Biostatistics (2013), *pp.* 1–31
doi:10.1093/biostatistics/kxt006

Test for interactions between a genetic marker set and environment in generalized linear models Supplementary Materials

XINYI LIN, SEUNGGUEN LEE

Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA

DAVID C. CHRISTIANI

Department of Environmental Health, Harvard School of Public Health, Boston, MA, USA

XIHONG LIN*

Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA

xlin@hsph.harvard.edu

*To whom correspondence should be addressed.

A. Asymptotic Bias of the Single Marker GE Interaction Test

A.1. Derivation of the asymptotic limit $\left(\tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\alpha}_{3j}, \tilde{\beta}_j\right)$ *of the single marker GE interaction test*

Assume that $g(\cdot)$ is an identity link function and $\boldsymbol{G}$ and E are binary. Define $\pi_E = \mathcal{E}(E), \pi_j = \mathcal{E}(G_j), \pi_{jk} = \mathcal{E}(G_j G_k), \pi_{jE} = \mathcal{E}(G_j E), \pi_{jkE} = \mathcal{E}(G_j G_k E)$. Then Equation (3.4) simplifies to a system of the following four equations:

$$\begin{aligned}
&\alpha_1 - \alpha_1^* + (\alpha_2 - \alpha_2^*)\,\pi_E + \sum_{k=1}^{p} \pi_k \alpha_{3k} + \sum_{k=1}^{p} \pi_{kE} \beta_k - \pi_j \alpha_{3j}^* - \pi_{jE} \beta_j^* = 0 \\
&(\alpha_1 - \alpha_1^* + \alpha_2 - \alpha_2^*)\,\pi_E + \sum_{k=1}^{p} \pi_{kE} \left(\alpha_{3k} + \beta_k\right) - \pi_{jE}\left(\alpha_{3j}^* + \beta_j^*\right) = 0 \\
&\left(\alpha_1 - \alpha_1^* - \alpha_{3j}^*\right)\pi_j + \sum_{k=1}^{p} \pi_{jk} \alpha_{3k} + \sum_{k=1}^{p} \pi_{jkE} \beta_k + \left(\alpha_2 - \alpha_2^* - \beta_j^*\right)\pi_{jE} = 0 \\
&\left(\alpha_1 - \alpha_1^* + \alpha_2 - \alpha_2^* - \alpha_{3j}^* - \beta_j^*\right)\pi_{jE} + \sum_{k=1}^{p} \pi_{jkE} \left(\alpha_{3k} + \beta_k\right) = 0
\end{aligned}$$

Solving the system of four equations for $\left(\alpha_1^*, \alpha_2^*, \alpha_{3j}^*, \beta_j^*\right)$ as a function of $(\alpha_1, \alpha_2, \boldsymbol{\alpha}_3, \boldsymbol{\beta})$ gives the asymptotic limits, $\left(\tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\alpha}_{3j}, \tilde{\beta}_j\right)$, Equation (3.5), presented in Section 3.1 of the main manuscript.

A.2. Violation of homoscedasticity assumption in linear regression

Let $g(\cdot)$ be the identity link function. The multi-marker GE interaction model (3.1) is then

$$\mu_i = \alpha_1 + \alpha_2 E_i + \sum_{k=1}^{p} G_{ik}\alpha_{3k} + \sum_{k=1}^{p} G_{ik}E_i\beta_k. \tag{A-1}$$

The corresponding single marker GE interaction test assumes the following misspecified model using only the j^{th} genetic marker $(j = 1, \cdots, p)$

$$\mu_i = \alpha_1^* + \alpha_2^* E_i + G_{ij}\alpha_{3j}^* + G_{ij}E_i\beta_j^*. \tag{A-2}$$

In Section 3 of the main manuscript, we noted that under the null hypothesis $H_0 : \boldsymbol{\beta} = 0$ in the multi-marker GE interaction model (A-1), if (i) (G_j, E) is independent of $\{G_k\}_{k\neq j}$ or (ii) G_j is independent of $\left(E, \{G_k\}_{k\neq j}\right)$ or (iii) $\{G_k\}_{k=1}^p$ is independent of E, we have that $\tilde{\beta}_j = 0$. This means the single marker GE interaction coefficient MLE is asymptotically unbiased under the null.

In fitting models (A-1) or (A-2), it is common to make the homoscedasticity assumption, for example, in linear regression. Specifically for multi-marker GE interaction model (A-1), it is common to assume

$$\text{Var}\left(Y_i | E_i, G_{i1}, \cdots, G_{ip}\right) = \sigma^2, \tag{A-3}$$

and for the misspecified single marker GE interaction model (A-2),

$$\text{Var}\left(Y_i | E_i, G_{ij}\right) = \sigma_*^2. \tag{A-4}$$

In general if models (A-1) and (A-3) hold, then (A-4) is generally not true. When models (A-1) and (A-3) and the null hypothesis $H_0 : \boldsymbol{\beta} = 0$ hold, $\text{Var}\left(Y_i | E_i, G_{ij}\right)$ is given by

$$\begin{aligned}\text{Var}\left(Y_i | E_i, G_{ij}\right) &= \mathcal{E}\left[\text{Var}\left(Y_i | E_i, G_{i1}, \cdots, G_{ip}\right) | E_i, G_{ij}\right] + \text{Var}\left[\mathcal{E}\left(Y | E_i, G_{i1}, \cdots, G_{ip}\right) | E_i, G_{ij}\right] \\ &= \sigma^2 + \text{Var}\left[\sum_{k=1}^{p} G_{ik}\alpha_{3k} | E_i, G_{ij}\right]. \end{aligned} \tag{A-5}$$

Thus in general, under the null hypothesis, if the true model is the multi-marker GE interaction models (A-1) and (A-3), then $\mathrm{Var}\,(Y_i|E_i, G_{ij})$ is a function of E_i, G_{ij}. This means that the single marker GE interaction test violates the homoscedasticity assumption and model (A-4) does not hold. Hence even if the single marker GE interaction coefficient MLE is asymptotically unbiased under the null, inference based on the single marker GE models (A-2) and (A-4) can still be wrong as the standard error estimate for the MLE can still be biased since the homoscedasticity assumption is violated. We note that if (i) (G_j, E) is independent of $\{G_k\}_{k\neq j}$ holds or (ii) there is no main SNP effects (i.e. $\alpha_{3k} = 0$ for all $k \neq j$) under the null hypothesis, then the homoscedasticity assumption (A-4) for the misspecified single marker GE interaction model (A-2) is still satisfied.

B. Gene-Environment Set Association Test (GESAT)

B.1. Proof of Equation (4.3)

It is known that the ridge estimator $\hat{\boldsymbol{\alpha}}^{\hat{\lambda}}$ is a $\sqrt{n}$-consistent estimator of $\boldsymbol{\alpha}_0$ with $\hat{\lambda} = o(\sqrt{n})$ (Knight and Fu, 2000). A Taylor expansion of the estimating equation for $\boldsymbol{\alpha}$ around $\boldsymbol{\alpha}_0$ gives

$$\sqrt{n}\left(\hat{\boldsymbol{\alpha}}^{\hat{\lambda}} - \boldsymbol{\alpha}_0\right) = \sqrt{n}\boldsymbol{\Omega}^{\hat{\lambda}}\left(\boldsymbol{\alpha}_0\right)^{-1}\boldsymbol{U}^{\hat{\lambda}}\left(\boldsymbol{\alpha}_0\right) + o_p(1). \tag{B-6}$$

Since $\hat{\lambda} = o(\sqrt{n})$, we have

$$n^{-\frac{1}{2}}\boldsymbol{S}^{\mathsf{T}}\boldsymbol{\mu}^{'}\left(\boldsymbol{\alpha}_0\right)\boldsymbol{\Omega}^{\hat{\lambda}}\left(\boldsymbol{\alpha}_0\right)^{-1}\hat{\lambda}\boldsymbol{I}_2\boldsymbol{\alpha}_0 = \left[n^{-1}\boldsymbol{S}^{\mathsf{T}}\boldsymbol{\Delta}_0^{-1}\tilde{\boldsymbol{X}}\right]\left[n^{-1}\tilde{\boldsymbol{X}}^{\mathsf{T}}\boldsymbol{\Delta}_0^{-1}\tilde{\boldsymbol{X}} + n^{-1}\hat{\lambda}\boldsymbol{I}_2\right]^{-1} n^{-\frac{1}{2}}\hat{\lambda}\boldsymbol{I}_2\boldsymbol{\alpha}_0 = o_p(1), \tag{B-7}$$

where $\boldsymbol{\mu}^{'}\left(\boldsymbol{\alpha}_0\right) = \boldsymbol{\Delta}_0^{-1}\tilde{\boldsymbol{X}}$, $\Omega^{\hat{\lambda}}\left(\boldsymbol{\alpha}_0\right) = \tilde{\boldsymbol{X}}^{\mathsf{T}}\boldsymbol{\Delta}_0^{-1}\tilde{\boldsymbol{X}} + \hat{\lambda}\boldsymbol{I}_2$, and $\boldsymbol{I}_2$ is defined in Section 4.1 of the main manuscript. It follows that

$$\begin{aligned} n^{-\frac{1}{2}}\boldsymbol{S}^{\mathsf{T}}\left(\boldsymbol{Y} - \hat{\boldsymbol{\mu}}\right) &= n^{-\frac{1}{2}}\boldsymbol{S}^{\mathsf{T}}\left[\boldsymbol{Y} - \boldsymbol{\mu}\left(\boldsymbol{\alpha}_0\right) - \left\{\boldsymbol{\mu}\left(\hat{\boldsymbol{\alpha}}^{\hat{\lambda}}\right) - \boldsymbol{\mu}\left(\boldsymbol{\alpha}_0\right)\right\}\right] \\ &= n^{-\frac{1}{2}}\boldsymbol{S}^{\mathsf{T}}\left(\boldsymbol{Y} - \boldsymbol{\mu}\left(\boldsymbol{\alpha}_0\right)\right) - n^{-1}\boldsymbol{S}^{\mathsf{T}}\boldsymbol{\mu}^{'}\left(\boldsymbol{\alpha}_0\right)\sqrt{n}\left(\hat{\boldsymbol{\alpha}}^{\hat{\lambda}} - \boldsymbol{\alpha}_0\right) + o_p(1). \end{aligned}$$

Using (B-6) and (B-7), we have

$$\begin{aligned} n^{-\frac{1}{2}}\boldsymbol{S}^{\mathsf{T}}\left(\boldsymbol{Y} - \hat{\boldsymbol{\mu}}\right) &= n^{-\frac{1}{2}}\boldsymbol{S}^{\mathsf{T}}\left(\boldsymbol{Y} - \boldsymbol{\mu}\left(\boldsymbol{\alpha}_0\right)\right) - n^{-1}\boldsymbol{S}^{\mathsf{T}}\boldsymbol{\mu}^{'}\left(\boldsymbol{\alpha}_0\right)\left[\sqrt{n}\boldsymbol{\Omega}^{\hat{\lambda}}\left(\boldsymbol{\alpha}_0\right)^{-1}\boldsymbol{U}^{\hat{\lambda}}\left(\boldsymbol{\alpha}_0\right) + o_p(1)\right] + o_p(1) \\ &= n^{-\frac{1}{2}}\boldsymbol{S}^{\mathsf{T}}\left[\boldsymbol{I}_{n\times n} - \boldsymbol{\mu}^{'}\left(\boldsymbol{\alpha}_0\right)\boldsymbol{\Omega}^{\hat{\lambda}}\left(\boldsymbol{\alpha}_0\right)^{-1}\tilde{\boldsymbol{X}}^{\mathsf{T}}\right]\left[\boldsymbol{Y} - \boldsymbol{\mu}\left(\boldsymbol{\alpha}_0\right)\right] + o_p(1) \\ &= n^{-\frac{1}{2}}\boldsymbol{S}^{\mathsf{T}}\left[\boldsymbol{I}_{n\times n} - \boldsymbol{H}_*^{\hat{\lambda}}\right]\left[\boldsymbol{Y} - \boldsymbol{\mu}\left(\boldsymbol{\alpha}_0\right)\right] + o_p(1) \\ &= n^{-\frac{1}{2}}\boldsymbol{S}^{\mathsf{T}}\boldsymbol{\Delta}_0^{-1}\left[\boldsymbol{I}_{n\times n} - \boldsymbol{H}^{\hat{\lambda}}\right]\left[\boldsymbol{y} - \tilde{\boldsymbol{X}}\boldsymbol{\alpha}_0\right] + o_p(1). \end{aligned}$$

The results in Equation (4.3) follow immediately.

B.2. Selection of the tuning parameter

Estimation of the tuning parameter λ is important. One approach is to use cross-validation, which is computationally intensive. Instead, we estimate λ by minimizing the generalized cross validation (GCV) (O'Sullivan *and others*, 1986), where

$$\hat{\lambda} = \arg\min \frac{(\boldsymbol{Y} - \hat{\boldsymbol{\mu}})^{\intercal} \boldsymbol{\Delta}_0 (\boldsymbol{Y} - \hat{\boldsymbol{\mu}})}{n \left\{1 - \operatorname{tr}\left(\boldsymbol{H}_{*}^{\lambda}\right)/n\right\}^2},$$

and $\boldsymbol{\Delta}_0$ is evaluated under the null hypothesis. In simulation studies and the data example, we search for the optimal λ within a range $[0, C]$, where we set $C = \sqrt{n}/\log(n)$, since the approximate null distribution of the test statistic Q is obtained under the assumption that $\hat{\lambda} = o(\sqrt{n})$.

C. Simulation Studies

This section serves two purposes. Firstly, we provide additional details on the simulations reported in the main manuscript. Secondly, we report additional simulation results.

C.1. min test

The *min* test is conducted in three steps. Firstly, using only the genotypes of the SNPs within the tested SNP-set, we estimate the effective number of SNPs within the tested SNP-set using Gao *and others* (2008). This effective number of SNPs is typically less than the total number of tested SNPs within the SNP-set as it accounts for the LD among the SNPs in the region. This effective number of SNPs is thus an estimate of the effective number of tests conducted, taking into account that the tests are dependent due to LD between the SNPs. Second, one fits an individual marker GE test to each SNP in the tested SNP-set. Lastly, one obtains a p-value for the SNP-set by multiplying the minimum of these individual marker GE p-values by the effective number of SNPs (modified Bonferroni correction). This modified Bonferroni correction corrects the minimum p-value from the dependent tests using the number of effective tests.

C.2. Selection of Causal SNPs for the 15q24-25.1 region in simulations

We simulated 166 HapMap SNPs in the 15q24-25.1 region using the LD structure of the CEU population in the HapMap project (Gibbs *and others*, 2003; Thorisson *and others*, 2005). See the top panel of Supplementary Figure 1 for LD structure of this region. To mimic the Harvard lung cancer data, only the 26 typed variants on Illumina 610-Quad array in this region are used for analysis. The LD structure of these 26 typed SNPs is shown in the bottom panel of Supplementary Figure 1. We restricted the analysis to common variants (MAF $\geqslant 0.05$), giving $p = 25 - 26$. To select causal SNPs, we used Haploview (Barrett *and others*, 2005) to identify a group of candidate untyped causal SNPs such that all common variants (MAF $\geqslant 0.2$) in the 15q24-25.1 region are in LD ($R^2 \geqslant 0.5$) with at least one SNP in the candidate causal group, and each SNP in the candidate causal SNP group is correlated with at least one of the 26 typed SNPs ($R^2 \geqslant 0.7$). Thus the effects of the untyped candidate causal SNPs are partially captured by analyzing the 26 typed SNPs. This gives a total of 5 candidate SNPs from which the causal SNPs are chosen. In other words, as the candidate causal SNPs are not typed, they are not used in analysis. Their effects are captured through their partial LD with the typed SNPs, which are used in analysis. This is a typical feature of GWAS. The LD plot of these five candidate untyped causal SNPs (in rectangular boxes) together with the 26 typed SNPs (not in rectangular boxes) are shown in the bottom panel of Supplementary Figure 1.

C.3. Additional Type 1 Error Simulations for Binary Trait

We studied the empirical Type 1 error rate of GESAT under a wide variety of scenarios. These results show that in all the scenarios we examined, the Type 1 error rate using GESAT is preserved, while the *min* test can be slightly conservative.

The 15q24-25.1 region is used and the untyped causal and tested SNPs are the same as that described in Supplementary Section C.2. Further details on simulation set-up are given in Section 5.1 of the main manuscript. We varied (a) sample size ($n = 2000$ and $n = 4000$), (b) environmental variable (binary w.p. 0.87, binary w.p. 0.5 and continuous), (c) minor allele frequency of the causal loci (the 5 candidate untyped causal SNPs have different MAF ranging from 0.26 to 0.37), (d) number of causal loci (0 ,1, 2) and (e) different α levels ($\alpha = 5e-02, 5e-03, 5e-04$). The Type 1 error rate for each scenario is evaluated using 10^5 simulations. Supplementary Tables 1 to 3 show the empirical Type 1 error rates for these set of simulations using the 15q24-25.1 region. Note that a subset of the results ($\alpha = 0.05$, $n = 2000$, No. of causal main effect = 0 or 2) in Supplementary Tables 1 to 3 are also reported in Table 2 of the main manuscript.

In addition, to vary (f) number of tested SNPs and LD of the tested SNPs, we conducted another set of simulations using the ASAH1 gene (Supplementary Tables 4 to 6). The ASAH1 gene has strong LD while the 15q24-25.1 region has moderate LD. The ASAH1 gene has a total of 62 SNPs. For each of the 10^5 simulations, we randomly selected one of the 62 SNPs to have causal main effect and used all SNPs that have MAF $\geqslant 0.05$ in the testing procedure for both GESAT and *min* test. Note that for the ASAH1 gene Type 1 error simulations, all the SNPs (including the selected causal SNP) with MAF $\geqslant 0.05$ are tested (giving $p = 41 - 53$). The covariates are the same as that used in the 15q24-25.1 region simulations above and described in Section 5.1 of the main manuscript and we also varied (a) to (e) as described above.

C.4. Comparing power of GESAT and min test when there is a single genotyped causal locus

A common current practice in GWAS is to impute all the HapMap/1000 genomes SNPs using the typed SNPs and HapMap/1000 genomes data. To reduce computational burden of imputation and data storage, this set of simulations uses a smaller gene, the ASAH1 gene. The ASAH1 gene consists of 62 HapMap SNPs and 13 typed SNPs (SNPs that are on the Illumina 610-Quad array). For each simulation, we randomly chose **one** of the 13 typed SNPs to have both causal SNP main effect and causal SNP-Environment effect. We first simulated genotype data for the 13 typed SNPs in ASAH1 gene based on the LD structure for the HapMap CEU population. We generated a binary outcome assuming:

$$\operatorname{logit}\left[P(Y_i = 1|X_{1i}, X_{2i}, E_i, \mathrm{SNPcausal}_i)\right] = \alpha_0 + 0.05X_{1i} + 0.057X_{2i} + 0.64E_i + 0.4\mathrm{SNPcausal}_i + \beta_1\mathrm{SNPcausal}_i \times E_i$$

where as in Section 5.1 of the main manuscript $\alpha_0 = \log(0.01/0.99)$, X_1 mimics age and is normally distributed with mean 62.4 and standard deviation 11.5, and X_2 mimics sex and takes on 1 and 2 with probability 0.52 and 0.48 respectively. The environmental variable E is a Bernoulli random variable taking 1 with probability 0.5, independent of all the SNPs. We then used only the 13 typed SNPs to impute for all 62 SNPs in the ASAH1 gene using the program MaCH (Li *and others*, 2010) and the CEU HapMap reference panel. The imputed dataset consisting of both typed and imputed SNPs (in the form of dosage) is then used for testing. Again we restricted testing to common variants with MAF $\geqslant 0.05$ (giving $p = 40 - 47$). We varied β_1 from 0 to 0.6 in a step of 0.05 and evaluated the power for each of the 13 values of β_1 using 500 datasets (giving a total of $500 \times 13 = 6500$ datasets), each with $n = 2000$ (1000 cases, 1000 controls). The power results for both GESAT and *min* test are shown in Supplementary Figure 2. We note that this is a scenario optimized for the *min* test since there is only a single causal SNP that is genotyped. When the effect size is modest, GESAT performs better than the *min* test, but when the effect size is strong, the *min* test performs better than GESAT.

C.5. Comparing GESAT and SKAT

In Section 4.1 of the main manuscript, we discussed the motivation for using ridge regression to estimate the null model. In particular, as the number of SNPs p in a set is likely to be large and some SNPs might be in high LD with each other, the regular MLE might not be stable or difficult to calculate, and hence we have chosen to apply a L_2 penalty on the main effects of the SNPs to estimate the null model. If there is no penalty on the main effects of the SNPs (i.e. tuning parameter $\lambda = 0$), our test is equivalent to using SNP-set Kernel Association Test (SKAT) with the linear kernel to test for the SNP-Environment effects while incorporating the main SNP effects and other non-genetic variables as covariates (Wu *and others*, 2010, 2011).

In this section, we compare GESAT (penalize main effects) with SKAT (do not penalize the main SNP effects and use a variance component test to test GE interaction effects). We report simulation results for SKAT for the scenarios presented in Table 2 and Figure 2 of the main manuscript. SKAT is implemented by using the SKAT R package (www.hsph.harvard.edu/research/skat). In cases where the SNPs in the SNP-set are in low LD and the number of SNPs in the set is not large, using SKAT to test for GE interaction effects performs well. However for SNP-sets with SNPs in high LD, SKAT can give unstable results or the model may not converge. This presents an additional challenge since in practice when one scans the genome, it is difficult to know whether or not a model is stable unless each SNP-set is checked individually. For example, for the 15q24-25.1 region which has moderate LD, we report in Supplementary Table 7 the number of simulations (out of 10^5) for Type 1 error simulations and the number of simulations (out of 6500) for power simulations that did not converge for SKAT. When SKAT converges, both GESAT and SKAT have similar performance both in terms of the Type 1 error rate (Supplementary Table 8) and power (Supplementary Figure 3).

C.6. Additional Discussion on Figure 3 of main manuscript

In this section we provide a more detailed discussion on Figure 3 of the main manuscript. The power of GESAT for the two scenarios ((a) $\rho_1 = \rho_2$ and (b) $\rho_1 = -\rho_2$) described in Section 5.2 of the main manuscript are plotted in the bottom panel of Figure 3 in the main manuscript. For each of the two scenarios, we used three different values of $\rho_1 = 0, 0.5, 1$ and varied $\beta_1 = \beta_2$ from 0 to 0.6 in steps of 0.05. In both scenarios, the environmental factor E is binary and is associated with the SNPs (see main manuscript Section 5.2 for description on how E is generated). When the environmental factor E is binary, one of the factors affecting power is the proportion of samples with $E = 1$. We expect the power to be maximized when $\mathcal{E}(E) = 0.5$ In Supplementary Table 9, we report the empirical value of $\mathcal{E}(E)$ for each scenario and different values of $\rho_1 = 0, 0.5, 1$, obtained by averaging over different values of β_1. When $\rho_1 = \rho_2$, as ρ_1 increases from 0 to 1, power decreases (bottom left panel of main manuscript Figure 3). This is because as ρ_1 increases from 0 to 1, $\mathcal{E}(E)$ moves away from 0.5 (first row of Supplementary Table 9). When $\rho_1 = -\rho_2$, as ρ_1 increases from 0 to 1, power stays approximately constant (bottom right panel of main manuscript Figure 3). This is because even though ρ_1 increases, $\mathcal{E}(E)$ remains almost constant (second row of Supplementary Table 9). Thus the power of GESAT seems fairly robust to the association between $\boldsymbol{G}$ and E.

We also report empirical Type 1 error and power rates in Supplementary Figure 4 for two additional scenarios: (c) $\rho_1 = 2\rho_2$ and (d) $\rho_1 = -2\rho_2$. All other simulation details are identical to those used for (a) $\rho_1 = \rho_2$ and (b) $\rho_1 = -\rho_2$ described in Section 5.2 of the main manuscript. The results obtained for (c) and (d) are similar to those for (a) and (b).

C.7. Simulations for Continuous Trait comparing GESAT and SIMreg

In this section, we report simulation results for the case where the outcome is continuous, comparing GESAT to similarity regression (SIMreg), the method proposed by Tzeng *and others* (2011). The 15q24-25.1 region is used and the causal and tested SNPs are the same as that described in Supplementary Section C.2. We generated a continuous outcome assuming a linear regression model

$$Y_i = 0.2X_{1i} - 0.4X_{2i} + 0.2E_i + \alpha_{\mathrm{SNP1}}\mathrm{SNP1}_i + \alpha_{\mathrm{SNP2}}\mathrm{SNP2}_i + \beta_1\mathrm{SNP1}_i \times E_i + \beta_2\mathrm{SNP2}_i \times E_i + \epsilon_i$$

where X_1, X_2, E, ϵ are standard normal random variables. For each dataset, SNP1 and SNP2 are randomly selected from the group of 5 candidate causal SNPs described in Supplementary Section C.2., independent of E. We calculated GESAT and SIMreg using X_1, X_2, E and the 26 typed SNPs. Each dataset had a sample size of $n = 1000$. We considered two distinct scenarios: (a) $\alpha_{\mathrm{SNP1}} = \alpha_{\mathrm{SNP2}} = 0$ and (b) $\alpha_{\mathrm{SNP1}} = \alpha_{\mathrm{SNP2}} = 0.15$. The empirical Type 1 error and power are evaluated using 5000 and 500 simulations respectively. To investigate the Type 1 error rate, we set $\beta_1 = \beta_2 = 0$; and to investigate the power we increased $\beta_1 = \beta_2$ from 0 to 0.25 in steps of 0.01.

Supplementary Table 10 and Supplementary Figure 5 give the empirical Type 1 error rates and power curves respectively. The results show that both GESAT and SIMreg have similar performance, which is not unexpected since both conduct a variance component score test on the GE interaction terms. However in the simulations, we find that GESAT is much faster than SIMreg (Supplementary Table 11).

D. Harvard Lung Cancer Genetic Study

Our dataset discussed in Section 6 of the main manuscript comprises of 26 typed SNPs in the 15q24-25.1 region from 76593078 bp to 76740642 bp. The LD structure of the 26 typed SNPs based on our study samples is shown in the bottom panel of Supplementary Figure 6, indicating that the LD in this region is moderate and that the LD pattern observed in our study is consistent with that in the HapMap CEU population which we used in simulations. Lung cancer case/control status, age, sex and smoking status (never vs. ever) of the subjects are also available. A person is defined as a never smoker if he/she has smoked fewer than 100 cigarettes in his lifetime. After quality control filtering of the SNPs, there are a total of 1954 samples, including 984 cases and 970 controls. Of the 1954 samples, 1941 samples (980 cases, 961 controls) who had non-missing genotypes in all 26 typed SNPs are used in the analysis. There are 92 never smokers of the 980 cases; and 159 never smokers of the 961 controls.

To illustrate that our data generates similar findings about individual SNP effects as the published GWAS studies, we fit logistic regression models to each of the 26 SNPs individually adjusting for age, sex, smoking status, and four principal components, which are used to control for population stratification (Price *and others*, 2006). The top panel of Supplementary Figure 6 illustrates the p-values of the 26 SNPs with their locations on chromosome 15. The three most significant SNPs are rs1051730/rs8034191 and rs578776, which have already been identified in previous GWAS studies (Hung *and others*, 2008).

We conducted additional analysis to test for GE interactions using only these top SNPs. Adjusting for age, sex, smoking status, four principal components and the SNP main effects of rs1051730, rs8034191, rs578776, the three DF likelihood ratio test of interactions between rs1051730, rs8034191, rs578776 and smoking had a p-value of 0.0526. Since rs1051730 and rs8034191 are in high LD with each other ($R^2 = 0.85$ in our dataset), we also conducted an analysis adjusting for age, sex, smoking status, four principal components and the SNP main effects of rs1051730 and rs578776: the 2 DF likelihood ratio test of interactions between rs1051730, rs578776 and smoking had a p-value of 0.0333. These results are similiar to those obtained by using all the 26 SNPs with GESAT, suggesting similar power without the need to restrict to the GWAS-validated SNPs.

Lastly, Supplementary Table 12 gives results from the *min* test using all 26 SNPs in the region. The smallest p-value for SNP-smoking interaction term out of the 26 tests is 0.0103 achieved by rs1051730. After correcting for multiple comparisons using the estimated 16 effective tests (Gao *and others*, 2008) and the Bonferroni method, the *min* test had a p-value of $0.0103 \times 16 = 0.165$, which is not significant.

References

Barrett, JC, Fry, B., Maller, J. and Daly, MJ. (2005). Haploview: analysis and visualization of ld and haplotype maps. *Bioinformatics* **21**(2), 263–265.

Gao, X., Starmer, J. and Martin, E.R. (2008). A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genetic Epidemiology* **32**(4), 361–369.

Gibbs, R.A., Belmont, J.W., Hardenbol, P., Willis, T.D., Yu, F., Yang, H., Ch'ang, L.Y., Huang, W., Liu, B., Shen, Y. *and others*. (2003). The international hapmap project. *Nature* **426**(6968), 789–796.

Hung, R.J., McKay, J.D., Gaborieau, V., Boffetta, P., Hashibe, M., Zaridze, D., Mukeria, A., Szeszenia-Dabrowska, N., Lissowska, J., Rudnai, P. *and others*. (2008). A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature* **452**(7187), 633–637.

Knight, K. and Fu, W. (2000). Asymptotics for lasso-type estimators. *The Annals of Statistics* **28**(5), 1356–1378.

Li, Y., Willer, C.J., Ding, J., Scheet, P. and Abecasis, G.R. (2010). Mach: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genetic epidemiology* **34**(8), 816–834.

O'Sullivan, F., Yandell, BS and Raynor Jr, WJ. (1986). Automatic smoothing of regression functions in generalized linear models. *Journal of the American Statistical Association* **81**(393), 96–103.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and

Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics* **38**(8), 904–909.

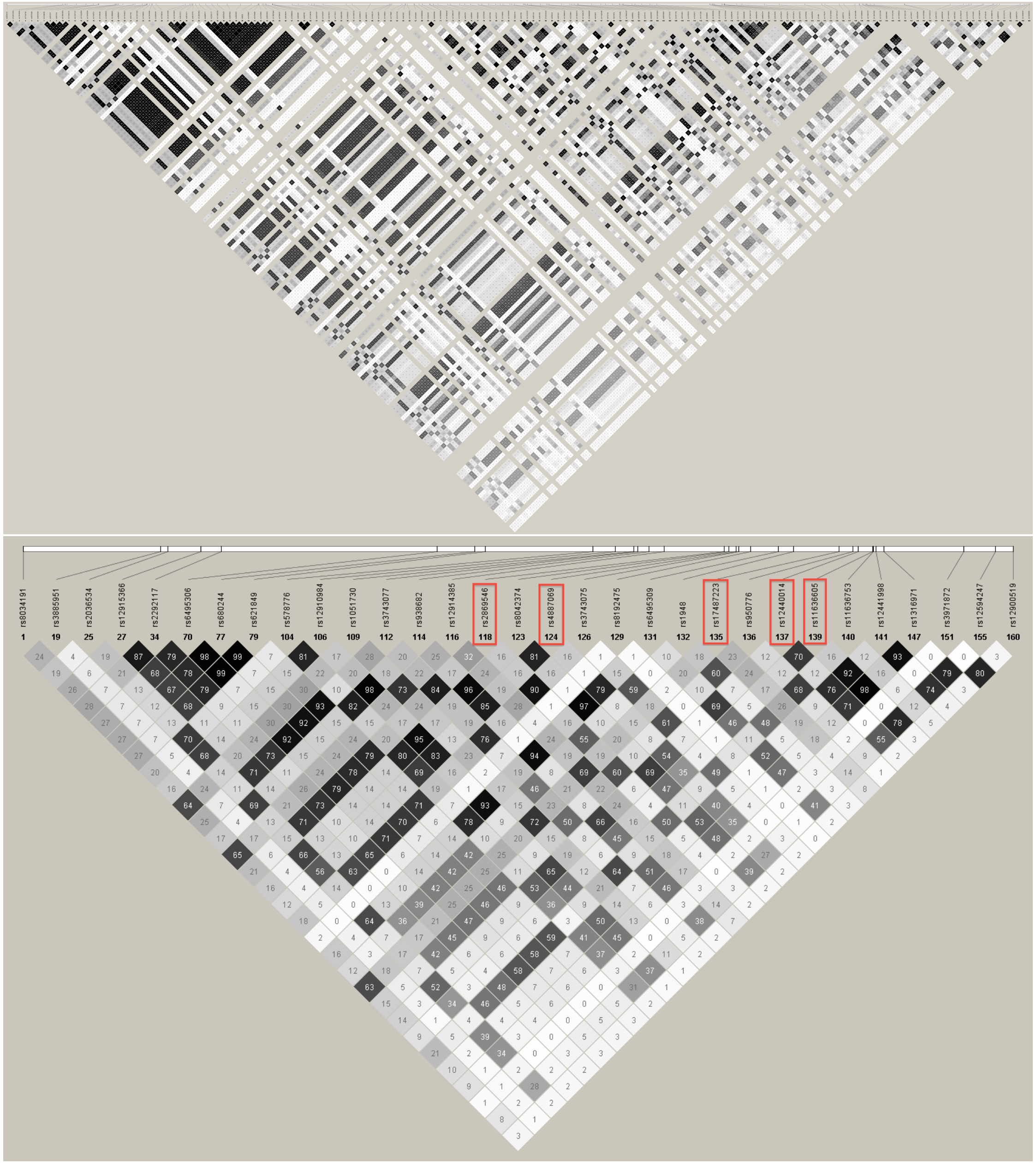
Thorisson, G.A., Smith, A.V., Krishnan, L. and Stein, L.D. (2005). The international hapmap project web site. *Genome research* **15**(11), 1592–1593.

Tzeng, J.Y., Zhang, D., Pongpanich, M., Smith, C., McCarthy, M.I., Sale, M.M., Worrall, B.B., Hsu, F.C., Thomas, D.C. and Sullivan, P.F. (2011). Studying gene and gene-environment effects of uncommon and common variants on continuous traits: a marker-set approach using gene-trait similarity regression. *The American Journal of Human Genetics* **89**(2), 277–288.

Wu, M.C., Kraft, P., Epstein, M.P., Taylor, D.M., Chanock, S.J., Hunter, D.J. and Lin, X. (2010). Powerful SNP-set analysis for case-control genome-wide association studies. *The American Journal of Human Genetics* **86**(6), 929–942.

Wu, M.C., Lee, S., Cai, T., Li, Y., Boehnke, M. and Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. *The American Journal of Human Genetics* **89**(1), 82–93.

Supplementary Figure 1: Linkage disequilibrium (LD), measured using R^2, of SNPs used in simulations from the HapMap CEU population. Top panel: LD for all 166 HapMap SNPs in the 15q24-25.1 region. Bottom panel: LD of the 26 typed common SNPs (SNPs on Illumina 610-Quad array) and a group of 5 candidate causal SNPs (indicated by rectangular box) used in simulating the data. In simulations, causal SNPs are chosen from the group of 5 candidate causal SNPs. The group of 5 candidate causal SNPs is identified using Haploview such that the effect of each SNP in the candidate causal group is partially captured by at least one of the 26 typed SNPs.



Supplementary Table 1: Empirical Type 1 error rates when $\boldsymbol{G}$ and $\boldsymbol{E}$ are independent - Type 1 error rate for GESAT is preserved, while that for *min* test can be slightly conservative. 15q24-25.1 region. Environmental Variable: Bernoulli w.p. 0.87. Description of simulation set-up are given in Section 5.1 of main manuscript and Supplementary Sections C.2. and C.3.

$\alpha_{\text{SNP 1}} = \alpha_{\text{SNP 2}} = 0$, $n = 2000$

	GESAT	*min* Test
5e-02	5.048e-02	3.088e-02
5e-03	5.460e-03	3.270e-03
5e-04	3.800e-04	2.000e-04

$\alpha_{\text{SNP 1}} = 0.4, \alpha_{\text{SNP 2}} = 0$, $n = 2000$

	GESAT	*min* Test
5e-02	5.123e-02	3.147e-02
5e-03	5.090e-03	3.080e-03
5e-04	6.600e-04	3.000e-04

$\alpha_{\text{SNP 1}} = \alpha_{\text{SNP 2}} = 0.4$, $n = 2000$

	GESAT	min Test
5e-02	5.220e-02	3.234e-02
5e-03	5.240e-03	2.970e-03
5e-04	5.100e-04	2.800e-04

$\alpha_{\text{SNP 1}} = \alpha_{\text{SNP 2}} = 0$, $n = 4000$

	GESAT	min Test
5e-02	5.222e-02	3.386e-02
5e-03	5.130e-03	3.880e-03
5e-04	4.800e-04	3.400e-04

$\alpha_{\text{SNP 1}} = 0.4, \alpha_{\text{SNP 2}} = 0$, $n = 4000$

	GESAT	*min* Test
5e-02	5.016e-02	3.333e-02
5e-03	5.220e-03	3.800e-03
5e-04	5.400e-04	4.500e-04

$\alpha_{\text{SNP 1}} = \alpha_{\text{SNP 2}} = 0.4$, $n = 4000$

	GESAT	min Test
5e-02	5.119e-02	3.386e-02
5e-03	5.250e-03	3.620e-03
5e-04	5.000e-04	3.400e-04

Supplementary Table 2: Empirical Type 1 error rates when $\boldsymbol{G}$ and $\boldsymbol{E}$ are independent - Type 1 error rate for GESAT is preserved, while that for *min* test can be slightly conservative. 15q24-25.1 region. Environmental Variable: Bernoulli w.p. 0.5. Description of simulation set-up are given in Section 5.1 of main manuscript and Supplementary Sections C.2. and C.3.

$\alpha_{\text{SNP 1}} = \alpha_{\text{SNP 2}} = 0$, $n = 2000$

	GESAT	*min* Test
5e-02	5.117e-02	3.582e-02
5e-03	5.240e-03	4.290e-03
5e-04	5.100e-04	5.900e-04

$\alpha_{\text{SNP 1}} = 0.4, \alpha_{\text{SNP 2}} = 0$, $n = 2000$

	GESAT	*min* Test
5e-02	5.203e-02	3.633e-02
5e-03	4.890e-03	4.030e-03
5e-04	5.100e-04	4.500e-04

$\alpha_{\text{SNP 1}} = \alpha_{\text{SNP 2}} = 0.4$, $n = 2000$

	GESAT	min Test
5e-02	5.153e-02	3.831e-02
5e-03	5.490e-03	4.240e-03
5e-04	5.600e-04	3.900e-04

$\alpha_{\text{SNP 1}} = \alpha_{\text{SNP 2}} = 0$, $n = 4000$

	GESAT	min Test
5e-02	5.105e-02	3.645e-02
5e-03	5.140e-03	4.430e-03
5e-04	5.500e-04	5.100e-04

$\alpha_{\text{SNP 1}} = 0.4, \alpha_{\text{SNP 2}} = 0$, $n = 4000$

	GESAT	*min* Test
5e-02	5.160e-02	3.702e-02
5e-03	5.200e-03	4.370e-03
5e-04	5.200e-04	4.400e-04

$\alpha_{\text{SNP 1}} = \alpha_{\text{SNP 2}} = 0.4$, $n = 4000$

	GESAT	min Test
5e-02	5.150e-02	3.770e-02
5e-03	5.020e-03	4.610e-03
5e-04	4.900e-04	4.300e-04

Supplementary Table 3: Empirical Type 1 error rates when $\boldsymbol{G}$ and $\boldsymbol{E}$ are independent - Type 1 error rate for GESAT is preserved, while that for *min* test can be slightly conservative. 15q24-25.1 region. Environmental Variable: standard normal random variable. Description of simulation set-up are given in Section 5.1 of main manuscript and Supplementary Sections C.2. and C.3.

$\alpha_{\text{SNP 1}} = \alpha_{\text{SNP 2}} = 0$, $n = 2000$

	GESAT	*min* Test
5e-02	5.179e-02	3.526e-02
5e-03	5.200e-03	4.010e-03
5e-04	5.600e-04	3.900e-04

$\alpha_{\text{SNP 1}} = 0.4, \alpha_{\text{SNP 2}} = 0$, $n = 2000$

	GESAT	*min* Test
5e-02	5.112e-02	3.487e-02
5e-03	4.980e-03	3.510e-03
5e-04	5.500e-04	4.200e-04

$\alpha_{\text{SNP 1}} = \alpha_{\text{SNP 2}} = 0.4$, $n = 2000$

	GESAT	min Test
5e-02	5.108e-02	3.576e-02
5e-03	5.060e-03	3.700e-03
5e-04	5.200e-04	3.000e-04

$\alpha_{\text{SNP 1}} = \alpha_{\text{SNP 2}} = 0$, $n = 4000$

	GESAT	min Test
5e-02	5.161e-02	3.595e-02
5e-03	5.290e-03	4.430e-03
5e-04	5.300e-04	4.200e-04

$\alpha_{\text{SNP 1}} = 0.4, \alpha_{\text{SNP 2}} = 0$, $n = 4000$

	GESAT	*min* Test
5e-02	5.069e-02	3.546e-02
5e-03	5.160e-03	4.170e-03
5e-04	4.600e-04	3.400e-04

$\alpha_{\text{SNP 1}} = \alpha_{\text{SNP 2}} = 0.4$, $n = 4000$

	GESAT	min Test
5e-02	5.088e-02	3.640e-02
5e-03	5.110e-03	4.180e-03
5e-04	5.300e-04	4.500e-04

Supplementary Table 4: Empirical Type 1 error rates when $\boldsymbol{G}$ and $\boldsymbol{E}$ are independent - Type 1 error rate for GESAT is preserved, while that for *min* test can be slightly conservative. ASAH1 gene. Environmental Variable: Bernoulli w.p. 0.87. Description of simulation set-up are given in Section 5.1 of main manuscript and Supplementary Section C.3.

$\alpha_{\text{SNP 1}} = \alpha_{\text{SNP 2}} = 0$, $n = 2000$

	GESAT	*min* Test
5e-02	5.205e-02	3.458e-02
5e-03	5.030e-03	3.010e-03
5e-04	4.000e-04	2.900e-04

$\alpha_{\text{SNP 1}} = 0.4, \alpha_{\text{SNP 2}} = 0$, $n = 2000$

	GESAT	*min* Test
5e-02	5.069e-02	3.455e-02
5e-03	4.920e-03	3.180e-03
5e-04	4.400e-04	2.400e-04

$\alpha_{\text{SNP 1}} = \alpha_{\text{SNP 2}} = 0.4$, $n = 2000$

	GESAT	*min* Test
5e-02	5.164e-02	3.519e-02
5e-03	5.320e-03	3.390e-03
5e-04	4.500e-04	3.000e-04

$\alpha_{\text{SNP 1}} = \alpha_{\text{SNP 2}} = 0$, $n = 4000$

	GESAT	*min* Test
5e-02	5.108e-02	3.898e-02
5e-03	5.290e-03	4.360e-03
5e-04	5.800e-04	5.000e-04

$\alpha_{\text{SNP 1}} = 0.4, \alpha_{\text{SNP 2}} = 0$, $n = 4000$

	GESAT	*min* Test
5e-02	5.176e-02	3.821e-02
5e-03	5.220e-03	4.050e-03
5e-04	6.000e-04	3.200e-04

$\alpha_{\text{SNP 1}} = \alpha_{\text{SNP 2}} = 0.4$, $n = 4000$

	GESAT	*min* Test
5e-02	5.082e-02	3.958e-02
5e-03	5.410e-03	4.560e-03
5e-04	6.100e-04	3.500e-04

Supplementary Table 5: Empirical Type 1 error rates when $\boldsymbol{G}$ and $\boldsymbol{E}$ are independent - Type 1 error rate for GESAT is preserved, while that for *min* test can be slightly conservative. ASAH1 gene. Environmental Variable: Bernoulli w.p. 0.5. Description of simulation set-up are given in Section 5.1 of main manuscript and Supplementary Section C.3.

$\alpha_{\text{SNP 1}} = \alpha_{\text{SNP 2}} = 0$, $n = 2000$

	GESAT	*min* Test
5e-02	5.147e-02	4.148e-02
5e-03	5.310e-03	4.550e-03
5e-04	4.200e-04	5.500e-04

$\alpha_{\text{SNP 1}} = 0.4, \alpha_{\text{SNP 2}} = 0$, $n = 2000$

	GESAT	*min* Test
5e-02	5.162e-02	4.105e-02
5e-03	5.220e-03	5.050e-03
5e-04	5.600e-04	5.700e-04

$\alpha_{\text{SNP 1}} = \alpha_{\text{SNP 2}} = 0.4$, $n = 2000$

	GESAT	*min* Test
5e-02	5.166e-02	4.241e-02
5e-03	5.400e-03	5.060e-03
5e-04	5.300e-04	5.300e-04

$\alpha_{\text{SNP 1}} = \alpha_{\text{SNP 2}} = 0$, $n = 4000$

	GESAT	*min* Test
5e-02	5.132e-02	4.161e-02
5e-03	4.760e-03	4.650e-03
5e-04	4.200e-04	5.200e-04

$\alpha_{\text{SNP 1}} = 0.4, \alpha_{\text{SNP 2}} = 0$, $n = 4000$

	GESAT	*min* Test
5e-02	4.889e-02	4.187e-02
5e-03	4.360e-03	4.670e-03
5e-04	4.400e-04	5.500e-04

$\alpha_{\text{SNP 1}} = \alpha_{\text{SNP 2}} = 0.4$, $n = 4000$

	GESAT	*min* Test
5e-02	5.027e-02	4.184e-02
5e-03	4.900e-03	4.610e-03
5e-04	4.300e-04	5.500e-04

Supplementary Table 6: Empirical Type 1 error rates when $\boldsymbol{G}$ and $\boldsymbol{E}$ are independent - Type 1 error rate for GESAT is preserved, while that for *min* test can be slightly conservative. ASAH1 gene. Environmental Variable: standard normal random variable. Description of simulation set-up are given in Section 5.1 of main manuscript and Supplementary Section C.3.

$\alpha_{\text{SNP 1}} = \alpha_{\text{SNP 2}} = 0$, $n = 2000$

	GESAT	*min* Test
5e-02	5.065e-02	3.921e-02
5e-03	4.930e-03	4.360e-03
5e-04	5.100e-04	4.400e-04

$\alpha_{\text{SNP 1}} = 0.4, \alpha_{\text{SNP 2}} = 0$, $n = 2000$

	GESAT	*min* Test
5e-02	5.086e-02	4.028e-02
5e-03	5.160e-03	4.380e-03
5e-04	6.700e-04	5.400e-04

$\alpha_{\text{SNP 1}} = \alpha_{\text{SNP 2}} = 0.4$, $n = 2000$

	GESAT	*min* Test
5e-02	5.017e-02	3.924e-02
5e-03	4.730e-03	4.000e-03
5e-04	5.100e-04	3.000e-04

$\alpha_{\text{SNP 1}} = \alpha_{\text{SNP 2}} = 0$, $n = 4000$

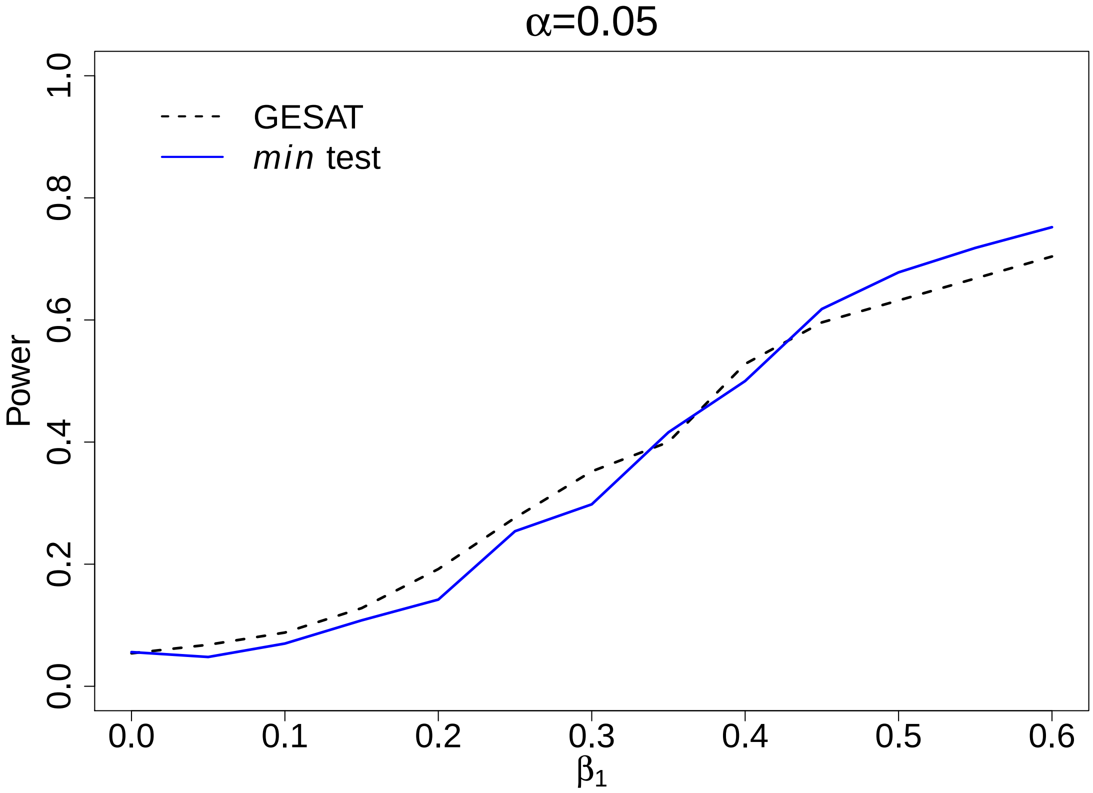
	GESAT	*min* Test
5e-02	4.994e-02	4.002e-02
5e-03	4.650e-03	4.540e-03
5e-04	4.100e-04	6.400e-04

$\alpha_{\text{SNP 1}} = 0.4, \alpha_{\text{SNP 2}} = 0$, $n = 4000$

	GESAT	*min* Test
5e-02	4.955e-02	4.027e-02
5e-03	4.950e-03	4.500e-03
5e-04	5.300e-04	4.000e-04

$\alpha_{\text{SNP 1}} = \alpha_{\text{SNP 2}} = 0.4$, $n = 4000$

	GESAT	*min* Test
5e-02	5.106e-02	4.117e-02
5e-03	5.070e-03	4.720e-03
5e-04	4.700e-04	4.100e-04



Supplementary Figure 2: Empirical power curves at $\alpha = 0.05$ level of significance for GESAT (dashed line) and *min* test (solid line) assuming $\boldsymbol{G}$ and $\boldsymbol{E}$ are independent for ASAH1 gene where the untyped SNPs are imputed and included in the testing procedure and there is only a single causal SNP that is typed. When the effect size is modest, GESAT performs better than the *min* test, but when the effect size is strong, the *min* test performs better than GESAT. Further details on simulation set-up are given in Supplementary Section C.4.

Supplementary Table 7: No. of simulations (out of 10^5) for Type 1 error simulations and No. of simulations (out of 6500) for power simulations that did not converge for SKAT for 15q24-25.1 region, a SNP-set with moderate LD. The Type 1 error simulations results for the remainder simulations that converged are reported in Supplementary Table 8. The power simulations results for the remainder simulations that converged are reported in Supplementary Figure 3. Further details on simulation set-up are given in Supplementary Section C.5.

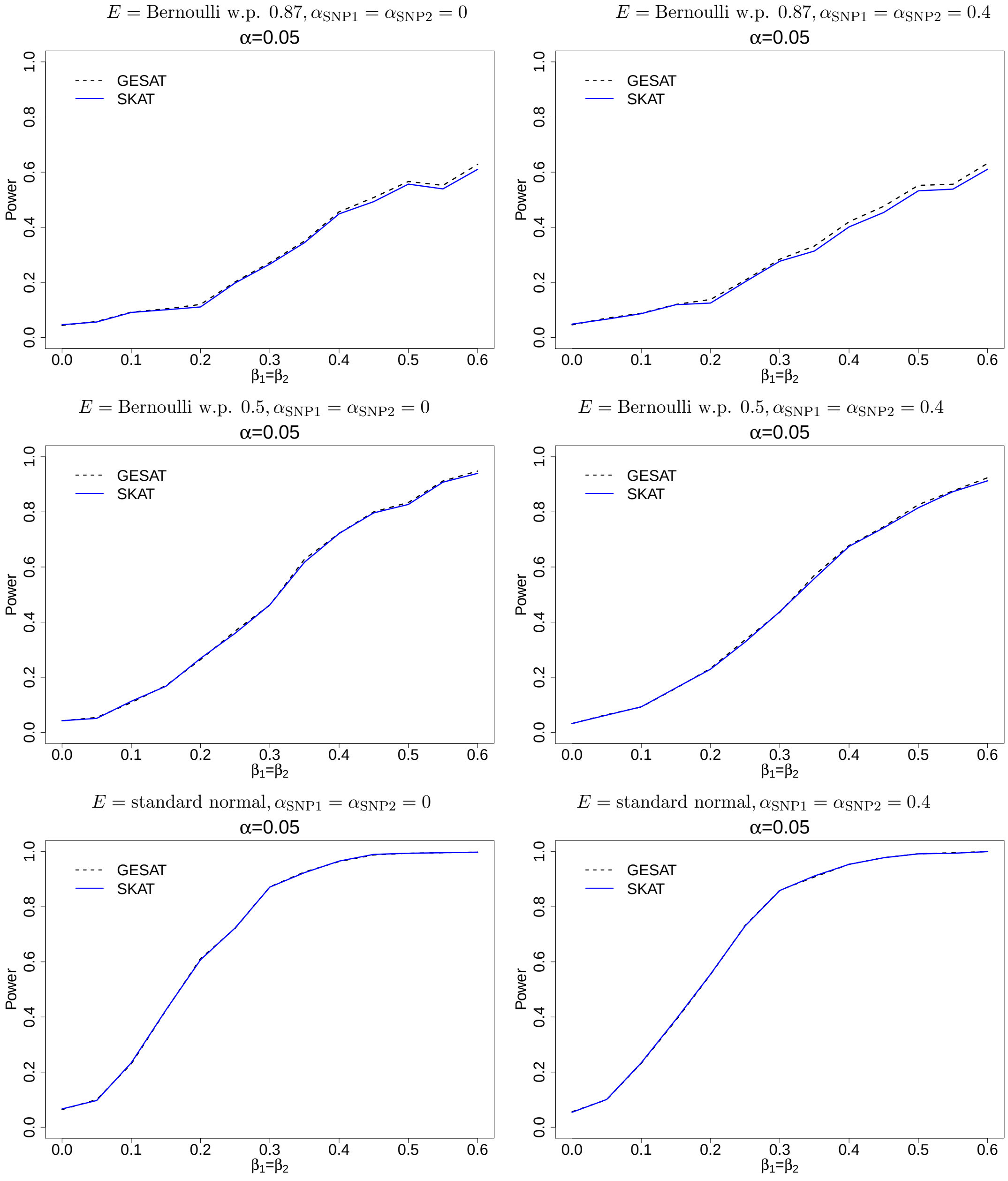
$\alpha_{\text{SNP1}}, \alpha_{\text{SNP2}}$	Environmental Variable	Type 1 error Simulations No. that did not converge (out of 10^5)	Power Simulations No. that did not converge (out of 6500)
0	Bernoulli w. prob 0.87	645	49
0	Bernoulli w. prob 0.5	615	41
0	Standard Normal	654	50
0.4	Bernoulli w. prob 0.87	662	62
0.4	Bernoulli w. prob 0.5	647	46
0.4	Standard Normal	752	42

Supplementary Table 8: Empirical Type 1 error rates for both GESAT and SKAT test at 0.05 level when $\boldsymbol{G}$ and $\boldsymbol{E}$ are independent. The results indicate that the Type 1 error rates are protected for both methods. Note that identical Type 1 error rates for GESAT are also reported in Table 2 of the main manuscript. Type 1 error rates for SKAT are calculated using only simulations that converged. Further details on simulation set-up are given in Supplementary Section C.5.

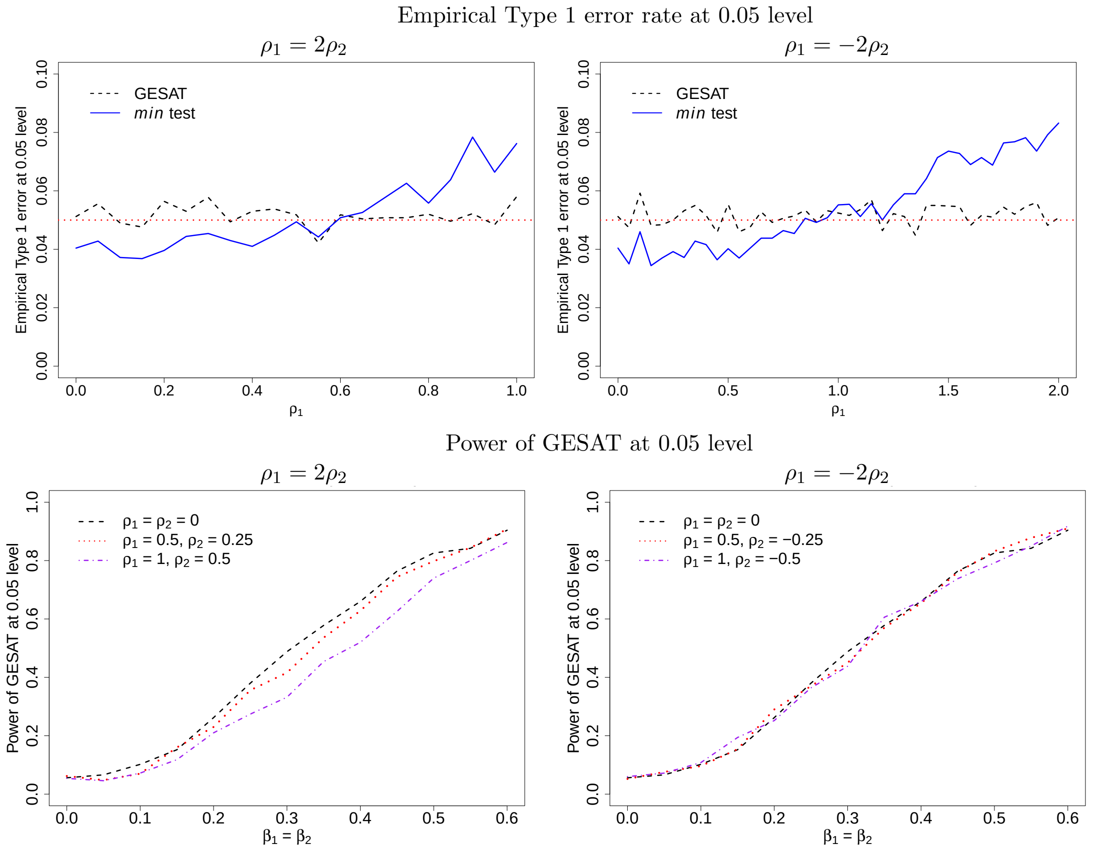
$\alpha_{\text{SNP1}}, \alpha_{\text{SNP2}}$	Environmental Variable	GESAT	SKAT
0	Bernoulli w. prob 0.87	5.05e-02	5.14e-02
0	Bernoulli w. prob 0.5	5.12e-02	5.17e-02
0	Standard Normal	5.18e-02	5.25e-02
0.4	Bernoulli w. prob 0.87	5.22e-02	5.23e-02
0.4	Bernoulli w. prob 0.5	5.15e-02	5.21e-02
0.4	Standard Normal	5.11e-02	5.19e-02

Supplementary Table 9: For each scenario and ρ_1, $\widehat{\mathcal{E}}(E)$ estimated by averaging over the different values of β_1 used. Further details on simulation set-up are given in Supplementary Section C.6.

	$\widehat{\mathcal{E}}(E)$ for $\rho_1 = 0$	$\widehat{\mathcal{E}}(E)$ for $\rho_1 = 0.5$	$\widehat{\mathcal{E}}(E)$ for $\rho_1 = 1$
Bottom left panel of Figure 3 ($\rho_1 = \rho_2$)	0.52	0.65	0.74
Bottom right panel of Figure 3 ($\rho_1 = -\rho_2$)	0.52	0.52	0.52
Bottom left panel of Supplementary Figure 4 ($\rho_1 = 2\rho_2$)	0.52	0.62	0.70
Bottom right panel of Supplementary Figure 4 ($\rho_1 = -2\rho_2$)	0.52	0.55	0.58



Supplementary Figure 3: GESAT and SKAT have similar power - Empirical power curves at $\alpha = 0.05$ level of significance for GESAT (dashed line) and SKAT (solid line) assuming $\boldsymbol{G}$ and $\boldsymbol{E}$ are independent. Top panel - Environmental factor is Bernoulli with probability 0.87; Middle panel - Environmental factor is Bernoulli with probability 0.5; Bottom panel - Environmental factor is standard normal. Left panel - SNPs have no main effect ($\alpha_{\mathrm{SNP1}} = \alpha_{\mathrm{SNP2}} = 0$); Right panel - SNPs have main effects ($\alpha_{\mathrm{SNP1}} = \alpha_{\mathrm{SNP2}} = 0.4$). Note that identical power for GESAT are also reported in Figure 2 of the main manuscript. Power for SKAT is calculated using only simulations that converged. Further details on simulation set-up are given in Supplementary Section C.5.



Supplementary Figure 4: Type 1 error of GESAT is robust to the dependence of $\boldsymbol{G}$ and $\boldsymbol{E}$ but *min* test can give inflated Type 1 error rate - Empirical Type 1 error rates at 0.05 level for GESAT (dashed line) and *min* test (solid line) when $\boldsymbol{G}$ and $\boldsymbol{E}$ are dependent, are given in the top panel. In the left panel, $\rho_1 = 2\rho_2$. In the right panel, $\rho_1 = -2\rho_2$. Power of GESAT is robust to association between $\boldsymbol{G}$ and $\boldsymbol{E}$ - Dashed, dotted, dashed-and-dotted lines give power of GESAT at 0.05 level when $\rho_1 = 0, 0.5, 1$ respectively in the bottom panel. The models for generating the data are given in Section 5.2 of main manuscript and Supplementary Section C.6. The parameters ρ_1, ρ_2 control the association between $\boldsymbol{G}$ and $\boldsymbol{E}$. Note that Figure 3 of the main manuscript gives results for $\rho_1 = \rho_2$ and $\rho_1 = -\rho_2$.

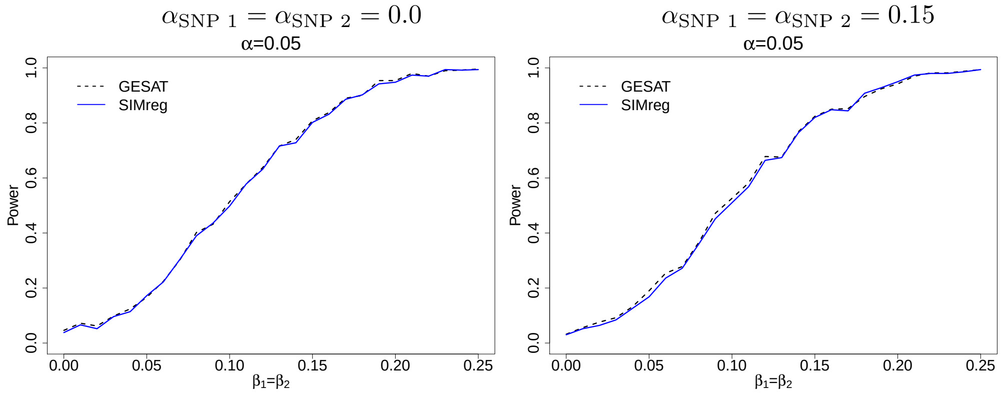
Supplementary Table 10: Empirical Type 1 error rates for both GESAT and SIMreg calculated using 5000 simulations at 0.05 level when the outcome is continuous. The results indicate that the Type 1 error rates are protected for both methods in this setting. Further details on simulation set-up are given in Supplementary Section C.7.

$\alpha_{\text{SNP 1}} = \alpha_{\text{SNP 2}} = 0$, $n = 1000$

	GESAT	SIMreg
5e-02	5.200e-02	4.740e-02
5e-03	4.200e-03	3.400e-03

$\alpha_{\text{SNP 1}} = \alpha_{\text{SNP 2}} = 0.15$, $n = 1000$

	GESAT	SIMreg
5e-02	5.260e-02	4.660e-02
5e-03	4.400e-03	3.200e-03

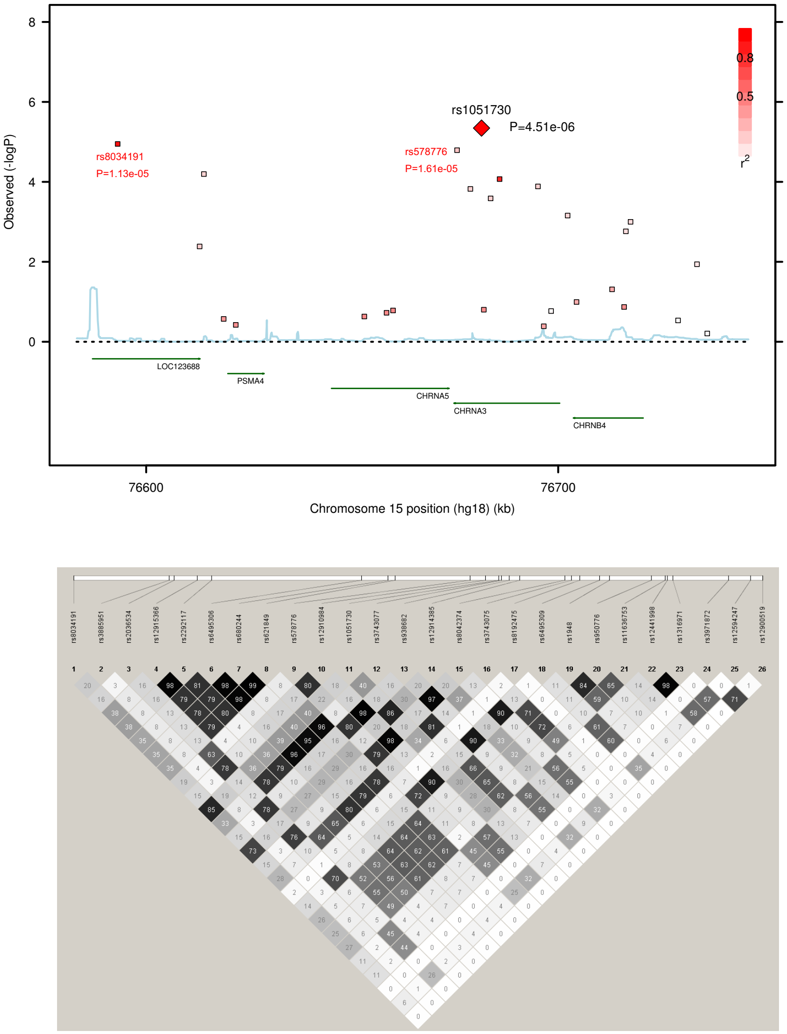


Supplementary Figure 5: GESAT and SIMreg have similar power - Empirical power curves at $\alpha = 0.05$ level for GESAT (dashed line) and SIMreg (solid line) when the outcome is continuous. Further details on simulation set-up are given in Supplementary Section C.7.

Supplementary Table 11: Average computational times for GESAT and SIMreg in seconds to test one SNP-set with $p = 26$ typed SNPs when the outcome is continuous. The computation times below are obtained from the Type 1 error simulations reported in Supplementary Table 10. Further details on simulation set-up are given in Supplementary Section C.7.

	GESAT	SIMreg
$\alpha_{\text{SNP 1}} = \alpha_{\text{SNP 2}} = 0.0$	0.0649	241
$\alpha_{\text{SNP 1}} = \alpha_{\text{SNP 2}} = 0.15$	0.0652	53.6

Supplementary Figure 6: Top panel: $-\log$(p-value) of the SNP main effect from a main effect logistic regression model fitted to each SNP individually for the 26 SNPs in the Harvard lung cancer data example. Each model adjusted for age, sex, smoking status and four principal components. The three SNPs with the smallest p-values (rs1051730, rs8034191, rs578776) have all been identified to be associated with lung cancer in previous studies. Bottom panel: Linkage disequilibrium (LD), measured using R^2, among the 26 typed SNPs used in the data example (LD based on 1941 study samples in the data example). The LD in this region is moderate. Further details are given in Supplementary Section D.



Supplementary Table 12: P-values for the SNP-smoking interaction term from a logistic regression GE interaction model fitted to each SNP individually for the 26 SNPs in data example. Each model had nine terms: age, sex, smoking status, four principal components, SNP and SNP-smoking interaction term. Of these 26 SNPs, rs1051730 gives the smallest SNP-smoking p-value of 0.01030. After correcting for multiple comparisons using estimated 16 effective degrees of freedom (Gao *and others*, 2008) and the Bonferroni method, the *min* test has a p-value of $0.0103 \times 16 = 0.165$. Further details are given in Supplementary Section D.

SNP	P-value for SNP-smoking interaction term
rs8034191	0.05373
rs3885951	0.03430
rs2036534	0.05927
rs12915366	0.6880
rs2292117	0.9086
rs6495306	0.2076
rs680244	0.1794
rs621849	0.3155
rs578776	0.1804
rs12910984	0.1346
rs1051730	0.01030
rs3743077	0.3019
rs938682	0.1478
rs12914385	0.01592
rs8042374	0.06443
rs3743075	0.04813
rs8192475	0.3892
rs6495309	0.2937
rs1948	0.03979
rs950776	0.1176
rs11636753	0.02368
rs12441998	0.4133
rs1316971	0.3767
rs3971872	0.3807
rs12594247	0.8052
rs12900519	0.5619