ARTICLE

Constrained Score Statistics Identify Genetic Variants Interacting with Multiple Risk Factors in Barrett's Esophagus

James Y. Dai,[1,2,*] Jean de Dieu Tapsoba,[1] Matthew F. Buas,[1] the BEACON Consortium, Harvey A. Risch,[3] and Thomas L. Vaughan[1,4]

Few gene-environment interactions (G × E) have been discovered in cancer epidemiology thus far, in part due to the large number of possible G × E to be investigated and inherent low statistical power of traditional analytic methods for discovering G × E. We consider simultaneously testing for interactions between several related exposures and a genetic variant in a genome-wide study. To improve power, constrained testing strategies are proposed for multivariate gene-environment interactions at two levels: interactions that have the same direction (one-sided or bidirectional hypotheses) or are proportional to respective exposure main effects (a variant of Tukey's one-degree test). Score statistics were developed to expedite the genome-wide computation. We conducted extensive simulations to evaluate validity and power performance of the proposed statistics, applied them to the genetic and environmental exposure data for esophageal adenocarcinoma and Barrett's esophagus from the Barretts Esophagus and Esophageal Adenocarcinoma Consortium (BEACON), and discovered three loci simultaneously interacting with gastresophageal reflux, obesity, and tobacco smoking with genome-wide significance. These findings deepen understanding of the genetic and environmental architecture of Barrett's esophagus and esophageal adenocarcinoma.

Introduction

With the exception of rare familial cancers, cancer risk is usually determined by a complex interplay between multiple genetic and environmental factors.[1] Alterations in somatic genome can result from exposure to external agents (e.g., tobacco smoking) and infectious agents (e.g., *Helicobacter pylori*) and from the multifaceted effects of certain host characteristics (e.g., obesity), all generally defined to be "environmental exposures" in this report. Individuals differ in their inherited efficiency to neutralize environmental insults and repair genomic damages. Efforts have long been focused on discovering susceptibility genes involved in carcinogenesis and on characterizing how these genes interact with environmental factors.[2] Agnostic searches for gene-environment interactions (G × E) have been routinely conducted in genome-wide studies,[3] but with very limited success.[4,5] Much discussion has been devoted to methodological issues,[6,7] such as design and estimation,[8–11] power and sample size,[12] and mathematical formulation of interactions.[13]

When multiple environmental factors have been established for a cancer and are available for G × E testing, the standard analytical approach is to interrogate one pair of genotype and exposure at a time and subsequently correct for multiple testing.[14,15] This strategy is well interpretable, but not optimal when environmental risk factors are related in cancer etiology, so that genotypes involved in the same etiological pathway may interact with all related exposures. This article develops constrained inferential methods for testing multiple G × E simultaneously, leveraging known biological knowledge on cancer etiology. The methods are motivated by a genome-wide G × E study examining esophageal adenocarcinoma and Barrett's esophagus (BE/EA [MIM: 614266]) from the Barretts Esophagus and Esophageal Adenocarcinoma Consortium (BEACON), an international consortium composed of studies in Australia, Europe, and North America, where the incidence of esophageal adenocarcinoma has been rising sharply in the past four decades.[16–19] Esophageal adenocarcinoma cases are believed to arise from Barrett's esophagus, a precursor lesion,[20] and both conditions share risk factors including gastresophageal reflux symptoms (GERD [MIM: 109350]), obesity, and tobacco smoking.[21,22] The results of genome-wide genetic association and a limited search for G × E among known susceptibility loci in BEACON have been reported elsewhere.[23,24]

For esophageal adenocarcinoma and its precursor lesion (Barrett's esophagus), all three well-established risk factors—obesity, gastresophageal reflux, and tobacco smoking—have been linked to local and systemic inflammation and the downstream consequence of oxidative stress that promotes DNA damage and chromosomal instability.[20] It is plausible that SNPs in inflammation pathways may modulate induced defense mechanisms against damage caused by these exposures, thereby interacting with all three exposures in a similar fashion. Furthermore, obesity is associated with gastresophageal reflux in the white population.[25] Adjustments of p values from testing one

exposure-genotype interaction at a time via Bonferroni correction can be overly conservative. In contrast, simultaneously testing for G × E among multiple environmental factors potentially pools small associations and increases power.

To improve power, we consider multivariate G × E testing strategies in which the directions of multiple G × E interactions under the alternative hypothesis are constrained. Suppose $\theta_1, ..., \theta_p$ are interaction parameters between a genetic variant and p established environmental factors in a logistic regression model for a cancer case-control study. The null hypothesis that there is no G × E for any of environmental factors is represented by

$$H_0: \quad \theta_j = 0 \text{ for } j = 1, ..., p. \qquad \text{(Equation 1)}$$

Classical multivariate tests such as Hotelling's T^2 test statistics are designed to test any departure from the null hypothesis; they therefore lack power to detect specific types of departure that may be considered plausible based on a priori scientific knowledge, e.g., the inflammation pathway in the etiology of esophageal adenocarcinoma. Restricting alternative hypothesis to be one-sided, $\theta_j < 0$ for some j, for example SNPs in inflammatory pathways may be associated with a lower odds ratio for any of abdominal obesity, gastresophageal reflux, and cigarette smoking, will presumably boost the power. The constraint can also be motivated for the other direction: $\theta_j > 0$. In the form of likelihood ratio tests, this kind of constraint has been investigated for clinical trials with multiple study endpoints.[26,27] However, the critical values for computing p values under the asymptotic null distribution of these one-sided statistics are generally difficult to obtain, and maximizing the likelihood under the one-sided constraint can be cumbersome for genome-wide testing.

In the same vein, a more restrictive, but parsimonious, model for interactions was proposed based on Tukey's 1-df test:[28–30] gene-gene or gene-environment interactions are proportional to main effects by a fixed constant. In essence, risk associations of genetic variants or environmental exposures are first summed together by a genetic risk score or an environmental risk score, assuming that each group of risk variables share common underlying etiological pathways. More specifically, this test can be motivated by a latent-variable model, where each group of measured markers serves as a surrogate of an underlying biological phenotype.[30] A single interaction parameter then summarizes the deviation from the additive contributions of two risk predispositions, possibly in a transformed risk scale.[30] Built on this parsimonious model, it has been proposed to jointly test for both genetic associations and gene-gene or gene-environment interactions.[30] A complication of this model is that, when there is no genetic association, as is true for the great majority of variants being examined in genome-wide studies, the interaction parameter vanishes and is not identifiable. The usual hypothesis testing statistics are no longer applicable and specialized procedures have to be developed with simulated null distributions.

In this article, we develop score statistics with aforementioned constraints for multiple G × E in a genome-wide study, because score statistics are computationally attractive as regression parameters are estimated only under the null and, further, they have the same local power as Wald and likelihood ratio statistics. We extend the classical one-sided test to a bidirectional test, in which multiple G × E can be either negative or positive, so that one does not have to specify the direction one at a time. For both one-sided test and bidirectional test, we propose score statistics that are easy to compute and have an exact or easily obtainable asymptotic distribution. For the proportional G×E testing, we modify Tukey's 1-df model to allow for the interaction not dependent on the genetic main association, and we derive finite-sample correction for the mean and the variance of the score statistic. Such correction, though often ignorable at the usual significance level for one test (type I error rate 0.05), is critical to control genome-wide type I errors, as we show in simulations and the BEACON study. We evaluate the performance of the proposed methods in extensive simulations and apply them in the genome-wide G × E study from BEACON.

Material and Methods

Consider a genetic association study with n participants. Let Y denote the disease status, for example cancer patients versus control subjects. Let $\mathbf{X}$ denote the vector of p known risk-associated exposures, and let G denote one of many SNPs under investigation, possibly in a genome-wide study. Typically, the study also assembles covariates including age, gender, and top principal components of genetic variation for controlling population stratification, collectively denoted by $\mathbf{W}$. The interest hereafter is to discover whether G modifies risk associations of some exposures in $\mathbf{X}$. The data therefore consist of n independent and identically distributed random vectors $(Y_i, \mathbf{X}_i, G_i, \mathbf{W}_i)$, for $i = 1, ..., n$.

Consider a regression model investigating p statistical interactions simultaneously,

$$g\{\mathbb{E}(Y \mid G, \mathbf{X}, \mathbf{W})\} = \beta_0 + \beta_1 G + \sum_{j=1}^{p} \beta_{2j} X_j + \beta_3 \mathbf{W} + \sum_{j=1}^{p} \theta_j X_j G,$$
(Equation 2)

where β_1 is the genetic association when all exposures are 0, β_{2j} is the association of the jth exposure when the genotype score is 0, θ_j is the interaction between G and X_j, and g is the logit function if Y is a dichotomous outcome or the identical function if Y is a quantitative trait modeled by linear regression.

The null hypothesis for multiple G × E in Equation 2 is expressed in Equation 1, stating that none of p exposure-risk associations changes with the genotype G. Classical multivariate tests are designed to be unrestricted in the alternative hypothesis

$$H_{1a}: \quad \theta_j \neq 0 \text{ for at least one } j.$$

Denote the parameters of interest $\boldsymbol{\theta} = (\theta_1, ..., \theta_p)$ and the nuisance parameters $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_{21}, ..., \beta_{2p}, \beta_3)$. Let $\boldsymbol{\alpha} = (\boldsymbol{\beta}, \boldsymbol{\theta})$, $f(y; \boldsymbol{\alpha})$ denote the density of Y given covariates, and let

$$\ell(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \log f(Y_i; \boldsymbol{\alpha}),$$
$$S(\boldsymbol{\alpha}) = (\partial/\partial\boldsymbol{\alpha})\ell(\boldsymbol{\alpha}),$$
$$\text{and } \mathcal{I}(\boldsymbol{\alpha}) = \mathbb{E}_{\boldsymbol{\alpha}}\left\{(\partial/\partial\boldsymbol{\alpha})\ell(\boldsymbol{\alpha})(\partial/\partial\boldsymbol{\alpha}^T)\ell(\boldsymbol{\alpha})\right\}$$

denote the log-likelihood, the score function, and the information matrix, respectively. $S(\boldsymbol{\alpha})$ is partitioned into (S_{β}, S_{θ}) to conform with the partition $(\boldsymbol{\beta}, \boldsymbol{\theta})$ of $\boldsymbol{\alpha}$. Similarly, $\mathcal{I}(\boldsymbol{\alpha})$ is partitioned into block matrices $\mathcal{I}_{\beta\beta}, \mathcal{I}_{\beta\theta}, \mathcal{I}_{\theta\beta}, \mathcal{I}_{\theta\theta}$. Let $\tilde{\boldsymbol{\alpha}}$ denote the maximum likelihood estimator under the null hypothesis H_0. Classical Rao's score statistic for testing H_{1a} is

$$T_a = \tilde{S}_{\theta}^T\left(\tilde{\mathcal{I}}_{\theta\theta} - \tilde{\mathcal{I}}_{\theta\beta}\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{\mathcal{I}}_{\beta\theta}\right)^{-1}\tilde{S}_{\theta},$$

where $\tilde{S}$ and $\tilde{\mathcal{I}}$ are evaluated at $\tilde{\boldsymbol{\alpha}}$, respectively. When the likelihood is correctly specified, T_a has a χ_p^2 distribution in large samples under the null hypothesis.

Testing for One-Sided or Bidirectional Hypothesis

As we elaborated in the Introduction, it is sometimes plausible in cancer epidemiology to test for one-sided multiple G × E simultaneously, for example

$$H_{1b}: \quad \theta_j \leq 0 \text{ for } j = 1, \ldots, p$$

with strict inequality for at least one j. The interest here is specifically in discovering genetic alleles that simultaneously link to a lower association for each of p exposures. The one-sided test can be motivated similarly for $\theta_j \geq 0$, aiming to discover genetic alleles that link to a higher association for exposures. In a complete agnostic search such as genome-wide testing for G × E, it is perhaps more sensible to formulate the following bidirectional hypothesis

$$H_{1c}: \quad \theta_j \leq 0 \text{ for } j = 1, \ldots, p \text{ or } \theta_j \geq 0 \text{ for } j = 1, \ldots, p$$

with strict inequality for at least one j. The multiple G × E are constrained to be directional, either all positive or all negative. This bidirectional test has not been seen in the statistical literature. In what follows, we derive score tests for H_{1b} and H_{1c} based on an approximated likelihood ratio test.[27]

Let $\tilde{\boldsymbol{\beta}}$ denote the maximum likelihood estimator of $\boldsymbol{\beta}$ under H_0, and let $\tilde{\boldsymbol{\alpha}} = (\tilde{\boldsymbol{\beta}}, \mathbf{0})$, where $\mathbf{0}$ is a length-p zero vector for $\boldsymbol{\theta}$. Let a sequence of local alternative hypotheses be defined by

$$H_{1n}: \quad \boldsymbol{\theta} = n^{-1/2}\boldsymbol{\delta},$$

where $\boldsymbol{\delta}$ is a fixed point in $\mathbb{R}^p$. By employing Taylor series expansions, we have

$$n^{-1/2}S_{\theta}(\tilde{\boldsymbol{\alpha}}) \to_d \mathcal{N}(\mathcal{I}_{\theta\theta,\beta}(\boldsymbol{\alpha}_0)\boldsymbol{\delta}, \mathcal{I}_{\theta\theta,\beta}(\boldsymbol{\alpha}_0))$$

and

$$n^{-1/2}\mathcal{I}_{\theta\theta,\beta}^{-1}(\tilde{\boldsymbol{\alpha}})S_{\theta}(\tilde{\boldsymbol{\alpha}}) \to_d \mathcal{N}\left(\boldsymbol{\delta}, \mathcal{I}_{\theta\theta,\beta}^{-1}(\boldsymbol{\alpha}_0)\right),$$

under H_{1n}, where $\boldsymbol{\alpha}_0 = (\boldsymbol{\beta}, 0)$, $\mathcal{I}_{\theta\theta,\beta} = \mathcal{I}_{\theta\theta} - \mathcal{I}_{\theta\beta}\mathcal{I}_{\beta\beta}^{-1}\mathcal{I}_{\beta\theta}$. Let $U = n^{-1/2}\mathcal{I}_{\theta\theta,\beta}^{-1}(\tilde{\boldsymbol{\alpha}})S_{\theta}(\tilde{\boldsymbol{\alpha}})$. Since $\boldsymbol{\theta} \leq 0$ is equivalent to $\boldsymbol{\delta} \leq 0$, one may define a score test as follows:

$$T_s = n\left[U^T\mathcal{I}_{\theta\theta,\beta}U - \inf\left\{(U - \boldsymbol{\delta})^T\mathcal{I}_{\theta\theta,\beta}(U - \boldsymbol{\delta}) : \boldsymbol{\delta} \leq 0\right\}\right].$$

This is essentially the likelihood ratio statistic for $\boldsymbol{\delta} \leq 0$ based on a single realization of U:[31] the first term is the log-likelihood of U under the null hypothesis $\boldsymbol{\delta} = 0$, and the second term is the

maximum log-likelihood of U under the alternative hypothesis $\boldsymbol{\delta} \leq 0$. If the information matrix $\mathcal{I}_{\theta\theta,\beta}$ is known, the asymptotic null distribution of T_s is a chi-bar-squared distribution that is equivalent to a weighted sum of a sequence of chi-square distributions, where the weights can be computed by algebra in closed forms when $p \leq 4$[31,32] and by simulations when $p > 4$. Alternatively, if $\mathcal{I}_{\theta\theta,\beta}$ is diagonal, close-form expressions of the weights are readily available for any p. In our case, however, $\mathcal{I}_{\theta\theta,\beta}$ is unknown, depending on $\boldsymbol{\beta}$ that has to be estimated. Furthermore, it is unlikely to be a diagonal matrix in the context of testing G × E. Consequently the asymptotic null distribution of T_s is difficult to obtain.

We therefore develop an approximate score statistic for one-sided hypotheses, in the same spirit as the approximate likelihood ratio statistic for a normal mean vector with nonnegative components.[27] The main advantage is that, after transformation and approximation, the asymptotic null distribution of the score statistic does not rely on the nuisance parameters involved in $\mathcal{I}_{\theta\theta,\beta}$, and therefore the significance level is much easier to compute, an appealing feature for genome-wide testing.

Let A be a $p \times p$ matrix such that

$$A^TA = \mathcal{I}_{\theta\theta,\beta}, \qquad \text{(Equation 3)}$$

and so $z = AU$ is asymptotically a normal vector with the identity covariance matrix. Such A is not unique, as we discuss the choice of A in the next paragraph. For a chosen A, the score statistic T_s becomes

$$n\left[z^Tz - \inf\left\{(z - A\boldsymbol{\delta})^T(z - A\boldsymbol{\delta}) : \boldsymbol{\delta} \leq 0\right\}\right].$$

Let $b = A\boldsymbol{\delta}$. Geometrically, the parameter space under the alternative is the negative orthant. To compute T_s, one has to minimize $(z - b)^T(z - b)$ for all vectors in the image space $A(\boldsymbol{\delta}) = \{A\boldsymbol{\delta} \mid \boldsymbol{\delta} \leq 0\}$, which can be cumbersome. Instead, the proposed test approximates the polyhedral space $A(\boldsymbol{\delta})$ by the negative orthant $\{b \mid b \leq 0\}$, with A chosen to make the center directions of the two spaces coincide. For the jth element in z, denoted by z_j, if $z_j \geq 0$, $\inf\{(z_j - b_j)^2 : b_j \leq 0\} = z_j^2$; if $z_j < 0$, $\inf\{(z_j - b_j)^2 : b_j \leq 0\} = 0$. Because z_j and $z_{j'}$ are independent, it follows that the approximate score statistic for one-sided hypothesis H_{1b} is

$$T_b = \sum_{j=1}^{p} \left(z_j \wedge 0\right)^2,$$

where $z_j \wedge 0$ denote the minimum of z_j and 0. Because of the independence among z_j values, the asymptotic null distribution is a special case of the chi-bar-squared distribution,[27] expressed as $\sum_{j=0}^{p}[\{p!/j!(p-j)!\}/2^p]\chi_j^2$, where χ_j^2 is the chi-square distribution with j degrees of freedom and χ_0^2 is defined as the constant zero.

The degree of approximation depends on the information matrix $\mathcal{I}_{\theta\theta,\beta}$. If it is diagonal, A can be chosen such that $A(\boldsymbol{\delta})$ is exactly the negative orthant. For an arbitrary information matrix, we can enhance the approximation by choosing the transformation A so that the center direction of $A(\boldsymbol{\delta})$, which is defined to be the direction that forms the same acute angle with any of the edge directions, coincides with the center direction of the negative orthant. For the negative orthant in the two-dimensional space, for instance, the center direction is $(-1, -1)$. For a given A, the p edges of the polyhedral cone are given by

$$f_j = Ae_j\big/\left(e_j^T\mathcal{I}_{\theta\theta,\beta}e_j\right)^{1/2},$$

with $j = 1, 2, \ldots, p$ and e_j is the unit vector for the jth coordinate. The center direction is then given by

$$d_A = \left\{ \left(f_1, \ldots, f_p \right)^T \right\}^{-1} J,$$

where $J = (-1, \ldots, -1)^T$, the center direction of the negative orthant. Therefore the center direction of A is required to be

$$d_A = cJ, \qquad \text{(Equation 4)}$$

for some $c > 0$. In Appendix A, we give the details of how to construct A to ensure $A(\delta)$ has the same center direction as the negative orthant.

Following the same derivation, the score statistic for the other one-sided test, $\theta_j \geq 0$ for $j = 1, \ldots, p$ can be constructed as $\sum_{j=1}^{p} (z_j \vee 0)^2$, where $z_j \vee 0$ denote the maximum of z_j and 0. The score statistic for the bidirectional hypothesis H_{1c} is therefore the maximum of the two test statistics,

$$T_c = \max \left(\sum_{j=1}^{p} \left(z_j \wedge 0 \right)^2, \sum_{j=1}^{p} \left(z_j \vee 0 \right)^2 \right).$$

The test statistic now depends on the angle of z_j in orthants other than positive and negative ones, and so the asymptotic null distribution is no longer a chi-bar-squared distribution. Since under the null hypothesis z_j is a vector of independent and standard normal distributions, the null distribution of T_c can be generated by simulation. For genome-wide testing, however, an enormous number of simulations need to be generated to achieve the accuracy needed for genome-wide significance (5×10^{-8}). Though this requires the generation of a large number of simulated statistics (in the order of ~10^9–10^{10}) only once, it is rather cumbersome for genome-wide testing. We propose a simple hybrid procedure that computes p values with adequate accuracy: first, a p value is computed by simulations to the accuracy of the third decimal place; if the simulation-based p value is less than 0.01, then we further compute an approximate, conservative p value by algebraic calculation. The detail of the algebraic approximation is shown in Appendix A. In Figures S1 and S2, we show some numerical analyses comparing simulation-based p values and algebraic approximations. This algebraic approximation works remarkably well: the smaller p value, the better approximation.

Testing for Proportional Interactions

Relative to the one-sided test, a more restrictive alternative hypothesis for multiple G × E in Equation 2 entails

$$H_{1d}: \quad \theta_j = \gamma \beta_{2j} \text{ for some } \gamma \neq 0 \text{ and } j = 1, \ldots, p,$$

so that in the log-odds scale, the interactions are proportional to the exposure main associations. One way to interpret this hypothesis is as follows: because of the shared etiology, the combinatory effect of environmental risk factors can be formulated as an environmental risk score, similar to genetic risk scores commonly used.[33] The genetic variant modifies the disease association of this environmental risk score as a whole. This test is slightly different from Tukey's 1-df test because the genetic main association, β_1, does not factor in the interactions.[28] For genome-wide association studies, β_1 and γ will be 0 for the majority of loci. Given that the exposures are all established risk factors, i.e., $\beta_{2j} \neq 0$, the interaction parameter γ is identifiable even if there is no genetic main effect under the null hypothesis $\beta_1 = 0$, avoiding the troublesome identifiability issue.[30]

The parameters are reduced to the vector $\alpha = (\beta, \gamma)$. The standard Rao score statistic for testing $\gamma = 0$, however, shows severe inflated type I errors in the BEACON genome-wide association study (see Figure S3, the quantile-quantile plot). The problem originates in the unique score function derived from the proportional interaction model. The usual expressions for the mean and variance of $\tilde{S}_\gamma$ need to be corrected for the finite-sample bias, as we derive below.

Let $\mathcal{X}_i = (1, G_i, \mathbf{X}_i, W_i)$ be the vector of covariates, and let $\tilde{\beta}$ denote the maximum likelihood estimator for β under the null hypothesis. The score function for β and γ under the null hypothesis can be written as

$$\tilde{S}_\beta = \sum_i \mathcal{X}_i \left(Y_i - \frac{\exp(\mathcal{X}_i \tilde{\beta})}{1 + \exp(\mathcal{X}_i \tilde{\beta})} \right),$$

$$\tilde{S}_\gamma = \sum_i \sum_j \tilde{\beta}_{2j} G_i X_{ij} \left(Y_i - \frac{\exp(\mathcal{X}_i \tilde{\beta})}{1 + \exp(\mathcal{X}_i \tilde{\beta})} \right).$$

Let $\tilde{\mathcal{H}}_{\gamma\beta} = \partial \tilde{S}_\gamma / \partial \beta$. Asymptotic distribution theories for maximum likelihood estimator $\tilde{\beta}$ lead to $\sqrt{n}(\tilde{\beta} - \beta) = \sqrt{n} \tilde{\mathcal{I}}_{\beta\beta}^{-1} \tilde{S}_\beta + o_p(1)$, and so the usual first-order Taylor expansion of $(1/\sqrt{n}) \tilde{S}_\gamma$ at β yields

$$\frac{1}{\sqrt{n}} \tilde{S}_\gamma = \frac{1}{\sqrt{n}} S_\gamma + \frac{1}{\sqrt{n}} \tilde{\mathcal{H}}_{\gamma\beta} (\tilde{\beta} - \beta) + o_p(1)$$

$$= \frac{1}{\sqrt{n}} S_\gamma + \frac{1}{\sqrt{n}} \tilde{\mathcal{H}}_{\gamma\beta} \tilde{\mathcal{I}}_{\beta\beta}^{-1} \tilde{S}_\beta + o_p(1)$$

The classical derivation for the large-sample distribution of a score statistic for a composite null hypothesis usually suffice. In this case, however, the first-order Taylor expansion turns out to be inadequate for a genome-wide significance level at 5×10^{-8}. More importantly, $\tilde{\mathcal{H}}_{\gamma\beta}$ for this proportional interaction model is a function of the outcome variable Y,

$$\tilde{\mathcal{H}}_{\gamma\beta} = \sum_i \left\{ -\mathcal{X}_i \left(\sum_j \tilde{\beta}_{2j} G_i X_{ij} \right) \mu_i (1 - \mu_i) + \mathcal{V}_i (Y_i - \mu_i) \right\},$$

where $\mu_i = \exp(\mathcal{X}_i \tilde{\beta}) / \{1 + \exp(\mathcal{X}_i \tilde{\beta})\}$ and $\mathcal{V}_i = (0, 0, G_i \mathbf{X}_i, 0)$. This induces correlation between $\tilde{\mathcal{H}}_{\gamma\beta}$ and $\tilde{S}_\beta$ which, if ignored, will result in bias when calculating the mean and the variance of score statistics in finite samples, and consequently inflated type I errors. For example, the expectation of $\tilde{S}_\gamma$ is no longer 0 in finite samples. We show in Figure S1 a heavily inflated quantile-quantile plot for the standard score test for the proportional interactions. In Appendix A, we show the algebraic development of the finite-sample correction for the mean (denoted by τ_d) and the variance (denoted by Σ_d) of $\tilde{S}_\gamma$ after a second-order Taylor expansion,

$$\frac{1}{\sqrt{n}} \tilde{S}_\gamma = \frac{1}{\sqrt{n}} S_\gamma + \frac{1}{\sqrt{n}} \tilde{\mathcal{H}}_{\gamma\beta} (\tilde{\beta} - \beta) + \frac{1}{2\sqrt{n}} (\tilde{\beta} - \beta)^T \tilde{\mathcal{H}}_{\gamma\beta\beta} (\tilde{\beta} - \beta)$$
$$+ o_p \left(\frac{1}{\sqrt{n}} \right),$$

where $\tilde{\mathcal{H}}_{\gamma\beta\beta}$ is the second derivative of $\tilde{S}_\gamma$ with respect to β. The corrected score statistic for testing $\gamma = 0$ is therefore expressed as

$$T_d = \left(\tilde{S}_\gamma - \tau_d \right) \Sigma_d^{-1} \left(\tilde{S}_\gamma - \tau_d \right).$$

The first-order correction is simpler but turns out to be inadequate to correct for the inflated type I error (Figure S4). In our simulation study and application to the BEACON data shown next, the second-order correction expressed as T_d shows proper control of the type I error in genome-wide significance.

Table 1.   The Ratio of the Observed Type I Error Rate versus the Nominal Value for the Bonferroni Test, the Unconstrained Score Test, and the Constrained Score Tests

n Cases/n Controls	p Value Level	Bonferroni	Unconstrained	One-Sided	Bidirectional	Proportional Standard	Proportional Corrected
Independent Exposures							
500/500	0.05	0.99	1.01	1.1	1.02	1.17	0.88
	0.001	0.97	1.01	1.57	1.07	4.4	1.05
	0.0001	0.88	1.14	1.98	1.32	14.24	1.5
2,500/2,500	0.05	0.98	1.00	1.03	1	1.04	0.98
	0.001	0.94	1.05	1.25	1.07	1.5	1.04
	0.0001	0.98	1.04	1.34	1.10	2.68	1.40
5,000/5,000	0.05	0.98	0.99	1.02	0.99	1.01	0.99
	0.001	0.96	1	1.14	0.97	1.25	1
	0.0001	0.76	1.02	1.18	1.02	1.56	1
Correlated Exposures							
500/500	0.05	0.94	0.99	1.09	0.99	1.10	0.91
	0.001	1.03	1.01	1.65	1.1	3.08	1.18
	0.0001	1.34	1.06	2.12	1.34	9.62	2.06
2,500/2,500	0.05	0.98	1.01	1.04	1.01	1.03	0.99
	0.001	1.04	0.99	1.38	1.06	1.48	1.1
	0.0001	0.92	0.80	1.54	0.90	1.92	0.98
5,000/5,000	0.05	0.97	0.99	1.01	0.99	1.01	0.99
	0.001	0.99	1.02	1.21	0.94	1.17	0.99
	0.0001	1.12	0.94	1.34	1.16	1.38	1.04

Results

Simulation

We conducted numerous simulations to evaluate the performance of the proposed score statistics. Specifically, a genotype (coded as 0, 1, 2) was generated with minor allele frequency 0.1. Three exposure variables were generated by a multivariate normal distribution with zero means and a covariance matrix (either a diagonal matrix with diagonal elements 1 or a compound symmetric matrix with diagonal elements 1 and off-diagonal elements 0.6). The second and third exposure variables were dichotomized at zero. A logistic model was used to generate a dichotomous disease outcome conditional on the genotype and the exposures. The logarithm of the odds ratio was set to -0.5 for the genotype and (0.7, 0.7, 0.7) for the three exposures. The intercept was set to -4 to generate a disease probability of approximately 5%, under the null hypothesis that none of the exposures interacts with the genotype.

The two correlation structures between exposures generate different levels of correlation in the score functions for the three interaction parameters: a moderate level of correlations for each pair of score functions (<0.5) or a high correlation for each pair of score functions (>0.5). The correlation would affect the performance of the one-side score test and bidirectional test, because it is based on an approximation of the correlation-transformed orthant.[27]

To evaluate the type I error rate, a cohort was first generated by these probabilistic distributions with a ladder of sample sizes (10^4, 5×10^4, 10^5). All case subjects were sampled for simulated genotyping, together with the same number of randomly sampled control subjects. Table 1 shows empirical type I error rates evaluated in 5×10^5 datasets simulated under the null hypothesis, when the nominal p value cut-off was set to be 0.05, 0.001, or 0.0001. The number of case subjects varies from 500 to 5,000, with the same number of control subjects. The upper part shows the scenario where the three environmental exposures are independent and the lower part shows the scenario where they are correlated. In all scenarios, the Bonferroni correction for three separate tests and the classical score test for the unconstrained alternative H_{1a} shows adequate performance in controlling type I errors. The one-sided test for H_{1b} shows proper control of type I error rates except slight inflation for significance level 0.001 and 0.0001 under 500 case subjects. The increased correlation hurts the performance of the one-sided score statistic. The bidirectional score statistic for H_{1c} shows superior control of the type I error rate,
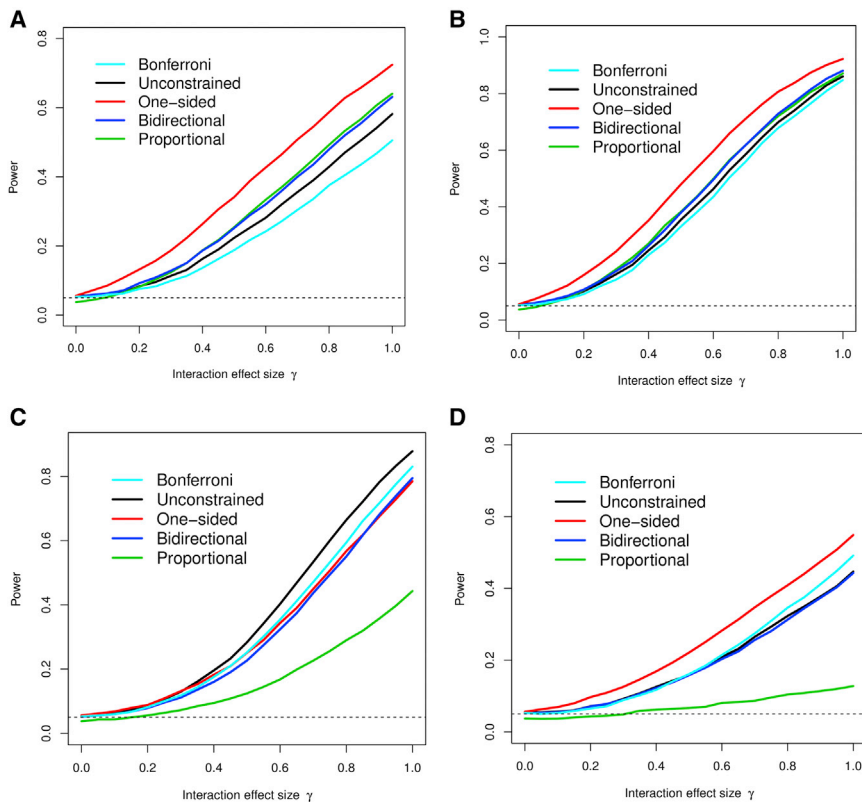
Figure 1. Statistical Power of the Four Methods Being Evaluated in Four Scenarios with a Small Sample Size

500 case subjects and 500 control subjects; there is no correlation between environmental exposures.
(A) The three interactions are proportional to respective main effects.
(B) The three interactions are not proportional, but one-sided.
(C) The three interactions have different directions.
(D) Two of the three interactions are zero.

because it uses a conservative approximation in the tail of the p value distribution. Under a relative smaller sample size, the naive score statistic for H_{1d} displays grossly inflated type I error rates when nominal significance is 0.001 or 0.0001, 4 times or 14 times more than the correct level when the exposures are independent. This inflation dissipates when sample size increases, suggesting that the problem is largely driven by erroneous small-sample properties, not the large-sample behavior. In comparison, the corrected score statistic for H_{1d} shows adequate control of type I error rates for all settings. In genome-wide testing for gene-environment interactions, the power is typically small for the genome-wide significance level (5×10^{-8}), and so the inflation of the type I error rate for the score statistic for H_{1d} can be concerning. See, for example, the quantile-quantile plot for the BEACON data in Figure S3. As we show in the data analysis later, the correction we proposed for H_{1d} is necessary to preserve adequate control of type I error.

In Figures 1 and 2 we show the power performance for the five methods, in which the Bonferroni correction and the 3-df score test are the benchmarks of the comparison. We consider two scenarios: 500 case and 500 control subjects, significance level 0.05 (Figure 1), and 2,500 case and 2,500 control subjects, significance level 0.0001 (Figure 2). In each scenario, we generate four different sets of interaction parameters: proportional, one-sided (negative), two-sided, and only one non-zero interaction. The latter two sets are designed to test the robustness of the constrained methods when respective modeling as-

sumptions are violated. We show the settings where the exposures are independent and omit the settings with correlated exposures, since the relative comparison between methods did not differ substantially from Figures 1 and 2.

Figure 1 shows power performance of various methods for 500 case subjects and 500 control subjects with the significance level 0.05. In Figure 1A, the interaction parameters were set to be truly proportional to the respective main effects, namely $\gamma\beta_{2j}$ with γ increasing from 0 to 1. Clearly, the Bonferroni method performs the worst among the five methods evaluated, losing 5%–10% power when compared to the unconstrained 3-df score test. The one-sided (negative) score statistic for H_{1b} shows the best performance in power, consistently higher over the unconstrained score test, with the maximal power gain reaching 15%–20%. The proportional score statistic with finite-sample correction shows the second best performance, slightly better than the bidirectional test. The advantage of the one-sided test over the proportional and the bidirectional test is expected, because the latter two examine differences in both positive and negative directions.

Figure 1B shows the setting in which the interactions are truly one-sided but not proportional to the main exposure effects, specifically, $\theta_j = \gamma(-0.75, -0.5, -0.25)$ with γ increasing from 0 to 1 and the main effects are (0.7, 0.7, 0.7). The order of the performance remains the same as that in Figure 1A, though the difference is smaller. The one-sided score test performs the best in this scenario, whereas the proportional interaction test and the bidirectional test appear to be next in power performance. The Bonferroni method yields nearly identical power as the unconstrained score test.

Figure 1C shows a scenario where the interactions do not have the same sign, nor are proportional to the main exposure effects. Specifically, $\theta_j = \gamma(-0.5, -0.75, -0.25)$ with δ increasing from 0 to 1. This is the setting where the modeling assumptions for the three constrained score tests are violated. Consequently, the unconstrained test and the
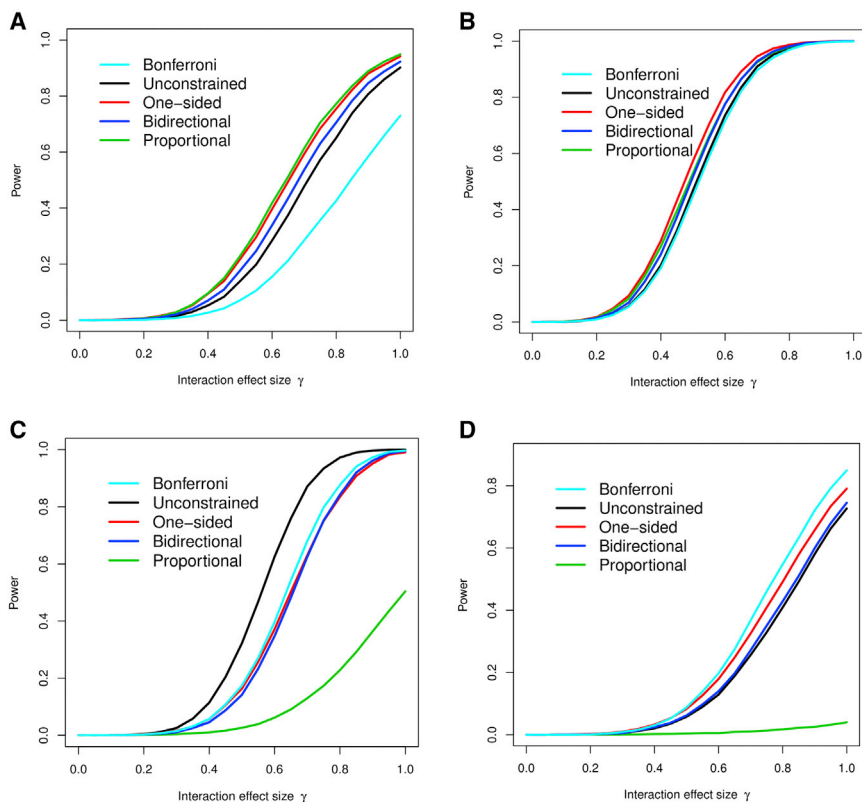
Figure 2. Statistical Power of the Four Methods Being Evaluated in Four Scenarios with a Large Sample Size

2,500 case subjects and 2,500 control subject; there is no correlation between environmental exposures.

(A) The three interactions are proportional to respective main effects.

(B) The three interactions are not proportional, but one-sided.

(C) The three interactions have different directions.

(D) Two of the three interactions are zero.

Bonferroni correction perform better than the constrained tests. The one-side (negative) test and the bidirectional test show some degree of robustness, with similarly lower power performance. The proportional test delivers a substantially inferior power performance relative to the other four tests.

In Figure 1D, we simulate a scenario where two out of the three interactions are 0, $\theta_j = \gamma(0, -1.0, 0)$ with δ increasing from 0 to 1. In this scenario, the Bonferroni method outperforms the unconstrained score test; the bidirectional test is similar to the constrained test; the proportional interaction test performs very poorly because this alternative hypothesis is not at all in the direction for which it is designated. Interestingly, the one-sided test still delivers the best power performance, because these interaction parameter values still conform to the alternative hypothesis H_{1b} that the interactions lie in the negative orthant.

Figure 2 shows the settings with 2,500 case and 2,500 control subjects, p value cut-off 0.0001. We want to examine the performance of the proposed methods in a large-scale case-control study, like the one in BEACON, and there are a large number of genotypes. The similar pattern as Figure 1 was observed in power comparison: when the alternative hypothesis is directional or proportional, the three constrained score statistics performs better than the standard 3-df test and the Bonferroni correction; notably, the proportional interaction outperforms the one-sided test, suggesting that pooling three exposures into a risk score for interaction really boosts the power when the significance level is far out in the

tail; the one-sided test and the bidirectional test can be quite robust against model violation, whereas the performance of the proportional test drops substantially when the alternative hypothesis are not in the same direction. We also note that another feature affecting the performance of the proportional test is the effect size of the main exposure effects (Figure S5). Strong exposure effects and the large sample size, such as the three exposures in the BEACON study, works best for the proportional test, because its performance depends on the variability of the main effect estimates. Based on these simulations, it is compelling to apply the three constrained score tests and the unconstrained score test in the BEACON study.

Data Application

All 1,516 esophageal adenocarcinoma case individuals and 2,416 Barrett's esophagus individuals in the discovery phase of the BEACON GWAS were included in this investigation, together with 2,187 control participants. A description of the genome-wide study has been published previously.[23] All recruited participants gave informed consent, and this study was approved by the ethics boards of each participating institution. Genetic data for which the authors have IRB permission to make public are available from the dbGaP database (accession numbers phs000869.v1.p1 and phs000187.v1.p1). Three established risk factors were investigated for potential interaction with 922,031 SNPs that were genotyped by the Illumina HumanOmni1-Quad platform and passed quality control: cigarette smoking, body-mass index (BMI), and gastresophageal reflux symptoms. These variables were coded as ever smoking (yes or no), BMI (<25, ≥ 25 to <30, ≥ 30 kg/m^2), and at least weekly heartburn or weekly reflux (yes or no). A logistic model was fitted for every SNP and the three exposures, separately for EA and BE. Each model included age, sex, the first four principal components derived from genome-wide SNP data to account for population stratification, the SNP main effect (coded as two indicator variables), the three exposure
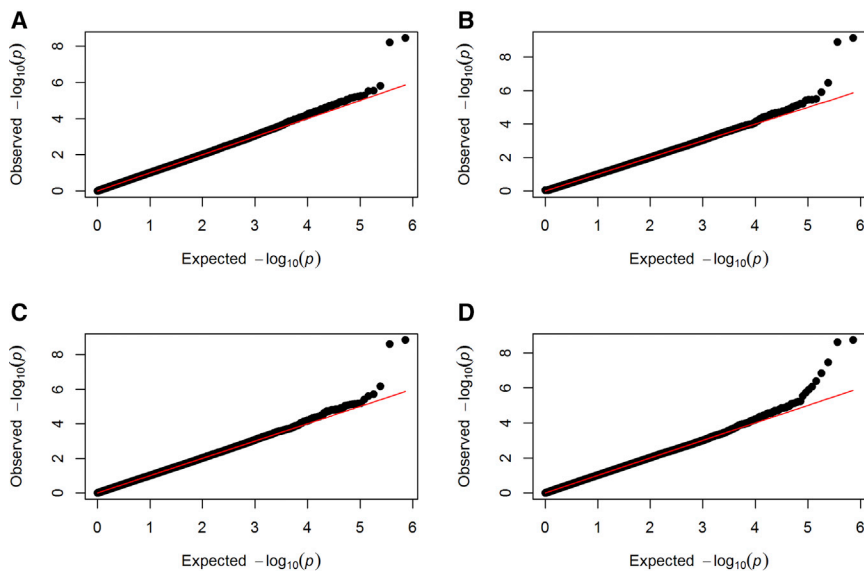
Figure 3. The q-q Plots for Genome-wide Testing for G × E on Barrett's Esophagus by the Three Score Statistics
(A) Unconstrained score test.
(B) One-sided score test for negative G × E.
(C) Bidirectional test.
(D) Score test for proportional G × E.

main effects (all coded categorically), and the product of the SNP value (0, 1, or 2, treated quantitatively) and each of the three exposures. The multivariate SNP-exposure interactions were tested by the unconstrained score test, the one-sided test, and the proportional test.

Figure 3 shows the q-q plots for the four genome-wide G × E testing methods for Barrett's esophagus. The one-sided results for the positive interaction for Barrett's esophagus, as well as all G × E tests for esophageal adenocarcinoma, did not yield any significant results and therefore were not presented. Figure 3A shows that the unconstrained 3-df score statistics yield two genome-wide significant SNPs on chromosome 7 that are highly correlated ($r^2 = 1$; rs11765529, p value 3.48×10^{-9}; rs7798662, p value 6.06×10^{-9}). The same two SNPs remain genome-wide significant when tested by one-sided (negative) score statistics, the bidirectional test, and the proportional interaction test shown in Figures 3B–3D. The proportional interaction test resulted in one additional genome-wide significant SNP, rs4930068 in chromosome 11 (p value 3.51×10^{-8}, Figure 3D). In Figure S3, we show the q-q plot with seriously inflated type I error rates when the finite-sample correction is not applied to score statistics for proportional interactions.

Table 2 shows the estimated odds ratios of the three environmental factors stratified by the number of alleles for the two genome-wide significant SNPs. Barrett's esophagus and esophageal adenocarcinoma were investigated as cases separately in logistic regression models with three SNP-environment interactions. The interpretation of these ORs is conditional on genotype, age, gender, and the other two exposures being fixed at certain levels. The right half of Table 2 shows the p values from various score tests for G × E testing. For almost every combination of SNP, exposure, and disease combination, the OR decreases with the number of increasing-risk alleles. This decreasing OR with the number of alleles is most strongly seen in

GERD, the strongest risk factor, and the trend we found for Barrett's esophagus is similar to that for esophageal adenocarcinoma, though the latter does not have a genome-wide significance level. Since both the cases of Barrett's esophagus and the cases of esophageal adenocarcinoma were compared to the same control group, the findings for esophageal adenocarcinoma are not independent replications. Nonetheless, the consistency between Barrett's esophagus and esophageal adenocarcinoma adds to the evidence for true G × E in the etiological pathway.

Association analyses were further conducted for imputed SNPs based on the 1000 Genomes Project around genotyped SNPs that showed significant interactions with one of the three risk factors. The details of the imputation procedure have been presented previously.[23] Figure 4 shows the regional interaction associations around rs11765529 and rs4930068, including both genotyped and imputed SNPs within a 1 Mb interval centering around the index SNPs. The index SNPs are marked as purple squares, other genotyped SNPs are marked with squares, and imputed SNPs are marked by circles. The rs11765529 variant is located in a gene-poor intergenic region, ~160–180 kb from the nearest protein-coding gene, POM121L12 (transmembrane nucleoporin-like 12 [MIM: 615753]). The rs4930068 variant at 11p15.5 is located 5.4 kb upstream (5′) of the transcriptional start site for ASCL2 (MIM: 601886). None of adjacent SNPs showed more significant p values than the index SNPs.

We conducted in silico functional characterization of these SNPs in several bioinformatics databases. The two variants at 7p12.1 (rs7798662 and rs11765529) are situated in a gene-poor intergenic region, ~160–180 kb from the nearest protein-coding gene, POM121L12 (Figure S6A). These two SNPs are separated by ~24 kb, are in strong linkage disequilibrium (LD) ($r^2 = 1$), and are highly correlated ($r^2 = 1$) with >100 additional variants in a 40-kb region, based on data from the 1000 Genomes Project[34] (Table S1A). rs7798662 modifies a predicted binding motif for the transcription factor Smad3 (MIM: 603109), whereas rs11765529 alters motifs for HDAC2 (MIM: 605164).[35] Several correlated SNPs lie in regions marked by enhancer histone modifications or DNase hypersensitivity in multiple tissues or cell lines, based on data from the NIH Roadmap Epigenome Project and ENCODE.[36,37] Recruitment of the transcription factor JUND (MIM: 165162) has been

Table 2. The Odds Ratio of the Three Environmental Factors Associated with SNP Genotypes for Barrett's Esophagus and Esophageal Adenocarcinoma, and the Three Score Tests for Multiple G × E Simultaneously

		OR Stratified by the Number of Minor Alleles			G × E p Value			
Disease	Risk Factor	0	1	2	Unconstrained	One-Sided	Bidirectional	Proportional
rs11765529								
BE	BMI	1.94 (1.74, 2.17)	1.64 (1.24, 2.16)	1.28 (0.08, 19.7)	–	–	–	–
	smoking	1.67 (1.41, 1.98)	0.79 (0.51, 1.21)	0.94 (0.80, 11.1)	–	–	–	–
	GERD	6.39 (5.35, 7.64)	2.09 (1.36, 3.21)	0.69 (0.06, 8.09)	3.48×10^{-9}	7.10×10^{-10}	1.42×10^{-9}	1.81×10^{-9}
EA	BMI	1.51 (1.31, 1.74)	1.17 (0.81, 1.69)	–	–	–	–	–
	smoking	1.87 (1.50, 2.35)	1.20 (0.64, 2.27)	–	–	–	–	–
	GERD	3.83 (3.07, 4.77)	2.34 (1.33, 4.11)	–	0.013	0.004	0.0072	0.003
rs4930068								
BE	BMI	2.11 (1.80, 2.49)	1.91 (1.64, 2.22)	1.32 (1.01, 1.72)	–	–	–	–
	smoking	1.56 (1.22, 1.99)	1.73 (1.38, 2.18)	0.85 (0.56, 1.28)	–	–	–	–
	GERD	7.91 (6.08, 10.30)	4.75 (3.76, 6.02)	2.86 (1.88, 4.35)	1.53×10^{-6}	3.38×10^{-7}	6.76×10^{-7}	3.51×10^{-8}
EA	BMI	1.46 (1.19, 1.79)	1.43 (1.17, 1.74)	1.41 (0.99, 1.99)	–	–	–	–
	smoking	2.14 (1.54, 2.99)	1.67 (1.23, 2.27)	1.08 (0.60, 1.95)	–	–	–	–
	GERD	4.80 (3.48, 6.60)	3.04 (2.26, 4.08)	2.21 (1.24, 3.92)	0.007	0.002	0.004	0.004

Abbreviations are as follows: BE, Barrett's esophagus; EA, esophageal adenocarcinoma.

reported ~6.5 kb away from *rs11765529*. Of potential interest, rs7798662 and rs11765529 lie within a 3.3 Mb region designated as a long-range epigenetic activation (LREA) domain in prostate cancer cells; this region, which encompasses a total of ten genes (including *POM121L12*) was identified as transcriptionally active in prostate cancer cell lines relative to normal primary prostate cells.[38] The POM121L12 protein is largely uncharacterized with respect to biological function, but is related to the POM121 transmembrane nucleoporin, a key component of the nuclear pore complex, which enables transport into and out of the nucleus.[39,40]

The rs4930068 variant at 11p15.5 is located 5.4 kb upstream (5′) of the transcriptional start site for *ASCL2*, which encodes the Achaete-Scute Family bHLH Transcription Factor 2 (Figure S6B). rs4930068 is in strong LD ($r^2 > 0.8$) with 41 additional variants located within 20 kb, including 30 SNPs within 8 kb ($r^2 > 0.90$).[34] rs4930068 lies in a 7,400-bp region characterized as heterochromatin in esophageal tissue, according to chromatin state segmentation data derived from the Roadmap Epigenome Project.[37] Multiple SNPs strongly correlated with rs4930068, rs11021967, rs7396048, rs2285567, rs11022026, and rs6578259 are located within segments marked by enhancer histone marks in the esophagus and/or distal gastrointestinal (GI) mucosa and appear to hold high regulatory potential (Table S1). Several of these variants lie within regions characterized by DNase hypersensitivity and transcription factor recruitment (*TAL1* [MIM: 187040], *USF1* [MIM: 191523], *E2F6* [MIM: 602944], *ELF1* [MIM: 189973], *MAX* [MIM: 154950]) in various cells/tissues and modify predicted regulatory binding motifs.[36] Some of these SNPs (rs7396048, rs2285567, rs11022026) are potential expression quantitative trait loci (eQTLs) in whole blood or other tissues.[36]

The ASCL2 protein is a member of the basic helix-loop-helix (HLH) family of transcription factors, which dimerize and bind to DNA via their HLH and basic domains, respectively.[40] bHLH proteins play important general roles in cell fate specification and lineage-specific transcriptional control. ASCL1 (MIM: 100790) and ASCL2 are essential factors in the development of the neuroectoderm and trophectoderm, respectively,[41,42] and both proteins are frequently expressed in human cancers. ASCL2, an apparent target gene of the Wnt signaling pathway, was shown to be upregulated in colorectal neoplasia[43] and has been linked to an intestinal stem cell expression signature that defines a majority of colorectal cancers (MIM: 114500)[44] and may be present in precancerous adenomas.[45] Low-level expression of ASCL2 mRNA has also recently been reported in BE[46] while studies in human esophageal cell lines have described induction of ASCL2 mRNA and protein in a sub-population of stem-like spheroid cells derived in culture.[47] Of particular interest, ASCL2 was shown to suppress expression of the homeobox transcription factor CDX2 (MIM: 600297),[48] a known early marker of BE,[49] subsequently downregulated during EA pathogenesis.[50]

Based on our in silico functional characterization of rs4930068 and correlated variants, a potential regulatory role for these SNPs in influencing ASCL2 expression levels appears plausible. However, future laboratory-based studies are clearly required to assess possible functional effects. Recent evidence has suggested that Barrett's esophagus may represent the outgrowth of a residual embryonic
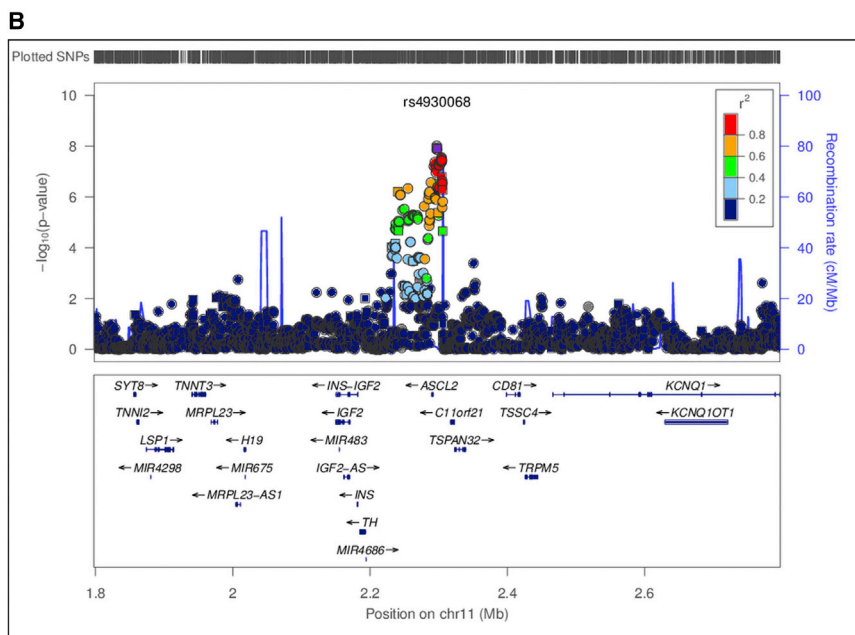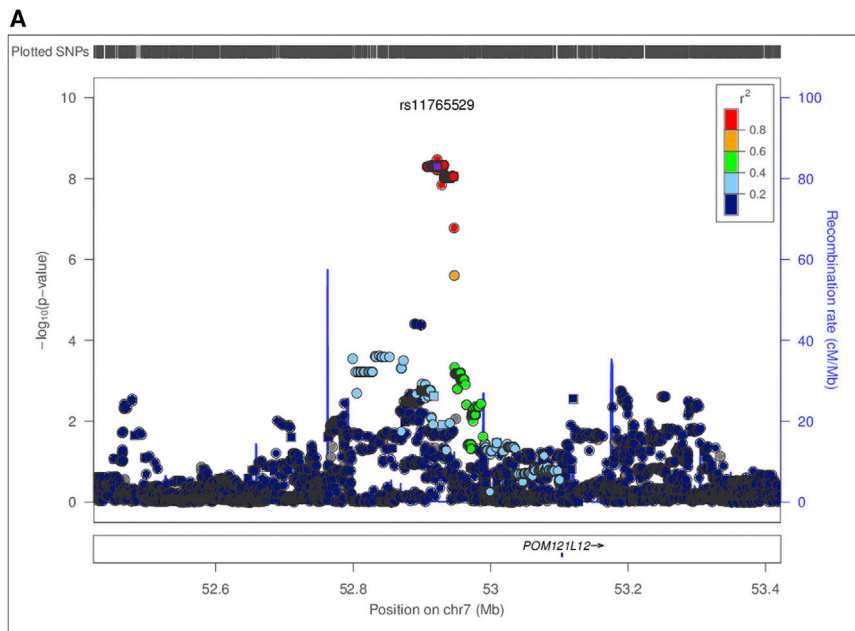
A

B

multiple related environmental exposures: the one-sided interactions, the bidirectional interactions, and the proportional interactions (on the log OR scale relative to respective main environmental associations). Several statistical contributions were made in this work. First, an approximate score statistic was developed for the one-sided test, which circumvents the difficulty of obtaining an exact asymptotic distribution, and further extended to the bidirectional test. Second, for the proportional interaction test, we proposed a second-order correction to the mean and the variance of the score statistic, which resolved the issue of the inflated type I error rate in genome-wide testing. Our proposed methods were applied to the BEACON study and identified several loci that interact with three environmental exposures in Barrett's esophagus with genome-wide significance.

Our work is exploratory in nature, as other methodological developments in genome-wide association testing. There is a paucity of established G × E, much due to lower power of such exploration. The goal of our work is to enrich G × E findings that may lead to further validation and functional studies. These G × E could help in risk prediction and stratification, or possibly individualized prevention for modifiable exposures.

Our simulation study suggests that the one-side test nearly always outperforms, if only slightly, the proportional interaction test, even in the scenario where interactions are generated to be proportional. This is primarily

epithelial cell population, rather than squamous-to-columnar trans-differentiation of resident stratified epithelium.[51,52] In this context, it is intriguing to consider the possibility that the observed interaction between rs4930068 and GERD in relation to risk of BE might in part reflect alterations in the expression level of ASCL2, a transcription factor with known roles in both development and cancer.

Discussion

In summary, we have proposed constrained testing strategies for identifying interactions between a genotype and

due to the proportional test being a two-sided test, and its increased variability because of its dependence on the main associations of environmental exposures. When the latter were estimated with quite some uncertainty, for example, when the sample size is small, the proportional test is not very appealing in power performance. However, in consortium studies where investigators can assemble a large sample size, the proportional test can be advantageous when the interactions are truly directional (Figures 2A and 2B). When such constraint is violated, the proportional test can be powerless as shown in Figures 1C, 1D, 2C, and 2D. The directional test strikes the balance between power and robustness, performing competitively under true model constraints and displaying robustness against model violations.

In cancer epidemiology, if several exposures share some etiological pathways, our simulation shows that testing one exposure at a time is inferior in power to simultaneously testing multiple G × E altogether. We show an example of applying such testing strategies in the BEACON study and we found several interesting genome-wide significant hits. We note that the suitability of the proposed constrained inference depends heavily on study context and a priori scientific knowledge. Other plausible settings include multiple measures of the same underlying exposure profile, such as measures of smoking and diet intake. Investigators should exercise caution when applying the constrained inference since, as we show in the simulation, violations of model constraints can result in deteriorated power.

Beyond the multivariate G × E testing we consider, constrained inference strategies for a single G × E test have been explored. For example, it is often reasonable to constrain the interaction to be quantitative, because epidemiologists believe such interaction is generally unnatural and rare, if it exists at all.[53] Some works have been developed along this line, possibly incorporating the powerful but controversial gene-environment independence assumption.[54,55] From a methodological point of view, the latter assumption could be incorporated into the constrained score statistics we developed in this article, with some additional work. For the BEACON study, however, this assumption does not hold because BMI and GERD are both phenotypes with substantial genetic susceptibility.[56,57]

We close with a brief discussion on generalizing the proposed methods. The Tukey's test has been formulated to study interactions between multiple genes and multiple environmental exposures. In principle, the one-sided and bidirectional testing strategies can be applied equally to one exposure of interest and a set of SNPs in the gene, or a set of known risk loci. The proportional interaction test naturally applies to this setting. One complexity for the one-sided test is that, in contrast to the BEACON study where all three exposures are known to be hazardous, it may be entirely unknown which alleles of the SNPs in a gene may modify the exposure association in the same di-

rection. Therefore it requires a bit more understanding of the function of the alleles prior to motivating the directional hypotheses for one exposure of interest and a set of SNPs in the gene. The other potential issue for one exposure and a set of SNPs is that the number of SNPs has to be limited, because the dimensionality would affect the degree of freedom in the chi-bar-squared distribution, which will in turn influence power performance.

Appendix A

Construction of the Transformation Matrix A

As stipulated in Material and Methods, $A^T A = \mathcal{I}_{\theta\theta,\beta}$ and its center direction should coincide with the center direction of the negative orthant. In addition, it would be appealing that $A(\delta)$ is invariant to the order of the elements in U, since the order of the interaction terms tested in the alternative hypothesis should not change the test statistic. We adopt a linear ordering method previously developed to rearrange U and $\mathcal{I}_{\theta\theta,\beta}$.[58] Next we find A that satisfies Equations 3 and 4. Let $H = (h_1, \ldots, h_p)$ be the orthogonal matrix obtained by applying the Gram-Schmidt orthonormalization to the independent set of vectors $(J, e_1, \ldots, e_p)$. Let $C^T C = \mathcal{I}_{\theta\theta,\beta}$ be the Choleski decomposition of $\mathcal{I}_{\theta\theta,\beta}$. Let d_c denote the center direction of the image space $C(\delta) = \{C\delta \mid \delta \leq 0\}$, and let $V = (v_1, \ldots, v_p)$ be the orthogonal matrix obtained by applying Gram-Schmidt orthonormalization to the independent set of vectors $(d_c, e_2, \ldots, e_k)$. Then $A = HV^T C$ satisfies the conditions of Equations 3 and 4.

Computation of Tail p Values for the Bidirectional Hypothesis by Algebraic Approximation

It is instructive to show the approximation when $p = 2$. The calculation in a higher dimension follows without difficulty. Specifically, let (z_1, z_2) denote the standard bivariate normal variables. The four quadrants are denoted by $\mathcal{Q}_1$, $\mathcal{Q}_2$, $\mathcal{Q}_3$, and $\mathcal{Q}_4$, respectively. The p value for the observed $T_c = t$ is the sum of components from each quadrant. The components of p value in $\mathcal{Q}_1$ and $\mathcal{Q}_3$ follow a standard χ_2^2 distribution. For components in $\mathcal{Q}_2$ and $\mathcal{Q}_4$,

$$
\begin{aligned}
\Pr(T_c > t \mid (z_1, z_2) &\in \mathcal{Q}_2 \cup \mathcal{Q}_4) = \Pr(z_2^2 > t, \mid z_2 \mid \\
&\geq \mid z_1 \mid \mid (z_1, z_2) \in \mathcal{Q}_2 \cup \mathcal{Q}_4) \\
&+ \Pr(z_1^2 > t, \mid z_1 \mid > \mid z_2 \mid \mid (z_1, z_2) \in \mathcal{Q}_2 \cup \mathcal{Q}_4) \\
&< \Pr(z_2^2 > t \mid (z_1, z_2) \in \mathcal{Q}_2 \cup \mathcal{Q}_4) \\
&+ \Pr(z_1^2 > t \mid (z_1, z_2) \in \mathcal{Q}_2 \cup \mathcal{Q}_4) \\
&= \Pr(\chi_1^2 > t) + \Pr(\chi_1^2 > t)
\end{aligned}
$$

It is clear that the bigger t, the better approximation. In Figures S1 and S2, we show a comparison of simulation-based p values and approximation-based p values. For p values < 0.01, the difference is essentially negligible.

Correcting the Mean and the Variance of the Score Statistic for the Proportional Interaction

For dichotomous outcomes, the exact computation of the mean of the second-order Taylor expansion of $\tilde{S}_\gamma$, namely

$$S_\gamma + \tilde{\mathcal{H}}_{\gamma\beta}\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{S}_\beta + \frac{1}{2}\tilde{S}_\beta^T\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{\mathcal{H}}_{\gamma\beta\beta}\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{S}_\beta,$$

can be derived algebraically, since

$$
\begin{aligned}
\tau_d &= \mathbb{E}\left(S_\gamma + \tilde{\mathcal{H}}_{\gamma\beta}\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{S}_\beta + \frac{1}{2}\tilde{S}_\beta^T\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{\mathcal{H}}_{\gamma\beta\beta}\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{S}_\beta\right) \\
&= \mathbb{E}\left[\left(\sum_i \tilde{\mathcal{K}}_{\gamma\beta,i}\right)\tilde{\mathcal{I}}_{\beta\beta}^{-1}\left(\sum_i \tilde{S}_{\beta,i}\right)\right. \\
&\quad \left. +\frac{1}{2}\left(\sum_i \tilde{S}_{\beta,i}\right)^T \tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{\mathcal{H}}_{\gamma\beta\beta}\tilde{\mathcal{I}}_{\beta\beta}^{-1}\left(\sum_i \tilde{S}_{\beta,i}\right)\right] \\
&= \sum_i \mathbb{E}\left(\tilde{\mathcal{K}}_{\gamma\beta,i}\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{S}_{\beta,i} + \frac{1}{2}\tilde{S}_{\beta,i}^T\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{\mathcal{H}}_{\gamma\beta\beta}\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{S}_{\beta,i}\right) \\
&= \sum_i \mu_i\left(\tilde{\mathcal{K}}_{\gamma\beta,i1}\tilde{\mathcal{I}}_{\beta\beta}\tilde{S}_{\beta,i1} + \frac{1}{2}\tilde{S}_{\beta,i1}^T\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{\mathcal{H}}_{\gamma\beta\beta}\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{S}_{\beta,i1}\right) \\
&\quad +(1-\mu_i)\left(\tilde{\mathcal{K}}_{\gamma\beta,i0}\tilde{\mathcal{I}}_{\beta\beta}\tilde{S}_{\beta,i0} + \frac{1}{2}\tilde{S}_{\beta,i0}^T\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{\mathcal{H}}_{\gamma\beta\beta}\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{S}_{\beta,i0}\right),
\end{aligned}
$$

where $\tilde{\mathcal{K}}_{\gamma\beta,i} = \mathcal{V}_i(Y_i - \mu_i)$ and $\tilde{S}_{\beta,i}$ are respective individual contribution from the ith subject, $\tilde{\mathcal{K}}_{\gamma\beta,i1}$ and $\tilde{S}_{\beta,i1}$ are contributions from the ith subject, evaluated at $Y_i = 1$, and $\tilde{\mathcal{K}}_{\gamma\beta,i0}$ and $\tilde{S}_{\beta,i0}$ are contributions from the ith subject, evaluated at $Y_i = 0$. The second-to-last equality uses the independence between the ith subject and the jth subject, if $i \neq j$. The last equality uses the unique feature that Y_i takes two values (1,0) with probability μ_i and $(1 - \mu_i)$, respectively. Similar, but more tedious, calculation applies to

$$\Sigma_d = \text{Var}\left(S_\gamma + \tilde{\mathcal{H}}_{\gamma\beta}\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{S}_\beta + \frac{1}{2}\tilde{S}_\beta^T\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{\mathcal{H}}_{\gamma\beta\beta}\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{S}_\beta\right).$$

Let $\Delta = \sum_i - \mathcal{X}_i(\sum_j \tilde{\beta}_{2j}G_iX_{ij})\mu_i(1 - \mu_i)$. The key terms involved in the calculation are expressed below:

$$
\mathbb{E}\left(S_\gamma + \tilde{\mathcal{H}}_{\gamma\beta}\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{S}_\beta\right)^2 = \sum_i \mathbb{E}\left(S_{\gamma,i} + (\Delta + \tilde{\mathcal{K}}_{\gamma\beta,i})\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{S}_{\beta,i}\right)^2 \\
+ \sum_{i \neq j}\mathbb{E}\left(\tilde{\mathcal{K}}_{\gamma\beta,i}\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{S}_{\beta,j}\right)^2
$$

$$
\mathbb{E}\left(\tilde{S}_\beta^T\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{\mathcal{H}}_{\gamma\beta\beta}\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{S}_\beta\right)^2 = \sum_i \sum_j \mathbb{E}\left(\tilde{S}_{\beta,i}^T\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{\mathcal{H}}_{\gamma\beta\beta}\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{S}_{\beta,i}\right)^2
$$

$$
\mathbb{E}\left(S_\gamma\tilde{S}_\beta^T\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{\mathcal{H}}_{\gamma\beta\beta}\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{S}_\beta\right) = \sum_i \left(S_{\gamma,i}\tilde{S}_{\beta,i}^T\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{\mathcal{H}}_{\gamma\beta\beta}\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{S}_{\beta,i}\right)
$$

$$
\mathbb{E}\left(S_\gamma\tilde{S}_\beta^T\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{\mathcal{H}}_{\gamma\beta\beta}\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{S}_\beta\right) = \sum_i \left(S_{\gamma,i}\tilde{S}_{\beta,i}^T\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{\mathcal{H}}_{\gamma\beta\beta}\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{S}_{\beta,i}\right)
$$

$$
\mathbb{E}\left(\left(\tilde{\mathcal{H}}_{\gamma\beta}\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{S}_\beta\right)\left(\tilde{S}_\beta^T\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{\mathcal{H}}_{\gamma\beta\beta}\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{S}_\beta\right)\right) = \sum_i \mathbb{E}\left(\left(\Delta\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{S}_{\beta,i}\right)\right. \\
\times \left(\tilde{S}_{\beta,i}^T\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{\mathcal{H}}_{\gamma\beta\beta}\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{S}_{\beta,i}\right)\right) + \sum_i \sum_j \mathbb{E}\left(\left(\tilde{\mathcal{K}}_{\gamma\beta,i}\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{S}_{\beta,j}\right)\right. \\
\left. \times \left(\tilde{S}_{\beta,i}^T\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{\mathcal{H}}_{\gamma\beta\beta}\tilde{\mathcal{I}}_{\beta\beta}^{-1}\tilde{S}_{\beta,j}\right)\right).
$$

Web Resources

dbGaP, http://www.ncbi.nlm.nih.gov/gap
OMIM, http://www.omim.org/

References

1. Brennan, P. (2002). Gene-environment interaction and aetiology of cancer: what does it mean and how can we measure it? Carcinogenesis *23*, 381–387.
2. Hunter, D.J. (2005). Gene-environment interactions in human diseases. Nat. Rev. Genet. *6*, 287–298.
3. Thomas, D. (2010). Gene–environment-wide association studies: emerging approaches. Nat. Rev. Genet. *11*, 259–272.
4. Wu, C., Kraft, P., Zhai, K., Chang, J., Wang, Z., Li, Y., Hu, Z., He, Z., Jia, W., Abnet, C.C., et al. (2012). Genome-wide association analyses of esophageal squamous cell carcinoma in Chinese identify multiple susceptibility loci and gene-environment interactions. Nat. Genet. *44*, 1090–1097.
5. Garcia-Closas, M., Rothman, N., Figueroa, J.D., Prokunina-Olsson, L., Han, S.S., Baris, D., Jacobs, E.J., Malats, N., De Vivo, I., Albanes, D., et al. (2013). Common genetic polymorphisms modify the effect of smoking on absolute risk of bladder cancer. Cancer Res. *73*, 2211–2220.
6. Prentice, R.L. (2011). Empirical evaluation of gene and environment interactions: methods and potential. J. Natl. Cancer Inst. *103*, 1209–1210.
7. Hutter, C.M., Mechanic, L.E., Chatterjee, N., Kraft, P., and Gillanders, E.M.; NCI Gene-Environment Think Tank (2013). Gene-environment interactions in cancer epidemiology: a National Cancer Institute Think Tank report. Genet. Epidemiol. *37*, 643–657.
8. Chatterjee, N., and Carroll, R.J. (2005). Semiparametric maximum likelihood estimation exploiting gene-environment indepdendence in case-control studies. Biometrika *92*, 399–418.

9. Mukherjee, B., and Chatterjee, N. (2008). Exploiting gene-environment independence for analysis of case-control studies: an empirical Bayes-type shrinkage estimator to trade off between bias and efficiency. Biometrics 64, 685–694.

10. Dai, J.Y., Kooperberg, C., Leblanc, M., and Prentice, R.L. (2012). Two-stage testing procedures with independent filtering for genome-wide gene-environment interaction. Biometrika 99, 929–944.

11. Hsu, L., Jiao, S., Dai, J.Y., Hutter, C., Peters, U., and Kooperberg, C. (2012). Powerful cocktail methods for detecting genome-wide gene-environment interaction. Genet. Epidemiol. 36, 183–194.

12. Kooperberg, C., Dai, J.Y., and Hsu, L. (2015). Methodological and Statistical Issues in Gene-Environment and Gene-Gene Interactions for Complex Phenotypes (Cambridge: MIT Press).

13. VanderWeele, T.J. (2009). Sufficient cause interactions and statistical interactions. Epidemiology 20, 6–13.

14. Campa, D., Kaaks, R., Le Marchand, L., Haiman, C.A., Travis, R.C., Berg, C.D., Buring, J.E., Chanock, S.J., Diver, W.R., Dostal, L., et al. (2011). Interactions between genetic variants and breast cancer risk factors in the breast and prostate cancer cohort consortium. J. Natl. Cancer Inst. 103, 1252–1263.

15. Hutter, C.M., Chang-Claude, J., Slattery, M.L., Pflugeisen, B.M., Lin, Y., Duggan, D., Nan, H., Lemire, M., Rangrej, J., Figueiredo, J.C., et al. (2012). Characterization of gene-environment interactions for colorectal cancer susceptibility loci. Cancer Res. 72, 2036–2044.

16. Blot, W.J., Devesa, S.S., Kneller, R.W., and Fraumeni, J.F.J., Jr. (1991). Rising incidence of adenocarcinoma of the esophagus and gastric cardia. JAMA 265, 1287–1289.

17. Brown, L.M., Devesa, S.S., and Chow, W.H. (2008). Incidence of adenocarcinoma of the esophagus among white Americans by sex, stage, and age. J. Natl. Cancer Inst. 100, 1184–1187.

18. Thrift, A.P., and Whiteman, D.C. (2012). The incidence of esophageal adenocarcinoma continues to rise: analysis of period and birth cohort effects on recent trends. Ann. Oncol. 23, 3155–3162.

19. Vaughan, T.L., and Fitzgerald, R.C. (2015). Precision prevention of oesophageal adenocarcinoma. Nat. Rev. Gastroenterol. Hepatol. 12, 243–248.

20. Reid, B.J., Li, X., Galipeau, P.C., and Vaughan, T.L. (2010). Barrett's oesophagus and oesophageal adenocarcinoma: time for a new synthesis. Nat. Rev. Cancer 10, 87–101.

21. Engel, L.S., Chow, W.H., Vaughan, T.L., Gammon, M.D., Risch, H.A., Stanford, J.L., Schoenberg, J.B., Mayne, S.T., Dubrow, R., Rotterdam, H., et al. (2003). Population attributable risks of esophageal and gastric cancers. J. Natl. Cancer Inst. 95, 1404–1413.

22. Olsen, C.M., Pandeya, N., Green, A.C., Webb, P.M., and Whiteman, D.C.; Australian Cancer Study (2011). Population attributable fractions of adenocarcinoma of the esophagus and gastroesophageal junction. Am. J. Epidemiol. 174, 582–590.

23. Levine, D.M., Ek, W.E., Zhang, R., Liu, X., Onstad, L., Sather, C., Lao-Sirieix, P., Gammon, M.D., Corley, D.A., Shaheen, N.J., et al. (2013). A genome-wide association study identifies new susceptibility loci for esophageal adenocarcinoma and Barrett's esophagus. Nat. Genet. 45, 1487–1493.

24. Dai, J.Y., de Dieu Tapsoba, J., Buas, M.F., Onstad, L.E., Levine, D.M., Risch, H.A., Chow, W.H., Bernstein, L., Ye, W., Lagergren, J., et al. (2015). A newly identified susceptibility locus near FOXP1 modifies the association of gastroesophageal reflux with Barrett's esophagus. Cancer Epidemiol. Biomarkers Prev. 24, 1739–1747.

25. Corley, D.A., Kubo, A., and Zhao, W. (2007). Abdominal obesity, ethnicity and gastro-oesophageal reflux symptoms. Gut 56, 756–762.

26. Pocock, S.J., Geller, N.L., and Tsiatis, A.A. (1987). The analysis of multiple endpoints in clinical trials. Biometrics 43, 487–498.

27. Tang, D.I., Gnecco, C., and Geller, N.L. (1989). An approximate likelihood ratio test for a normal mean vector with nonnegative components with application to clinical trials. Biometrika 76, 577–583.

28. Tukey, J.W. (1949). One degree of freedom for non-additivity. Biometrics 5, 232–242.

29. Mandel, J. (1961). Non-additivity in two-way analysis of variance. J. Am. Stat. Assoc. 56, 878–888.

30. Chatterjee, N., Kalaylioglu, Z., Moslehi, R., Peters, U., and Wacholder, S. (2006). Powerful multilocus tests of genetic association in the presence of gene-gene and gene-environment interactions. Am. J. Hum. Genet. 79, 1002–1016.

31. Kudo, A. (1963). A multivariate analogue of the one-sided test. Biometrika 50, 403–418.

32. Shapiro, A. (1985). Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. Biometrika 72, 133–144.

33. Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. PLoS Genet. 9, e1003348.

34. Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A.; 1000 Genomes Project Consortium (2012). An integrated map of genetic variation from 1,092 human genomes. Nature 491, 56–65.

35. Ward, L.D., and Kellis, M. (2012). HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 40, D930–D934.

36. Bernstein, B.E., Stamatoyannopoulos, J.A., Costello, J.F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M.A., Beaudet, A.L., Ecker, J.R., et al. (2010). The nih roadmap epigenomics mapping consortium. Nat. Biotechnol. 28, 1045–1048.

37. Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C., Doyle, F., et al.; ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74.

38. Bert, S.A., Robinson, M.D., Strbenac, D., Statham, A.L., Song, J.Z., Hulf, T., Sutherland, R.L., Coolen, M.W., Stirzaker, C., and Clark, S.J. (2013). Regional activation of the cancer genome by long-range epigenetic remodeling. Cancer Cell 23, 9–22.

39. Hallberg, E., Wozniak, R.W., and Blobel, G. (1993). An integral membrane protein of the pore membrane domain of the nuclear envelope contains a nucleoporin-like region. J. Cell Biol. 122, 513–521.

40. Jones, S. (2004). An overview of the basic helix-loop-helix proteins. Genome Biol. 5, 226.

41. Guillemot, F., Lo, L.C., Johnson, J.E., Auerbach, A., Anderson, D.J., and Joyner, A.L. (1993). Mammalian achaete-scute homolog 1 is required for the early development of olfactory and autonomic neurons. Cell 75, 463–476.

42. Guillemot, F., Nagy, A., Auerbach, A., Rossant, J., and Joyner, A.L. (1994). Essential role of Mash-2 in extraembryonic development. Nature 371, 333–336.

43. Jubb, A.M., Chalasani, S., Frantz, G.D., Smits, R., Grabsch, H.I., Kavi, V., Maughan, N.J., Hillan, K.J., Quirke, P., and Koeppen, H. (2006). Achaete-scute like 2 (ascl2) is a target of Wnt signalling and is upregulated in intestinal neoplasia. Oncogene *25*, 3445–3457.

44. Ziskin, J.L., Dunlap, D., Yaylaoglu, M., Fodor, I.K., Forrest, W.F., Patel, R., Ge, N., Hutchins, G.G., Pine, J.K., Quirke, P., et al. (2013). In situ validation of an intestinal stem cell signature in colorectal cancer. Gut *62*, 1012–1023.

45. Jang, B.G., Kim, H.S., Kim, K.J., Rhee, Y.Y., Kim, W.H., and Kang, G.H. (2016). Distribution of intestinal stem cell markers in colorectal precancerous lesions. Histopathology *68*, 567–577.

46. Jang, B.G., Lee, B.L., and Kim, W.H. (2015). Intestinal stem cell markers in the intestinal metaplasia of stomach and Barrett's esophagus. PLoS ONE *10*, e0127300.

47. Zhao, R., Quaroni, L., and Casson, A.G. (2012). Identification and characterization of stemlike cells in human esophageal adenocarcinoma and normal epithelial cell lines. J. Thorac. Cardiovasc. Surg. *144*, 1192–1199.

48. Shang, Y., Pan, Q., Chen, L., Ye, J., Zhong, X., Li, X., Meng, L., Guo, J., Tian, Y., He, Y., et al. (2015). Achaete scute-like 2 suppresses CDX2 expression and inhibits intestinal neoplastic epithelial cell differentiation. Oncotarget *6*, 30993–31006.

49. Moons, L.M.G., Bax, D.A., Kuipers, E.J., Van Dekken, H., Haringsma, J., Van Vliet, A.H., Siersema, P.D., and Kusters, J.G. (2004). The homeodomain protein CDX2 is an early marker of Barrett's oesophagus. J. Clin. Pathol. *57*, 1063–1068.

50. Hayes, S., Ahmed, S., and Clark, P. (2011). Immunohistochemical assessment for Cdx2 expression in the Barrett metaplasia-dysplasia-adenocarcinoma sequence. J. Clin. Pathol. *64*, 110–113.

51. Wang, X., Ouyang, H., Yamamoto, Y., Kumar, P.A., Wei, T.S., Dagher, R., Vincent, M., Lu, X., Bellizzi, A.M., Ho, K.Y., et al. (2011). Residual embryonic cells as precursors of a Barrett's-like metaplasia. Cell *145*, 1023–1035.

52. Xian, W., Ho, K.Y., Crum, C.P., and McKeon, F. (2012). Cellular origin of Barrett's esophagus: controversy and therapeutic implications. Gastroenterology *142*, 1424–1430.

53. Weiss, N.S. (2008). Subgroup-specific associations in the face of overall null results: should we rush in or fear to tread? Cancer Epidemiol. Biomarkers Prev. *17*, 1297–1299.

54. Han, S.S., Resenberg, P.S., and Chatterjee, N. (2012). Testing for gene-environment and gene-gene interaction under monotonity constraints. J. Am. Stat. Assoc. *107*, 1441–1452.

55. Song, M., and Nicolae, D.L. (2009). Restricted parameter space models for testing gene-gene interaction. Genet. Epidemiol. *33*, 386–393.

56. Locke, A.E., Kahali, B., Berndt, S.I., Justice, A.E., Pers, T.H., Day, F.R., Powell, C., Vedantam, S., Buchkovich, M.L., Yang, J., et al.; LifeLines Cohort Study; ADIPOGen Consortium; AGEN-BMI Working Group; CARDIOGRAMplusC4D Consortium; CKDGen Consortium; GLGC; ICBP; MAGIC Investigators; MuTHER Consortium; MIGen Consortium; PAGE Consortium; ReproGen Consortium; GENIE Consortium; International Endogene Consortium (2015). Genetic studies of body mass index yield new insights for obesity biology. Nature *518*, 197–206.

57. Gharahkhani, P., Tung, J., Hinds, D., Mishra, A., Vaughan, T.L., Whiteman, D.C., and MacGregor, S.; Barrett's and Esophageal Adenocarcinoma Consortium (BEACON); BEACON study investigators (2016). Chronic gastroesophageal reflux disease shares genetic background with esophageal adenocarcinoma and Barrett's esophagus. Hum. Mol. Genet. *25*, 828–835.

58. Tang, D.I., Geller, N.L., and Pocock, S.J. (1993). On the design and analysis of randomized clinical trials with multiple endpoints. Biometrics *49*, 23–30.

Supplemental Data

Constrained Score Statistics

Identify Genetic Variants Interacting

with Multiple Risk Factors in Barrett's Esophagus

James Y. Dai, Jean de Dieu Tapsoba, Matthew F. Buas, the BEACON Consortium, Harvey A. Risch, and Thomas L. Vaughan

Supplementary Materials

BEACON consortium authors

Wong-Ho Chow Department of Epidemiology, MD Anderson Cancer Center, Houston, TX, USA

Nicholas J. Shaheen

Division of Gastroenterology and Hepatology, University of North Carolina School of Medicine, University of North Carolina, Chapel Hill, North Carolina, USA

Lesley Anderson

Centre for Public Health, Queen's University, Belfast, Northern Ireland

Douglas A. Corley

Division of Research, Kaiser Permanente Northern California, Oakland, California, USA

San Francisco Medical Center, Kaiser Permanente Northern California, San Francisco, California,

Marilie D. Gammon

Department of Epidemiology, University of North Carolina at Chapel Hill, NC, USA

Laura J. Hardie

Division of Epidemiology, University of Leeds, Leeds, UK

Jesper Lagergren

Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden

Division of Cancer Studies, King's College London, London, United Kingdom

David C. Whiteman   Cancer Control, QIMR Berghofer Medical Research Institute, Brisbane QLD, Australia
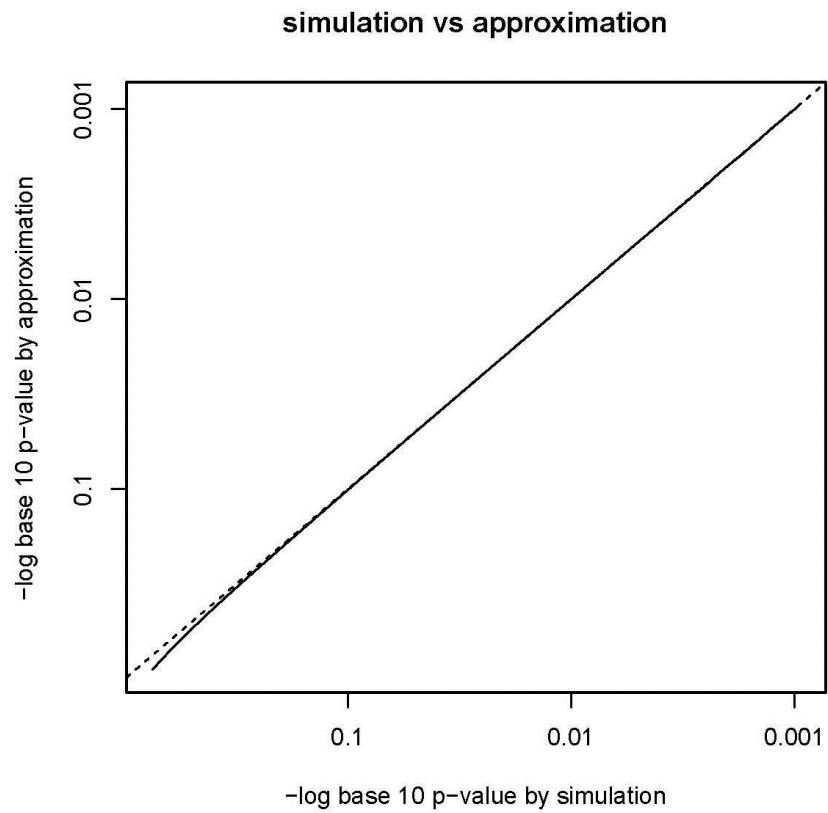
Figure S1. Comparison of the simulation-based p-value and the approximation based p-values for the score statistic for the bidirectional test. The p-value is computed for p=3, the 3-dimentional standard multivariate normal distribution.
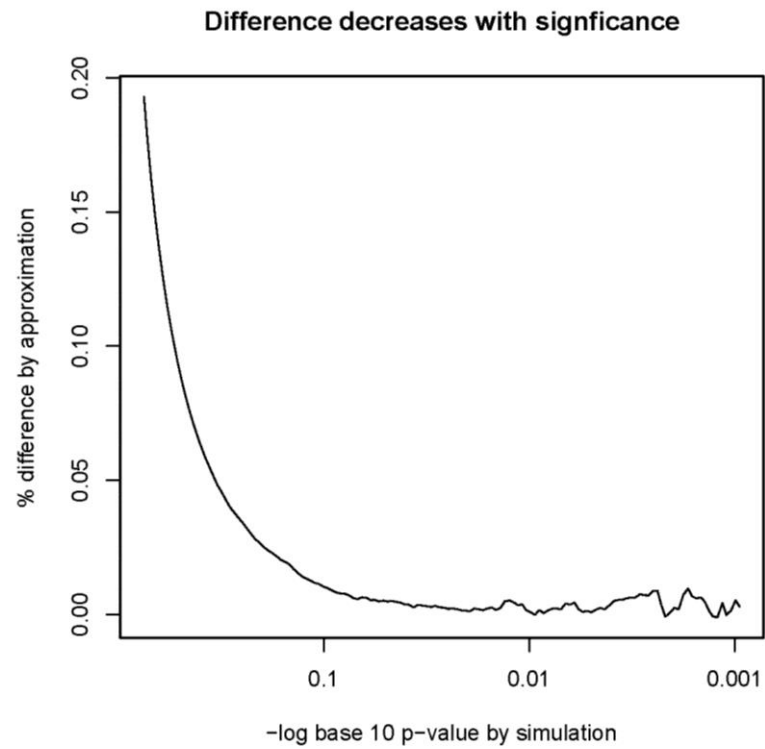
Figure S2. The percentage difference between the simulation-based p-value and the approximation based p-values decreases with the size of p-value. The p-value is computed for bidirectional test with p=3, the 3-dimentional standard multivariate normal distribution.

BE: Score Test for Proportional GxE effects - Without Correction
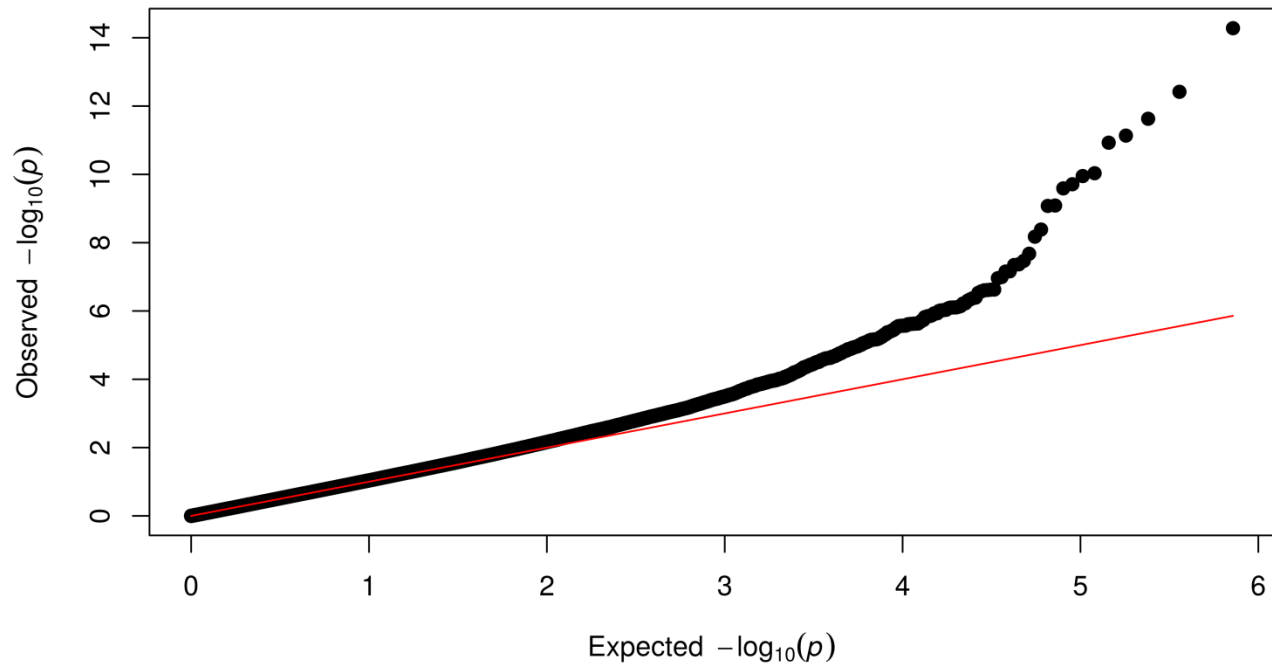
Figure S3. The q-q plot from the score test for 1df proportional interactions on Barrett's esophagus (without correction). The 1-df Chi-square test takes the standard forms of mean and variance for the Rao score statistic.

BE: Tukey test for proportional interactions (Correction 1)
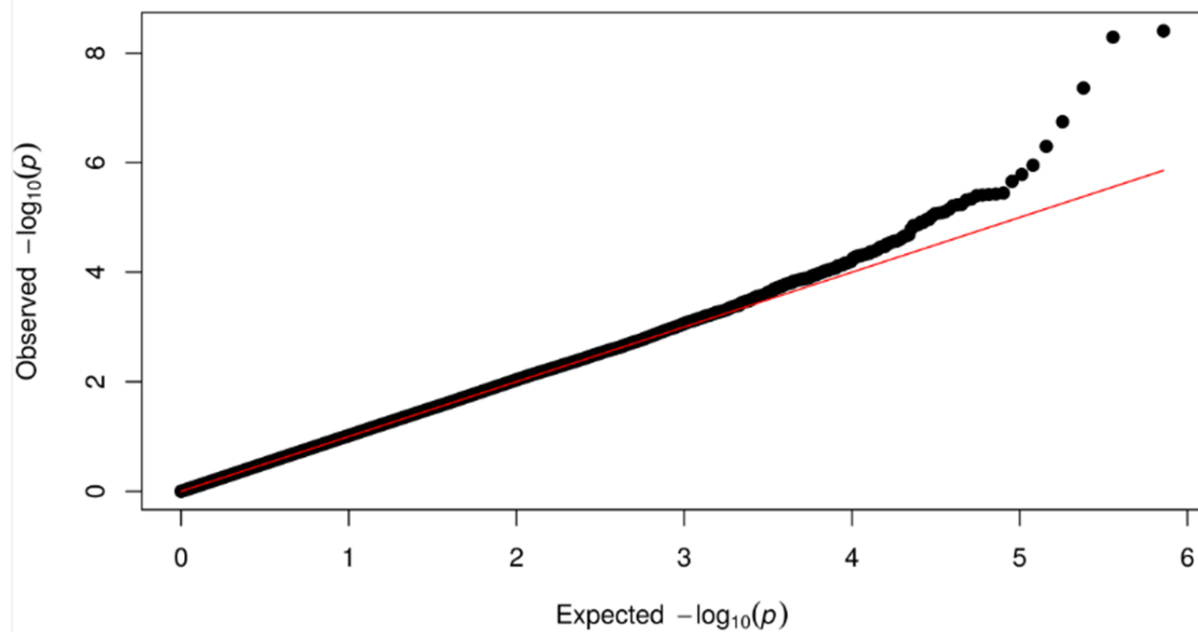
Figure S4. The q-q plot from the score test for 1df proportional interactions on Barrett's esophagus with the first-order correction. The 1-df Chi-square test takes the corrected forms of mean and variance for the Rao score statistic after the first-order Taylor expansion.
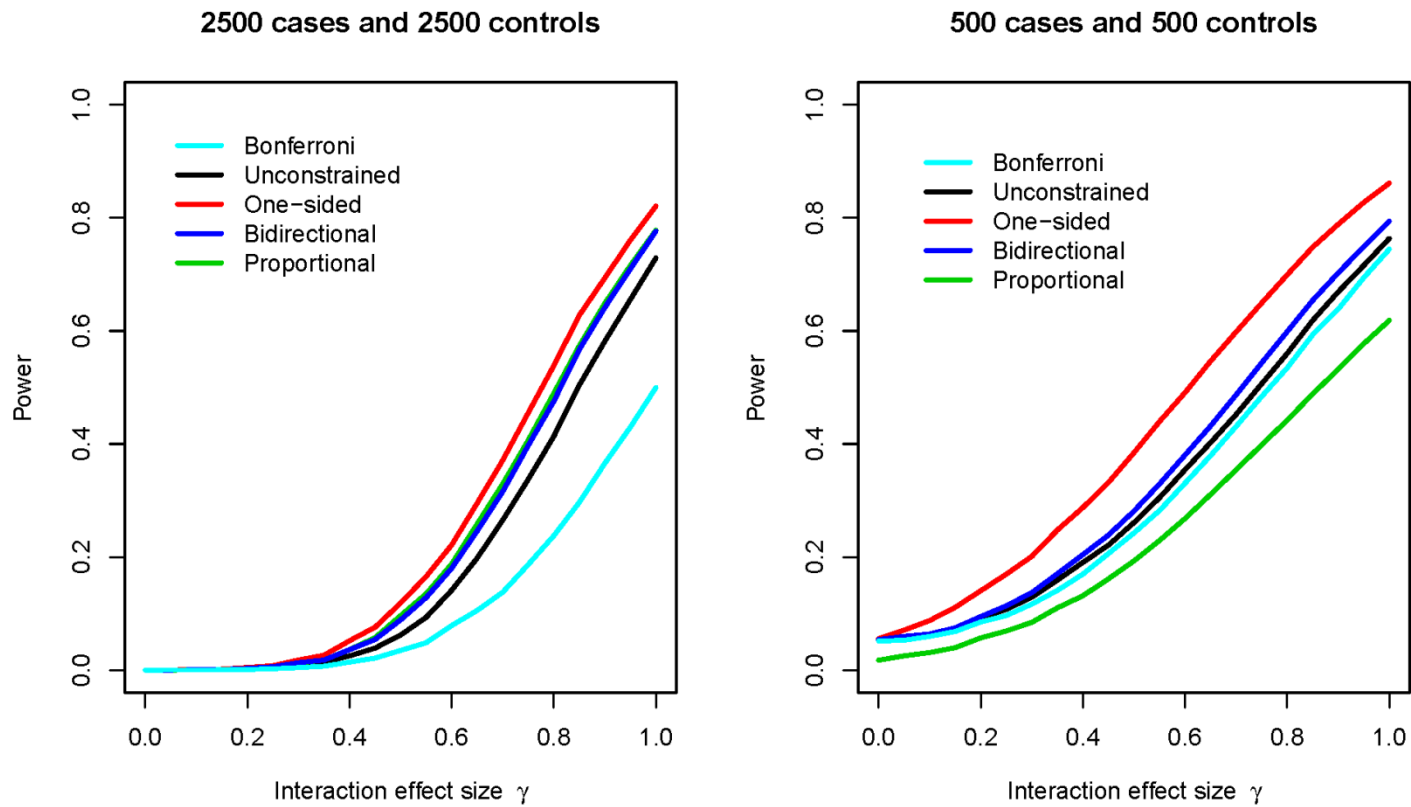
Figure S5. The power of the various test when the main environmental effect is (0.4,0.4,0.4), and all other parameters are the same as Figure 1(a) and Figure 1(b). The left panel shows the scenario with 2500 cases and 2500 controls, the right panel shows the scenario with 500 cases and 500 controls.
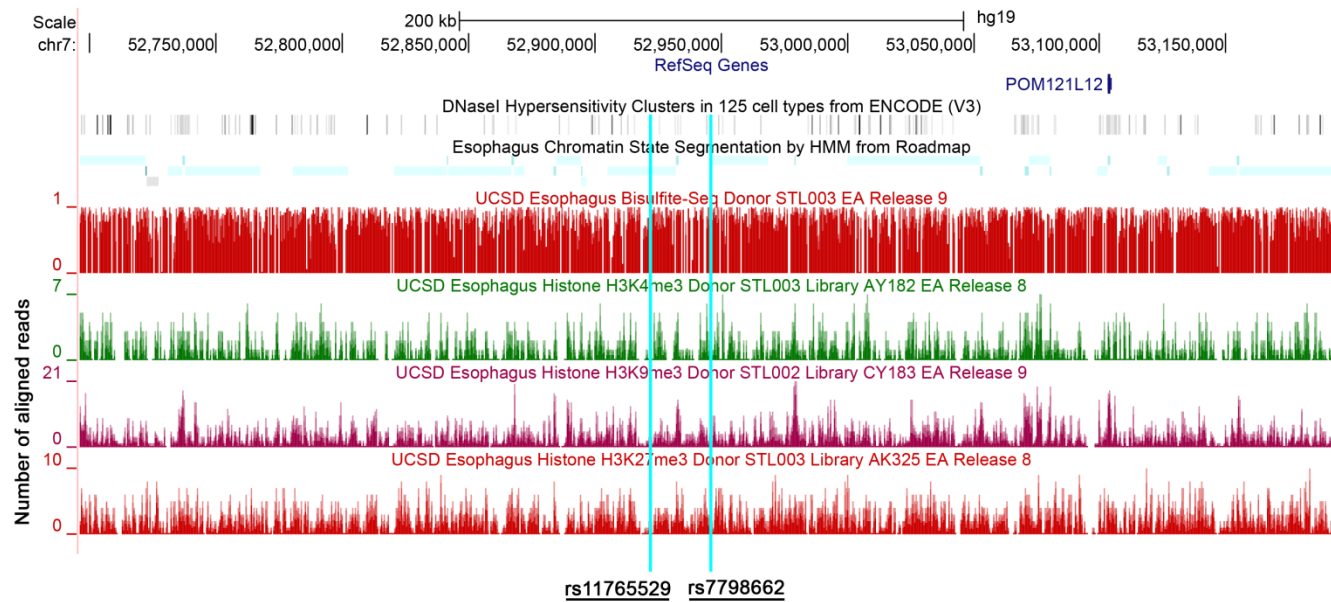
Figure S6A. NIH Roadmap Epigenome Atlas characterization of a ~500-kb region encompassing rs7798662 and rs11765529. Patterns of DNA methylation (Bisulfite-Seq) and key histone modifications (H3K4me3, H3K9me3, H3K27me3) detected in esophageal tissue are displayed. The locus is segmented into regions of "integrative chromatin states" based on a Hidden Markov Model (HMM)-based algorithm (http://www.roadmapepigenomics.org/). Genomic positions are indicated for the two SNPs of interest (underlined).
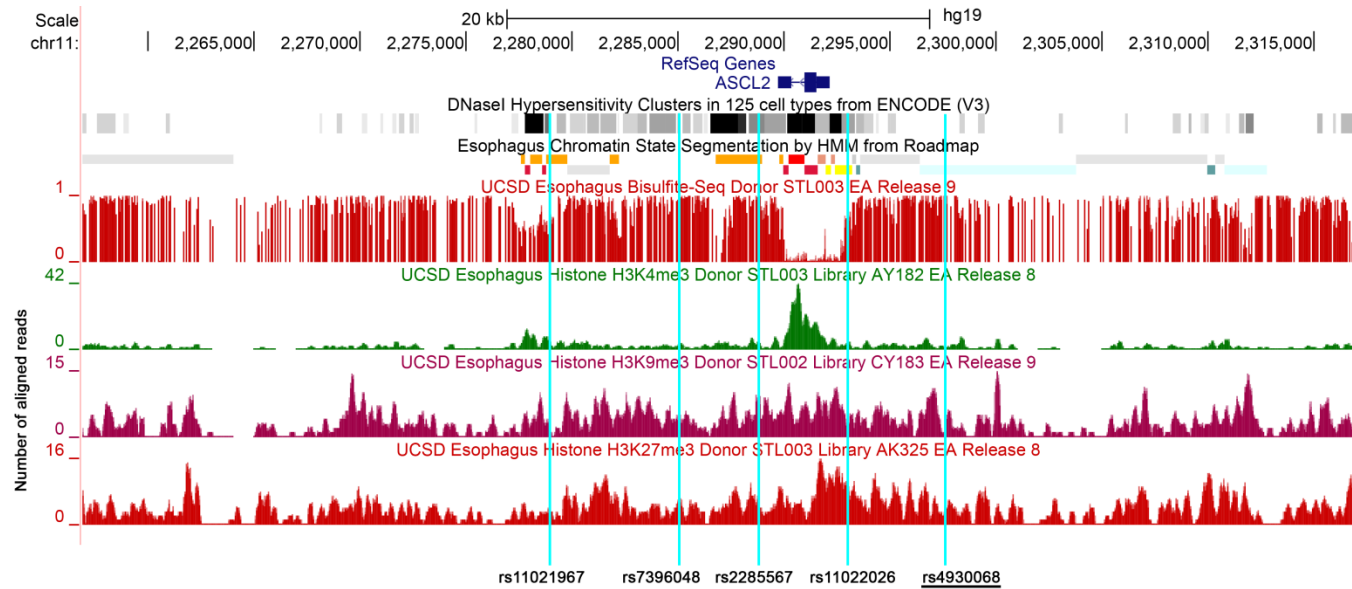
Figure S6B. NIH Roadmap Epigenome Atlas characterization of a ~60-kb region encompassing rs4930068. Patterns of DNA methylation (Bisulfite-Seq) and key histone modifications (H3K4me3, H3K9me3, H3K27me3) detected in esophageal tissue are displayed. The locus is segmented into regions of "integrative chromatin states" based on a Hidden Markov Model (HMM)-based algorithm (http://www.roadmapepigenomics.org/). Genomic positions are indicated for the index SNP (underlined) and four additional variants in high LD ($r^2 > 0.8$) with rs4930068.

Table S1. Functional annotations for variants in high LD (r^2>0.8) with rs7798662.

	SNP	$r^{2(a)}$	Distance[b]	Location	Promoter[c]	Enhancer[d]	DNase[e]	Proteins[f]	eQTL[g]	Motifs[h]
1	rs17691637	1	38367	196kb 5' of POM121L12		ESDR, **GI**, OVRY, PLCNT, VAS				ZBTB33
2	rs10488504	1	38256	196kb 5' of POM121L12						DMRT2,DMRT3,DMRT7,Sox,Zfp187
3	rs112884530	1	36857	194kb 5' of POM121L12						EWSR1-FLI1, Hbp1, MAZ, Mef2, Znf143
4	rs17613916	1	36086	193kb 5' of POM121L12		BRN				
5	rs11766226	1	32207	190kb 5' of POM121L12						E2F,Maf,NF-kappaB,Sox,Zfp410
6	rs112045720	1	31243	189kb 5' of POM121L12						Arid3a,CTCF,Foxq1,Pou3f2
7	rs75557256	1	30940	188kb 5' of POM121L12						CDP,RREB-1
8	rs79160007	1	30930	188kb 5' of POM121L12						GR,Sox
9	rs78469163	1	30884	188kb 5' of POM121L12						CDP
10	rs80103923	1	30882	188kb 5' of POM121L12						CDP,Pou1f1
11	rs111641968	1	30462	188kb 5' of POM121L12						CACD,NRSF,Zfx
12	rs113515559	0.93	30381	188kb 5' of POM121L12						GR,YY1
13	rs74620601	1	30124	188kb 5' of POM121L12			BRN	JUND		AP-1,Gfi1
14	rs112264971	1	29939	187kb 5' of POM121L12						Crx,XBP-1
15	rs79429223	1	29407	187kb 5' of POM121L12						AP-1
16	rs75314167	1	28589	186kb 5' of POM121L12						Gfi1,Pax-5,SP2
17	rs74676024	1	28144	186kb 5' of POM121L12			BLD			Irf,Pax-5,RXRA,p300
18	rs75276194	1	28114	186kb 5' of POM121L12						CDP
19	rs111642729	1	27892	185kb 5' of POM121L12						Irf,Pou2f2,Pou3f3,Pou5f1,Sox
20	rs113494626	1	27414	185kb 5' of POM121L12						Foxj1,Foxk1,RXRA
21	rs75710970	1	27357	185kb 5' of POM121L12						Cdc5,Pax-4,p53
22	rs76973260	1	26484	184kb 5' of POM121L12						ATF3,Arnt,BHLHE40,Ets,Gfi1,Myc,NF-E2,SIRT6,SREBP,Sin3Ak-20,TFE
23	rs80121386	1	26403	184kb 5' of POM121L12						E2F,Hoxa10,Nanog,Pou2f2,Pou3f2,Sox
24	rs17614229	1	26381	184kb 5' of POM121L12						HMG-IY,Irf,Nanog,Pbx-1, Pou1f1, Pou2f2, Pou3f2, Pou3f3, Pou5f1, TATA
25	rs76100117	1	24792	182kb 5' of POM121L12		ESC, OVRY				Foxp1,IRC900814,Mef2,NRSF,RFX5
26	rs75931245	1	24784	182kb 5' of POM121L12		ESC, OVRY				Foxf2,Mef2,Myf,NRSF,Sin3Ak-20,ZEB1

	SNP	$r^{2(a)}$	Distance[b]	Location	Promoter[c]	Enhancer[d]	DNase[e]	Proteins[f]	eQTL[g]	Motifs[h]
27	rs111256487	1	24624	182kb 5' of POM121L12		ESC, OVRY				Bcl6b,CEBPB,Hsf,STAT
28	rs74629733	1	24560	182kb 5' of POM121L12		ESC, OVRY				
	SNP	$r^{2(a)}$	**Distance**[b]	**Location**	**Promoter**[c]	**Enhancer**[d]	**DNase**[e]	**Proteins**[f]	**eQTL**[g]	**Motifs**[h]
29	rs77873291	1	24520	182kb 5' of POM121L12		OVRY				AIRE,Gfi1,Ik-2,NF-AT,Pou2f2
30	rs77842860	1	24385	182kb 5' of POM121L12		OVRY				HNF1,Mef2,Pbx-1
31	rs75082417	1	24370	182kb 5' of POM121L12		OVRY				Pou2f2
32	<u>**rs11765529**</u>	1	23722	181kb 5' of POM121L12						HDAC2,Zbtb3
33	rs11769735	1	23708	181kb 5' of POM121L12						
34	rs11772201	1	23704	181kb 5' of POM121L12						Arid5b
35	rs11769032	1	23691	181kb 5' of POM121L12						Arid5b,Fox,p300
36	rs78683211	1	23505	181kb 5' of POM121L12						Arid5a,Foxa,Pou5f1,Sox
37	rs75960827	1	23121	181kb 5' of POM121L12						Arid5b,Cdx2,Hoxa10,Hoxb13,Hoxb9,Hoxd10,Pou2f2,Sox
38	rs77416190	1	22654	180kb 5' of POM121L12						Foxa,Foxf1,TCF12,p300
39	rs111384348	1	22307	180kb 5' of POM121L12						
40	rs11767295	1	22045	179kb 5' of POM121L12						CEBPB,GATA,XBP-1,p300
41	rs11767394	1	21753	179kb 5' of POM121L12						Foxa
42	rs11767396	1	21736	179kb 5' of POM121L12						
43	rs11760456	1	21678	179kb 5' of POM121L12						Cdx,Evi-1,Foxj1,TATA
44	rs112287825	1	21195	179kb 5' of POM121L12						HNF4,Pou2f2,RAR,RXR::LXR,RXRA,TATA
45	rs112743003	1	21086	178kb 5' of POM121L12						AhR,Egr-1,Pax-6,RFX5
46	rs77462617	1	20423	178kb 5' of POM121L12						
47	rs77737967	1	20296	178kb 5' of POM121L12						RXRA
48	rs111231293	1	19389	177kb 5' of POM121L12						Evi-1,FAC1
49	rs11773824	1	18959	176kb 5' of POM121L12						Foxk1,Foxp1,Irf,Mef2
50	rs11763206	1	18930	176kb 5' of POM121L12						Arnt,E2A,Mtf1,Myc,Rad21
51	rs11770293	1	18761	176kb 5' of POM121L12						GATA
52	rs75553733	1	17969	175kb 5' of POM121L12						Foxj2,PLZF,Pou3f2,Sox,TATA
53	rs113629168	1	17823	175kb 5' of POM121L12						
54	rs17692657	1	17523	175kb 5' of POM121L12						AP-1,E2F,Foxp1,Irf,NF-Y,Pbx3,RFX5,SP1,SP2,Znf143
55	rs79478382	1	17334	175kb 5' of POM121L12						LUN-1,Roaz
56	rs74496138	1	17287	175kb 5' of POM121L12						Hdx,NF-Y,STAT
57	rs77799175	1	17267	175kb 5' of POM121L12						TATA

	SNP	$r^{2(a)}$	Distance[b]	Location	Promoter[c]	Enhancer[d]	DNase[e]	Proteins[f]	eQTL[g]	Motifs[h]
58	rs111781807	1	16981	174kb 5' of POM121L12						ERalpha-a,ZID
59	rs112685756	1	16615	174kb 5' of POM121L12						AP-1,ELF1
60	rs10488501	1	16218	174kb 5' of POM121L12						Evi-1,GATA,Hltf
61	rs113188931	1	15714	173kb 5' of POM121L12						CTCF,E2A,ERalpha-a,Esr2,Mrg,SREBP,Tgif1
62	rs80001306	1	15387	173kb 5' of POM121L12		VAS				HNF4,Pax-4
63	rs77106104	1	15257	173kb 5' of POM121L12						Bbx,Bsx,Cart1,Dbx2,Esx1,Foxj2,Gbx1,HMG-IY, Hoxa5, Hoxa7, Hoxb7, Hoxd8, Lhx3, Ncx, Nkx6-1, Pax-4, Pou2f2, Pou3f2, Pou3f4, Pou6f1
64	rs113303892	1	15075	172kb 5' of POM121L12						Crx,Gsc,Obox3,Otx2,Pitx3,RORalpha1,RREB-1,SIX5
65	rs111886202	1	14855	172kb 5' of POM121L12						Mef2
66	rs75350842	1	14712	172kb 5' of POM121L12						CDP,Pax-4,Pou2f2,Pou3f2
67	rs112225991	1	14669	172kb 5' of POM121L12						Pax-4,Spz1
68	rs4358759	1	14140	172kb 5' of POM121L12						
69	rs11762984	1	13796	171kb 5' of POM121L12			PANC			ATF3
70	rs77503606	1	13578	171kb 5' of POM121L12						
71	rs77874117	1	13456	171kb 5' of POM121L12						Mrg1::Hoxa9
72	rs112393748	1	13093	171kb 5' of POM121L12			IPSC			Foxp3,Nkx3,SIX5
73	rs11771017	1	12782	170kb 5' of POM121L12						Arid5b,LBP-1,TATA
74	rs11760967	1	12545	170kb 5' of POM121L12						AIRE
75	rs11761682	1	12258	170kb 5' of POM121L12						
76	rs112670870	1	11785	169kb 5' of POM121L12						Dobox4,Nkx3
77	rs77846464	1	11602	169kb 5' of POM121L12						AIRE,CIZ,DMRT1
78	rs77201830	1	11275	169kb 5' of POM121L12	VAS					
79	rs76176075	1	11228	169kb 5' of POM121L12	VAS					CTCF,Mef2,Rad21
80	rs76947106	1	11085	168kb 5' of POM121L12	VAS					STAT,Sox,ZEB1
81	rs113653466	1	10431	168kb 5' of POM121L12						Foxm1,Nkx2,Nkx3,YY1
82	rs75547585	1	10159	168kb 5' of POM121L12						Crx,Rad21
83	rs112321404	1	9929	167kb 5' of POM121L12						Tgif1
84	rs113117967	1	9599	167kb 5' of POM121L12						FAC1,Foxo,Foxp1,Pax-4,RREB-1, SIX5, STAT, Zfp105
85	rs11765486	1	8758	166kb 5' of POM121L12						Irx,TR4
86	rs111437435	1	8067	165kb 5' of POM121L12						
87	rs76909095	1	7891	165kb 5' of POM121L12						GATA,Lhx8

	SNP	$r^{2(a)}$	Distance[b]	Location	Promoter[c]	Enhancer[d]	DNase[e]	Proteins[f]	eQTL[g]	Motifs[h]
88	rs113691255	1	7799	165kb 5' of POM121L12						CDP,CTCF,SEF-1,Sox
89	rs77466972	1	7733	165kb 5' of POM121L12			**GI**			PLZF,Pou6f1,Zfp187
90	rs77581067	1	7725	165kb 5' of POM121L12						ATF2,Ik-2,NF-AT1, PLZF, Pou2f2, YY1, Zfp410
91	rs74687520	1	7303	165kb 5' of POM121L12						EWSR1-FLI1,Ik-1,TFII-I
92	rs117249146	1	6987	164kb 5' of POM121L12						Foxp1,HDAC2,HMG-IY,HNF1,Mef2,TATA
93	rs75267662	1	6106	164kb 5' of POM121L12			IPSC			Maf,Sox
94	rs77267459	1	6084	163kb 5' of POM121L12			IPSC			
95	rs78645775	1	5085	162kb 5' of POM121L12						RFX5,ZEB1,Zfp105
96	rs76627272	1	5025	162kb 5' of POM121L12						DBP,Fox,Foxa,Foxd1,Foxd3,Foxi1,Foxj1,Foxj2,Foxq1,HDAC2,Hmx,NF-Y,p300
97	rs6977746	1	4547	162kb 5' of POM121L12						CEBPB,Evi-1,HNF1,Mef2,Osf2,PEBP,Pax-4
98	rs6958403	1	4474	162kb 5' of POM121L12						ERalpha-a,PLZF,Pax-5,RXRA
99	rs6943949	1	4455	162kb 5' of POM121L12						GATA,ZEB1
100	rs112665251	1	4001	161kb 5' of POM121L12			BLD			Nanog
101	rs111700951	1	3684	161kb 5' of POM121L12						BAF155,Mef2,Myc
102	rs111787575	1	3404	161kb 5' of POM121L12						HNF4,Maf,NR4A,SREBP
103	rs111972542	1	3322	161kb 5' of POM121L12						Ets,Myf,ZEB1,Znf143
104	rs111553333	1	3275	161kb 5' of POM121L12						ATF3,ZBTB33
105	rs112424724	1	3245	161kb 5' of POM121L12						
106	rs112779901	1	3212	161kb 5' of POM121L12						Ik-1,NRSF,Zfx
107	rs77470508	1	2627	160kb 5' of POM121L12						PLAG1,RREB-1,TBX5
108	rs112208947	1	2575	160kb 5' of POM121L12						ERalpha-a,KAP1
109	rs111677539	1	2089	160kb 5' of POM121L12						LBP-9,TAL1
110	rs113408419	1	2066	159kb 5' of POM121L12						HNF1,Pax-4
111	rs77618686	1	1532	159kb 5' of POM121L12						Homez
112	rs76355749	1	1219	159kb 5' of POM121L12						Evi-1,Pou1f1
113	rs75588263	1	1082	158kb 5' of POM121L12						VDR
114	rs6978398	1	630	158kb 5' of POM121L12						CDP
115	rs7779800	1	169	158kb 5' of POM121L12						Foxq1,HNF4
116	**rs7798662**	1	0	157kb 5' of POM121L12						Smad3
117	rs74777222	1	339	157kb 5' of POM121L12						Evi-1

| 118 | rs117339222 | 0.87 | 900 | 157kb 5' of POM121L12 | | | | | | Foxc1,Pax-4 |

[a]Pair-wise correlation (r^2) with rs7798662, [b]Distance from rs7798662 (bp), [c,d]Tissue/cell types in which indicated chromatin marks were detected (NIH Roadmap Epigenomics Consortium), [e]DNAse hypersensitivity (Roadmap), [f]Protein recruitment determined by chromatin-immunoprecipitation (ENCODE), [g]Evidence for expression quantitative trait locus (eQTL) (Y: Yes), [h]Predicted binding motifs altered. Annotations extracted from HaploReg-v4 and Regulome-DB. Abbreviations for tissue/cell types accessible online.