of the three tests is similar to the previous scenario displayed in figure 1. For instance, in the presence of marginal and interaction effects where both explain at least 1% of variation in the outcome, the M-ME test outperforms both the M and ME tests even in the presence of fairly strong methylation-environment correlation ($\alpha_e < 0.7$), making the M-ME test still an attractive choice in the presence of methylation-environment dependence.

As shown in figure 3, even in the presence of methylation-environment correlation, all three tests have the nominal 0.05 type 1 error rate when there is no measurement error in the exposure ($\sigma_{\text{error}} = 0$). However, when the environmental exposure is measured with error, our models can only adjust for the mismeasured exposure, E^* . Because DNA methylation and the environmental exposure are correlated, this leads to residual confounding that we are unable to correct for. There is an inflation in the type 1 error rates of the M and M-ME tests as the measurement error in the exposure (σ_{error}) increases, and this is exacerbated as the methylation-environment correlation (α_e) increases across the three plot windows in figure 3. The M test suffers the largest inflation of type 1 error followed by the M-ME test. The ME test maintains the nominal 0.05 type 1 error rate, but this is due to the centering of the simulated environmental exposure. When the distribution of E is uncentered, the ME test does have some inflation in type 1 error, but it still remains the least inflated of the three tests, as shown in the online supplementary material. Very little type 1 error inflation occurs

if methylation and the environmental exposure are only weakly correlated ($(\alpha_e)^2 > 0.01$), even when there is a large amount of measurement error in the exposure ($\sigma_{\text{error}} \in [0, 1]$). However, because many environmental exposures have been shown to have strong correlation with DNA methylation, it is important to consider the impact that measurement error will have on the type 1 error rate in these studies. Furthermore, the inflation of type 1 error also increases as β_e increases. This elevates the importance of considering the impact of measurement error when working with an exposure known to have a large main effect on the trait of interest.

To examine the impact of environmental measurement error on the power of the M, ME, and M-ME tests, we limit our models to those where methylation and the environmental exposure are independent ($\alpha_e = 0$). Although many environmental exposures have been shown to be correlated with DNA methylation, we set $\alpha_e = 0$ to prevent an inflation of type 1 error from driving the patterns seen in the power of the three tests. The power of these tests when there is methylation-environment independence is shown in figure 4 under three different true models: one with an interaction but no main effect of methylation, one with a main effect of methylation but no interaction, and one where both the main effect of methylation and the interaction with the environment each explain 1% of the variation in the outcome. As expected, the ME test suffers a loss of power as measurement error in the exposure increases, but the power of the M test is unaffected by environmental measurement error. The M-ME test has some loss of power as environmental measurement error increases when there is truly an interaction effect ($\beta_{me} > 0$). Despite this loss, we once again see that the M-ME test has optimal or nearly optimal power across a wide range of models, even when the environmental exposure is mismeasured.

In all of these results, the relative ranking of the three tests is not affected by small sample size.

Discussion

When the goal of a study is to discover a CpG site where subjects' DNA methylation at that site has any association with the outcome, be it marginally or perhaps only in the presence of a particular exposure, one has multiple tests at their disposal. We have shown through simulation that the standard marginal test (M test) has no power to detect the methylation effect when it exists only through the interaction with a centered environmental exposure. Furthermore, the standard interaction test (ME test) has lower power than the M test to detect the methylation effect when the effect of methylation is very similar across exposure strata. Thus, when little is known about the epigenetic associations a priori, performing just one of these two tests presents a risk of falling into one of these low-powered scenarios. Performing both the M and ME tests does avoid this problem, but it is not an ideal solution as it worsens the multiple-testing issue that arises when these tests are performed at many CpG sites across the epigenome.

The joint M-ME test is able to avoid the problem of low power across all of the simulated underlying models with a single test. Even in the presence of methylation-environment dependence and measurement error of the environmental exposure, there are many scenarios

in which the M-ME test has higher power than both the M and ME tests. In situations when either the M test or ME test outperforms the M-ME test, the difference in power is small, making the joint M-ME test an attractive option when little is known about the underlying epigenetic associations a priori.

This flexibility is important in epigenetic studies because of the complex relationship of DNA methylation with both exposures and outcomes. It is difficult to determine the true underlying model or causal pathway because DNA methylation could be marginally associated with an outcome, only be associated with the outcome in a certain exposure group, mediate the environmental effect, or quite possibly have a more complex combination of these scenarios. Furthermore, it is common for DNA methylation to be correlated with environmental exposures, and environmental exposures are frequently measured with some amount of error, making it imperative to also consider both of these phenomena in epigenetic studies. It is well-known that failure to correct for measurement error in the exposure leads to biased point estimates of the exposure-outcome association, so care should be taken to use one of the many methods available for correction of this error [16]. When there is also methylation-exposure correlation, the bias due to unobserved confounding can also be seen in the methylation-outcome relationship. The inflation of type 1 error seen in figure 3 demonstrates the importance of considering both of these factors.

The M-ME test is intended for use in a ‘discovery’ EWAS where interest lies in identifying a subset of CpG sites that are associated with the outcome in any way. These particular sites could be further characterized with an independent sample to determine the form of the association using the M and ME tests in a second stage. In this second stage, the multiple-testing burden is significantly reduced, and more subjects can be recruited. To have a valid test of methylation-environment interaction, the second stage must use a sample independent of that used in the ‘discovery’ EWAS.

In some situations, there may be a strong known methylation-outcome relationship, and a ‘discovery’ EWAS may be of less importance. In this case, one may be interested in identifying any environmental exposures that are associated with the outcome either marginally or through exposure-methylation interaction. This effectively reverses the roles of the environmental exposure and DNA methylation in our setting of interest, but our conclusions about the flexibility of the M-ME test still apply. Thus, the M-ME test could be used in this setting to identify exposures that are methylome-specific.

This analysis focused on the performance of the M, ME, and M-ME tests in identifying DNA methylation association when there is a single environmental exposure of interest. If there is reason to believe that the association between DNA methylation at a CpG site and some outcome of interest may be influenced by more than one environmental exposure, these tests can be adapted to accommodate that scenario. One option would be to perform the tests presented here separately for each environmental exposure of interest and correct for the multiple testing. Another option would be to include multiple environmental exposures in a single multivariate test, accounting for each of the appropriate pairwise and higher-order interactions. However, caution must be used as the number of parameters in the model will increase quickly, and an increase in the degrees of freedom may lead to low

power. How best to incorporate multiple candidate exposures is an open question for future research.

Future research should also consider the role that the genome plays in the relationship between the epigenome, environmental exposures, and disease. Our analysis could easily be modified to consider epigenome-genome interaction allowing for DNA methylation-SNP correlation, or more complex causal frameworks could be considered relating disease risk to DNA methylation, genotype, and environmental exposures. Furthermore, we considered the effect of measurement error in the exposure, but one should also take care to consider error in the measurement of DNA methylation.

This paper shows that the joint M-ME test has optimal or nearly optimal power to detect a DNA methylation effect in a wide range of scenarios including those involving methylation-environment dependence and measurement error in the exposure. This makes the M-ME test an attractive choice to seek out a specific region of the epigenome where changes in DNA methylation affect disease. We provide an R function at <http://www.hsph.harvard.edu/peter-kraft/software/> which computes the power of the M test, ME test, and M-ME test for a user-specified sample size and underlying model. This tool can assist in the planning of a study by allowing investigators to explore the relative performance of the tests for their scenario of interest.

Acknowledgments

E.S. was supported by NIH grants 2T32AI007358-26 and 2T32GM074897-11. P.K. was supported by NIH/NCI grant R21 CA165920-01.

Appendix

Let $\sigma_E^2 = \text{Var}(E)$ and $\sigma_M^2 = \text{Var}(M) = \alpha_e^2 \sigma_E^2 + \sigma_m^2$. For all simulations, E and M are centered so that $E[E] = E[M] = 0$.

Power of the M test

When fitting the model $Y \sim E + M$, the design matrix, $\mathbf{X}$, is $[1 \ E \ M]^T$ and

$$\begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_e \\ \widehat{\beta}_m \end{bmatrix} \sim N \left(\begin{bmatrix} \beta_0 \\ \beta_e \\ \beta_m \end{bmatrix}, \Sigma \right),$$

where

$$\Sigma \propto (E[\mathbf{X}\mathbf{X}^T])^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sigma_E^2 & \alpha_e \sigma_E^2 \\ 0 & \alpha_e \sigma_E^2 & \alpha_e^2 \sigma_E^2 + \sigma_m^2 \end{bmatrix}^{-1}.$$

For all simulations, we set $\sigma_E^2 = 1$ and $\sigma_m^2 = 1 - \alpha_e^2$. This implies

$$\Sigma \propto \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{1-\alpha_e^2} & \frac{-\alpha_e}{1-\alpha_e^2} \\ 0 & \frac{-\alpha_e}{1-\alpha_e^2} & \frac{1}{1-\alpha_e^2} \end{bmatrix}.$$

Thus, the non-centrality parameter for the M test is

$$\frac{\beta_m^2}{\left(\frac{1}{1-\alpha_e^2}\right)} = \beta_m^2 (1 - \alpha_e^2).$$

As methylation-environment correlation increases, α_e increases, so the non-centrality parameter of the M test decreases. This gives theoretical justification for the decrease in power of the M test as methylation-environment correlation increases.

Power of the ME test

In fitting the model $Y \sim E + M + M \times E$, the design matrix, $\mathbf{X}^*$, is $[1 \ E \ M \ ME]^T$ and

$$\begin{bmatrix} \tilde{\beta}_0 \\ \tilde{\beta}_e \\ \tilde{\beta}_m \\ \tilde{\beta}_{me} \end{bmatrix} \sim N \left(\begin{bmatrix} \beta_0^* \\ \beta_e^* \\ \beta_m^* \\ \beta_{me}^* \end{bmatrix}, \Sigma^* \right).$$

Writing the i -th moment of E as $m_E(i)$, we have

$$\Sigma^* \propto \left(E \left[\mathbf{X}^* \mathbf{X}^{*T} \right] \right)^{-1} = \begin{bmatrix} 1 & 0 & 0 & \alpha_e \sigma_E^2 \\ 0 & \sigma_E^2 & \alpha_e \sigma_E^2 & \alpha_e m_E(3) \\ 0 & \alpha_e \sigma_E^2 & \alpha_e^2 \sigma_E^2 + \sigma_m^2 & \alpha_e^2 m_E(3) \\ \alpha_e \sigma_E^2 & \alpha_e m_E(3) & \alpha_e^2 m_E(3) & \alpha_e^2 m_E(4) + \sigma_m^2 \sigma_m^2 \end{bmatrix}^{-1}.$$

For all simulations, we set $\sigma_E^2 = 1$, $\sigma_m^2 = 1 - \alpha_e^2$, and $E \sim N(0, 1)$, so $m_E(3) = 0$ and $m_E(4) = 3$. This implies

$$\Sigma^* \propto \begin{bmatrix} \frac{1+2\alpha_e^2}{1+\alpha_e^2} & 0 & 0 & \frac{\alpha_e^3 - \alpha_e}{1-\alpha_e^4} \\ 0 & \frac{1}{1-\alpha_e^2} & \frac{-\alpha_e^3 - \alpha_e}{1-\alpha_e^4} & 0 \\ 0 & \frac{-\alpha_e^3 - \alpha_e}{1-\alpha_e^4} & \frac{1}{1-\alpha_e^2} & 0 \\ \frac{\alpha_e^3 - \alpha_e}{1-\alpha_e^4} & 0 & 0 & \frac{1}{1+\alpha_e^2} \end{bmatrix}.$$

Thus, the non-centrality parameter for the ME test is

$$\frac{\beta_{me}^*}{\left(\frac{1}{1+\alpha_e^2}\right)} = \beta_{me}^{*2} (1 + \alpha_e^2).$$

As methylation-environment correlation increases, α_e increases, so the non-centrality parameter of the ME test increases. This gives theoretical justification for the increase in power of the ME test as methylation-environment correlation increases.

References

1. Tsaprouni LG, Yang TP, Bell J, Dick KJ, Kanoni S, Nisbet J, Vinuela A, Grundberg E, Nelson CP, Meduri E, Buil A, Cambien F, Hengstenberg C, Erdmann J, Schunkert H, Goodall AH, Ouwehand WH, Dermitzakis E, Spector TD, Samani NJ, Deloukas P. Cigarette smoking reduces DNA methylation levels at multiple genomic loci but the effect is partially reversible upon cessation. *Epigenetics*. 2014; 9:1382–1396. [PubMed: 25424692]
2. Christensen BC, Houseman EA, Marsit CJ, Zheng S, Wrensch MR, Wiemels JL, Nelson HH, Karagas MR, Padbury JF, Bueno R, Sugarbaker DJ, Yeh RF, Wiencke JK, Kelsey KT. Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLoS Genet*. 2009; 5:e1000602. [PubMed: 19680444]
3. Baccarelli A, Wright RO, Bollati V, Tarantini L, Litonjua AA, Suh HH, Zanobetti A, Sparrow D, Vokonas PS, Schwartz J. Rapid DNA methylation changes after exposure to traffic particles. *Am J Respir Crit Care Med*. 2009; 179:572–578. [PubMed: 19136372]
4. Gunawardhana LP, Gibson PG, Simpson JL, Benton MC, Lea RA, Baines KJ. Characteristic DNA methylation profiles in peripheral blood monocytes are associated with inflammatory phenotypes of asthma. *Epigenetics*. 2014; 9:1302–1316. [PubMed: 25147914]
5. Feinberg AP, Tycko B. The history of cancer epigenetics. *Nat Rev Cancer*. 2004; 4:143–153. [PubMed: 14732866]
6. Sharma P, Garg G, Kumar A, Mohammad F, Kumar SR, Tanwar VS, Sati S, Sharma A, Karthikeyan G, Brahmachari V, Sengupta S. Genome wide DNA methylation profiling for epigenetic alteration in coronary artery disease patients. *Gene*. 2014; 541:31–40. [PubMed: 24582973]
7. Kupers LK, Xu X, Jankipersadsing SA, Vaez A, la Bastide-van Gemert S, Scholtens S, Nolte IM, Richmond RC, Relton CL, Felix JF, Duijts L, van Meurs JB, Tiemeier H, Jaddoe VW, Wang X, Corpeleijn E, Snieder H. DNA methylation mediates the effect of maternal smoking during pregnancy on birthweight of the offspring. *Int J Epidemiol*. 2015; 44:1224–1237. [PubMed: 25862628]
8. Bind MA, Baccarelli A, Zanobetti A, Tarantini L, Suh H, Vokonas P, Schwartz J. Air pollution and markers of coagulation, inflammation, and endothelial function: associations and epigenvironment interactions in an elderly cohort. *Epidemiology*. 2012; 23:332–340. [PubMed: 22237295]
9. Kraft P, Yen YC, Stram DO, Morrison J, Gauderman WJ. Exploiting gene-environment interaction to detect genetic associations. *Hum Hered*. 2007; 63:111–119. [PubMed: 17283440]
10. Lindström S, Yen YC, Spiegelman D, Kraft P. The impact of gene-environment dependence and misclassification in genetic association studies incorporating gene-environment interactions. *Hum Hered*. 2009; 68:171–181. [PubMed: 19521099]
11. Ferfour F, Boitrelle F, Ghout I, Albert M, Molina Gomes D, Wainer R, Bailly M, Selva J, Vialard F. A genome-wide DNA methylation study in azoospermia. *Andrology*. 2013; 1:815–821. [PubMed: 23996935]
12. Lleras RA, Smith RV, Adrien LR, Schlecht NF, Burk RD, Harris TM, Childs G, Prystowsky MB, Belbin TJ. Unique DNA methylation loci distinguish anatomic site and HPV status in head and neck squamous cell carcinoma. *Clin Cancer Res*. 2013; 19:5444–5455. [PubMed: 23894057]
13. Schlecht NF, Ben-Dayan M, Anayannis N, Lleras RA, Thomas C, Wang Y, Smith RV, Burk RD, Harris TM, Childs G, Ow TJ, Prystowsky MB, Belbin TJ. Epigenetic changes in the CDKN2A locus are associated with differential expression of P16INK4A and P14ARF in HPV-positive oropharyngeal squamous cell carcinoma. *Cancer Med*. 2015; 4:342–353. [PubMed: 25619363]
14. Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, Lin SM. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics*. 2010; 11:587. [PubMed: 21118553]

15. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015; 518:197–206. [PubMed: 25673413]
16. Keogh RH, White IR. A toolkit for measurement error correction, with a focus on nutritional epidemiology. *Stat Med*. 2014; 33:2137–2155. [PubMed: 24497385]

Fig. 1.

The power of the ME and M-ME tests increases with the effect of methylation-environment interaction (β_{me}), and the power of the M and M-ME tests increases with the marginal effect of methylation (β_m) when there is no methylation-environment dependence and no measurement error in the exposure.

Fig. 2.

When there is no marginal methylation effect ($\beta_m = 0$), the ME and M-ME tests gain power as the methylation-environment correlation (α_e) increases. When there is no interaction effect ($\beta_{me} = 0$), the M and M-ME tests lose power as the methylation-environment correlation increases. Nevertheless, the M-ME test still outperforms the M and ME tests in the presence of both marginal and interaction effects.

Fig. 3.

The type 1 error rates in the M and M-ME tests increase with measurement error in the exposure (σ_{error}), and the rate of inflation increases with methylation-environment correlation (α_e).

Fig. 4.

Under methylation-environment independence, the power levels of the ME and M-ME tests decrease with measurement error in the exposure (σ_{error}). Despite the M test being robust to this reduction in power, the M-ME test still outperforms the M test in the presence of both a marginal effect of methylation (β_m) and an interaction effect (β_{me}).