


Polygenic approaches to detect gene–environment interactions when external information is unavailable

Wan-Yu Lin , Ching-Chieh Huang, Yu-Li Liu, Shih-Jen Tsai and Po-Hsiu Kuo

Corresponding author: Wan-Yu Lin, Institute of Epidemiology and Preventive Medicine & Department of Public Health, College of Public Health, National Taiwan University, Room 501, No. 17, Xu-Zhou Road, Taipei 100, Taiwan. Tel.: +886-2-33668106; Fax: +886-2-33668106; E-mail: linwy@ntu.edu.tw

Abstract

The exploration of ‘gene–environment interactions’ ($G \times E$) is important for disease prediction and prevention. The scientific community usually uses external information to construct a genetic risk score (GRS), and then tests the interaction between this GRS and an environmental factor (E). However, external genome-wide association studies (GWAS) are not always available, especially for non-Caucasian ethnicity. Although GRS is an analysis tool to detect $G \times E$ in GWAS, its performance remains unclear when there is no external information. Our ‘adaptive combination of Bayes factors method’ (ADABF) can aggregate $G \times E$ signals and test the significance of $G \times E$ by a polygenic test. We here explore a powerful polygenic approach for $G \times E$ when external information is unavailable, by comparing our ADABF with the GRS based on marginal effects of SNPs (GRS-M) and GRS based on $SNP \times E$ interactions (GRS-I). ADABF is the most powerful method in the absence of SNP main effects, whereas GRS-M is generally the best test when single-nucleotide polymorphisms main effects exist. GRS-I is the least powerful test due to its data-splitting strategy. Furthermore, we apply these methods to Taiwan Biobank data. ADABF and GRS-M identified gene \times alcohol and gene \times smoking interactions on blood pressure (BP). BP-increasing alleles elevate more BP in drinkers (smokers) than in nondrinkers (nonsmokers). This work provides guidance to choose a polygenic approach to detect $G \times E$ when external information is unavailable.

Key words: diastolic blood pressure; systolic blood pressure; gene–alcohol interaction; gene–smoking interaction; Taiwan Biobank; multiple-testing correction

Introduction

Evidence of ‘gene–environment interactions’ ($G \times E$) has been found in some phenotypes and complex diseases [1–5]. Genes and environmental exposure may jointly influence disease liability. The identification of $G \times E$ is important to improve the accuracy and precision of assessing genetic and environmental influences [6]. The $G \times E$ identification is an active research area that generates high expectations, but usually leads to

great disappointment [7]. With the shift toward genome-wide association studies (GWAS), genome-wide $G \times E$ studies are fairly common in human genetic research [8]. However, few significant and replicated $G \times E$ have been found from GWAS to date [9–12]. Popular choices for genome-wide $G \times E$ analyses include single marker analysis and set-based (or gene-based) analysis [13–16]. However, with the large number of single-nucleotide

Wan-Yu Lin is an associate professor at the Institute of Epidemiology and Preventive Medicine & Department of Public Health, College of Public Health, National Taiwan University.

Ching-Chieh Huang is a Masters student at the Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University. He is under Wan-Yu Lin’s supervision.

Yu-Li Liu is an investigator at the Center for Neuropsychiatric Research, National Health Research Institutes.

Shih-Jen Tsai is the Chairman at the Department of Psychiatry, Taipei Veterans General Hospital and a professor at the Division of Psychiatry, National Yang-Ming University.

Po-Hsiu Kuo is a professor at the Institute of Epidemiology and Preventive Medicine & Department of Public Health, College of Public Health, National Taiwan University.

Submitted: 5 June 2018; Received (in revised form): 14 August 2018

© The Author(s) 2018. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

polymorphisms (SNPs) or genes in the human genome, both approaches suffer from low statistical power due to a harsh penalty from multiple hypothesis correction.

Many G × E studies avoid the harsh multiple-testing penalty by constructing a genetic risk score (GRS) according to previous GWAS findings and then testing the significance of the interaction term (GRS × E) in a regression model [3–5, 17–19]. The GRS or polygenic risk score (PRS) is commonly used to summarize genetic effects among an ensemble of SNPs that do not individually achieve the genome-wide significance level, i.e. 5×10^{-8} [20–22]. Information on L disease-associated SNPs that are nearly in linkage equilibrium is aggregated by defining a GRS of the i^{th} subject as $GRS_i = \sum_{l=1}^L \hat{\beta}_l g_{il}$, where g_{il} is the number of risk alleles (those associated with disease liability according to previous studies) at the l^{th} SNP of that subject and $\hat{\beta}_l$ is the weight given to the l^{th} SNP ($l = 1, \dots, L; i = 1, \dots, n$, where n is the sample size).

If all the $\hat{\beta}_l$ s are specified to be 1, the GRS_i is the overall genetic burden of the i^{th} subject by summing his/her total number of risk alleles. For example, the gene × physical activity interaction in obesity was identified using an unweighted GRS [3, 4]. This unweighted GRS was composed of 12 body mass index (BMI) associated SNPs that were discovered by previous GWAS [23–27]. However, the unweighted GRS (all $\hat{\beta}_l$ s = 1) should be used only if the L disease-associated SNPs are considered of equal importance.

When a weighted GRS is considered, researchers often extract $\hat{\beta}_l$ s from external GWAS that focus on the same ethnicity [28]. Because GWAS usually fit a regression model for each SNP and provide summary statistics for the scientific community, $\hat{\beta}_l$ s come from the regression estimates regarding SNP effects. For a continuous trait, $\hat{\beta}_l$ ($l = 1, \dots, L$) is the effect size of the risk allele at the l^{th} SNP [5]. For a binary trait, $\hat{\beta}_l$ is the log odds ratio (OR) of the risk allele at the l^{th} SNP [19, 29–31]. For example, the interactions of gene × physical activity, gene × alcohol consumption and gene × socioeconomic status were detected using a weighted GRS [5] composed of 94 BMI-associated SNPs that were identified by a previous GWAS [32]. Mullins et al. found that childhood trauma has a greater effect in subjects with lower genetic liability for major depressive disorder [19], by constructing a weighted GRS using results from the Psychiatric Genomics Consortium [33, 34]. These G × E findings rely on previous GWAS discoveries. However, an appropriate external GWAS is not always available, especially for non-Caucasian ethnicity.

When there is no appropriate external information, the weights have to come from the current study. Hüls et al. have proposed a ‘GRS-marginal-internal approach’ and a ‘GRS-interaction-training approach’ for pathway-oriented G × E studies [28, 35]. That is, G × E interactions are investigated in a pre-selected set of candidate SNPs, e.g. SNPs from a biological pathway. The weights of these SNPs are then estimated by a multivariate elastic net regression [36]. However, take Taiwan Biobank (TWB) GWAS as an example, there are 601 531 autosomal SNPs passing the quality control stage. After pruning SNPs in high linkage disequilibrium (LD), we still have 143 574 SNPs. Fitting a multivariate elastic net regression on such a large number of SNPs is computationally infeasible. Therefore, in this work, we borrow the idea from Hüls et al. [28, 35] and list two GRS-based tests that can be implemented in GWAS.

The GRS weights can be determined in the following two ways: (1) based on SNP marginal effects (we call it ‘GRS-M’) and (2) based on SNP × E interaction effects (we call it ‘GRS-I’). The abovementioned [3–5, 19] and many other G × E studies [17, 18, 29–31, 37–39] were all applications of the

GRS-M approach, because their GRSs were constructed by SNPs with larger marginal effects on the phenotype. In most GRS-M applications, the weights $\hat{\beta}_l$ s ($l = 1, \dots, L$) required to build GRS came from external studies [3–5, 17–19, 29–31].

Another way to construct a GRS for detecting G × E is to use the weights from SNP × E interaction effects themselves, and we call it the GRS-I approach. To preserve the validity, GRS-I should be performed by splitting the whole sample into a training subset and a testing subset. The weights $\hat{\beta}_l$ s are decided by the regression coefficients of SNP × E interaction term using the training subset. Then, GRS-I is calculated for the testing subset, and the significance of GRS-I × E is assessed.

Although G × E have been found for several phenotypes, many interaction effects may remain hidden due to the lack of a powerful polygenic test. In this work, we use the ‘adaptive combination of Bayes factors method’ (ADABF) [40] to aggregate G × E signals and to test the significance of G × E by a polygenic test. More than GRS-M and GRS-I, if the ADABF test result is significant, we can further pinpoint individual SNPs that interact with E. We compare the performance of ADABF, GRS-M and GRS-I with extensive simulations. Regarding the ability to pinpoint individual SNP × E, we calculate the positive predictive value (PPV) and sensitivity of ADABF and compare these values with those of single marker analysis. We then apply these approaches to TWB data to explore gene × alcohol consumption and gene × smoking interactions on blood pressure (BP) levels.

Methods

Adaptive combination of Bayes factors method

A pruning stage:

There are a pruning stage and a screening stage prior to using ADABF. Both the two stages are also commonly used in GRS methods [12, 20, 22, 41]. We first prune SNPs in high LD to eliminate a large degree of redundancy in SNPs. Suppose we have a GWAS dataset called ‘TWBGWAS’, the PLINK command ‘plink -bfile TWBGWAS -chr 1-22 -indep 50 5 2’ is used to prune SNPs in high LD [42]. This command removes SNPs with a variance inflation factor (VIF) > 2 within a sliding window of size 50. The sliding window is shifted at each step of five SNPs. VIF is calculated by $(1 - R^2)^{-1}$, where R^2 is the multiple correlation coefficient when an SNP is regressed on all other SNPs simultaneously. A VIF of 1 indicates that $R^2 = 0$ and the SNP is completely independent of all other SNPs. According to the PLINK guideline of SNP pruning (<http://zzz.bwh.harvard.edu/plink/summary.shtml#prune>), a VIF between 1.5 and 2 should be used in practice.

A screening stage:

Moreover, to improve the statistical power of G × E tests, the remained SNPs are then screened according to their marginal associations with the phenotype. The generalized linear model (GLM) for the l^{th} SNP ($l = 1, \dots, L$) is described as follows:

$$g[E(Y_i)] = \beta_0 + \beta_{G_l} G_{il} + \beta'_{X_l} X_i, i = 1, \dots, n, \quad (1)$$

where $g[\cdot]$ is the link function; Y_i is the phenotype, G_{il} is the number of minor alleles at the l^{th} SNP (0, 1 or 2) and X_i is the vector of covariates of the i^{th} subject. In this screening stage, we test $H_0: \beta_{G_l} = 0$ versus $H_1: \beta_{G_l} \neq 0$ ($l = 1, \dots, L$). The SNPs passing the screening at the desired significance level ($P < 0.05$) are

then analyzed using ADABF. This screening stage that reduces the number of SNPs tested for interactions can substantially increase the power of genome-wide $G \times E$ studies [8, 12, 43–48].

Suppose that in a GWAS there are L autosomal SNPs retained after the pruning and screening stages. We assess the interaction between the l^{th} SNP ($l = 1, \dots, L$) and E by the following GLM:

$$g[E(Y_i)] = \beta_0 + \beta_{G_l}G_{il} + \beta_E E_i + \beta_{GE_l}G_{il}E_i + \beta'_X \mathbf{X}_i, i = 1, \dots, n; \quad (2)$$

where E_i is the environmental factor (E) of the i^{th} subject, and the other notations have been described under Equation (1). Let $\hat{\beta}_{GE_l}$ be the maximum likelihood estimate (MLE) of β_{GE_l} . According to the asymptotic normality of MLE, $\hat{\beta}_{GE_l}$ follows a normal distribution with a mean of β_{GE_l} and a variance of $V_{l,i}$, i.e. $\hat{\beta}_{GE_l} \sim N(\beta_{GE_l}, V_{l,i})$.

Sizes of interaction effects (β_{GE_l} s) depend heavily on the scale of an E . An E_i ranging from 0 to 1 and an E_i ranging from 0 to 100 should be linked to different prior distributions of β_{GE_l} s. To propose a prior that can be used in most situations, we first rescale E_i to range from 0 to 1. Therefore, $G_{il}E_i$ will be between 0 and 2, the same as G_{il} . In Equation (2), a binary E_i (e.g. smoking versus nonsmoking) is coded as 1 or 0 and a continuous E_i is first scaled to a range from 0 to 1. Let E_{\min} and E_{\max} be the minimum and maximum of a continuous E_i , where $i = 1, \dots, n$. The continuous E_i is scaled to be $E_i' = (E_i - E_{\min}) / (E_{\max} - E_{\min})$, where $i = 1, \dots, n$.

The Wellcome Trust Case Control Consortium (WTCCC) GWAS specified a normal distribution with a mean of 0 and a variance of $W = 0.2^2 = 0.04$ as the prior of SNP main effects, i.e. $\beta_{G_l} \sim N(0, W = 0.04)$. Because $G_{il}E_i$ has been scaled to range from 0 to 2 (the same as G_{il}), we may consider the appropriateness of using $N(0, W = 0.04)$ as the prior of β_{GE_l} s. Similarly with SNP main effects, most reported SNP $\times E$ interactions are of modest effect sizes that can be positive or negative [49–51]. $N(0, W = 0.04)$ may be a reasonable prior for β_{GE_l} s as well. In a binary trait analysis such as the WTCCC GWAS [52], this prior implies that we believe 95% of ORs range from $\exp(-2 \times 0.2) = 0.67$ to $\exp(2 \times 0.2) = 1.49$. For continuous traits, this prior implies that 95% of β_{GE_l} s range from -0.4 to 0.4 . We consider this prior suitable for a standardized continuous trait with a mean of 0 and an SD of 1. Therefore, for a continuous trait analysis, our R code (<http://homepage.ntu.edu.tw/~linwy/ADABFGEPoly.html>) standardizes the trait before implementing the ADABF method.

The prior variance, $W = 0.2^2 = 0.04$ is originally designed for SNP main effects. Empirical evidence has shown that SNP $\times E$ interaction effects are usually modest [49–51], and therefore this prior variance may be slightly large for β_{GE_l} s. However, a larger prior variance can reflect our uncertainty of the prior information [53]. If investigators believe that very few SNP $\times E$ interactions may exist in their own study, they can specify a smaller prior variance that provides more shrinkage toward zero and favors more coefficients to be zero [53].

To test whether the l^{th} SNP interacts with E , the hypothesis is $H_{0,l} : \beta_{GE_l} = 0$ versus $H_{1,l} : \beta_{GE_l} \neq 0$ ($l = 1, \dots, L$). The BF is described as follows [54, 55]:

$$BF_l = \frac{\Pr(\text{Data}|H_{1,l})}{\Pr(\text{Data}|H_{0,l})} = \sqrt{\frac{\hat{V}_l}{\hat{V}_l + W}} \exp\left(\frac{\hat{\beta}_{GE_l}^2 W}{2\hat{V}_l(\hat{V}_l + W)}\right), l = 1, \dots, L, \quad (3)$$

where $\hat{\beta}_{GE_l}$ and \hat{V}_l have been estimated from the GLM in Equation (2).

The hypothesis of interest in ADABF is $H_0 : \beta_{GE_1} = \dots = \beta_{GE_L} = 0$ (none of the L SNPs interact with E) versus H_1 : at least one $\beta_{GE_l} \neq 0$ for $l = 1, \dots, L$. ADABF tests the interaction between E and all the L SNPs, by combining Bayes factors (BFs) of the L SNP $\times E$ signals. After obtaining BF_1, \dots, BF_L from Equation (3), we sort these L BFs from largest to smallest and denote them as $BF_{(1)} \geq BF_{(2)} \geq \dots \geq BF_{(L)}$. We calculate a summary score that aggregates the leading k BFs, $S_k = \sum_{l=1}^k \log(BF_{(l)})$, where $k = 1, \dots, L$. We then perform B resampling replicates, e.g. $B = 1000$ in our simulation and $B = 10^5$ in the following real data analysis. In each resampling, we draw an L -length vector of $\hat{\beta}_{GE_l, H_0}$ containing the MLEs of β_{GE_l} s ($l = 1, \dots, L$) under $H_0 : \beta_{GE_1} = \dots = \beta_{GE_L} = 0$. With the typical large sample size of a GWAS, the asymptotic normality of MLE holds so that $\hat{\beta}_{GE_l}$ follows a normal distribution with a mean of 0 (under H_0). Therefore, $\hat{\beta}_{GE, H_0}$ follows the multivariate normal distribution $N(\mathbf{0}_{L \times 1}, \mathbf{V}_{L \times L})$. The $(i, j)^{\text{th}}$ element of the variance–covariance matrix $\mathbf{V}_{L \times L}$ is $R_{i,j} \sqrt{\hat{V}_i \hat{V}_j}$, where \hat{V}_i and \hat{V}_j are the estimated variances of $\hat{\beta}_{GE_i}$ and $\hat{\beta}_{GE_j}$, respectively. Because the correlation among association statistics can be well approximated by the correlation among genotypes [56], $R_{i,j}$ is the correlation coefficient of the genotypes at the i^{th} and j^{th} SNPs. $\mathbf{V}_{L \times L}$ incorporates the pairwise correlations among SNPs and therefore the pruning of SNPs in high LD is not necessary in ADABF (but is still recommended to reduce the computational burden).

Let the summary score from the b^{th} resampling be $S_k^{(b)}$, where $k = 1, \dots, L$ and $b = 1, \dots, B$. P -value is the probability of obtaining a statistic as extreme as or more extreme than the observed statistic under the null hypothesis. Therefore, the P -value of S_k is estimated by $\frac{1}{B} \sum_{b=1}^B I(S_k^{(b)} \geq S_k)$, where $k = 1, \dots, L$ and $I(S_k^{(b)} \geq S_k)$ is an indicator variable with an outcome of 1 if $S_k^{(b)} \geq S_k$ or 0 if otherwise. Likewise, we calculate the P -value of $S_k^{(b)}$ by $\frac{1}{B-1} \sum_{b' \neq b} I(S_k^{(b')} \geq S_k^{(b)})$, where $k = 1, \dots, L$ and $b = 1, \dots, B$. Denote the minimum P -value (across $k = 1, \dots, L$) of the observed sample by $\min P$ and its counterparts from the B resampling replicates by $\min P^{(b)}$, $b = 1, \dots, B$. Therefore, the P -value of the ADABF test is $\frac{1}{B} \sum_{b=1}^B I(\min P^{(b)} \leq \min P)$, which is compared with the usual nominal significance level of 0.05 (or 0.01). No multiple hypothesis correction is required because all the L SNPs are considered in an overall test. If the ADABF P -value is less than 0.05 (or 0.01), the null hypothesis of $H_0 : \beta_{GE_1} = \dots = \beta_{GE_L} = 0$ is rejected, and we conclude that at least one SNP interacts with E . Because a P -value < 0.05 (or 0.01) leads to the rejection of H_0 of no polygenic SNP $\times E$ interactions [57], 1000 resampling replicates ($B = 1000$) is sufficient for our ADABF. This is different from the gene-based ADABF test where P -values are compared with the genome-wide significance level of $2.5 \times 10^{-6} = 0.05/20,000$ ($\sim 20,000$ genes in the genome) [40]. Given a significant ADABF test, the subsequent step is to pinpoint which SNPs interact with E .

From the above resampling procedure, we can also obtain the resampling false discovery rate (FDR). In the b^{th} resampling replicate, we have $BF_1^{(b)}, \dots, BF_L^{(b)}$, for the L SNP $\times E$. Based on the B resamples, the average number of false positives in a resampling replicate (while claiming significance of the leading k SNP $\times E$) is estimated by $FP_{(k)} = \frac{1}{B} \sum_{b=1}^B \sum_{l=1}^L I(BF_l^{(b)} \geq BF_{(k)})$. The corresponding FDR is thus estimated as $FDR_{(k)} = \frac{1}{k} FP_{(k)}$, i.e. the number of false positives over the number of significant findings [58, 59]. We find the maximum k that satisfies $FDR_{(k)} < 5\%$ and the SNPs corresponding to the leading k BFs are suggested to have interactions with E . The R code of the

ADABF approach can be downloaded from <http://homepage.ntu.edu.tw/~linwy/ADABFGEPoly.html>. We also provide the pipeline from PLINK to ADABF, at <http://homepage.ntu.edu.tw/~linwy/ADABFGEPolyPLINK.html>.

The Gaussian (normal) prior is the most common choice for a prior distribution [52, 60]. Under this setting, Wakefield has derived a simple and convenient BF formula, as shown in Equation (3) [55]. It is hard to justify that the prior really follows a normal distribution. However, deviation from the Gaussian prior may not make much difference to ADABF results, because this method makes inference through the resampling procedure. For the observed data and for each of the resampling replicates, we obtain BFs according to the same prior. As shown by the following simulations, ADABF performs well although the true interaction effects (β_{GE_i}) never really come from a Gaussian distribution.

GRS based on marginal effects of SNPs

We compare ADABF with GRS-M and GRS-I. Regarding GRS-M, the phenotype is first regressed on each of the L SNPs, as shown by Equation (1). The regression coefficients ($\hat{\beta}_{G_i}$ s) of the SNPs that are more associated with the phenotype (P -value less than a certain threshold) are treated as the weights of the GRS. To be specific, the pre-scaled GRS-M of the i^{th} subject is defined as follows:

$$GRS_{Mi,t}^{pre} = \sum_{l=1}^L \hat{\beta}_{G_i} G_{il} I(P_{G_i} < P_t), i = 1, \dots, n; t = 1, \dots, 10, \quad (4)$$

where $\hat{\beta}_{G_i}$ is estimated by the GLM in Equation (1), G_{il} is the number of minor alleles at the l^{th} SNP of the i^{th} subject, $I(\cdot)$ is the indicator variable, P_{G_i} is the P -value of testing $H_0 : \beta_{G_i} = 0$ versus $H_1 : \beta_{G_i} \neq 0$ and P_t is the t^{th} P -value threshold. Most investigators use a P -value threshold to select a subset of SNPs for a GRS [20, 22, 57, 61–64]. We used 10 thresholds to explore the strength of GRS: 0.0001, 0.00025, 0.0005, 0.001, 0.0025, 0.005, 0.01, 0.025, 0.05 and 0.1.

For example, if the phenotype is binary ($Y = 1$ means diseased whereas $Y = 0$ indicates non-diseased) and the minor allele at the l^{th} SNP is a risk allele, $\hat{\beta}_{G_i}$ estimated by the GLM (i.e. a logistic regression) in Equation (1) will be positive. Subjects with more minor alleles at this SNP will get an increase in their $GRS_{Mi,t}^{pre}$. However, if the minor allele at the l^{th} SNP is a protective allele, $\hat{\beta}_{G_i}$ will be negative and subjects with more minor alleles at this SNP will get a decrease in their $GRS_{Mi,t}^{pre}$. Therefore, summing the information of the L SNPs, a larger $GRS_{Mi,t}^{pre}$ indicates a larger disease liability. $GRS_{Mi,t}^{pre}$ is then rescaled to calibrate the number of phenotype-increasing alleles [18]:

$$GRS_{Mi,t} = \frac{GRS_{Mi,t}^{pre} \times \text{number of available SNPs}}{\text{sum of } |\hat{\beta}_{G_i}| \text{ of available SNPs}}. \quad (5)$$

Denote $WA = \frac{GRS_{Mi,t}^{pre}}{\text{sum of } |\hat{\beta}_{G_i}| \text{ of available SNPs}}$ as the weighted average of the number of phenotype-increasing alleles. (For a set of values x_i s with nonnegative weights w_i s, the weighted average of this set is $\frac{\sum w_i x_i}{\sum w_i}$.) The weights here, $\hat{\beta}_{G_i}$ s, can be positive or negative. Nonetheless, positive and negative $\hat{\beta}_{G_i}$ s will not be cancelled out in the formula of $GRS_{Mi,t}^{pre}$. As explained in the previous paragraph, regardless of the sign

of each $\hat{\beta}_{G_i}$, a larger $GRS_{Mi,t}^{pre}$ indicates a larger disease liability. Therefore, the denominator of WA is the sum of the absolute values of $\hat{\beta}_{G_i}$ s.

Because $\hat{\beta}_{G_i}$ s can be positive or negative, $GRS_{Mi,t}^{pre}$ and WA can be positive or negative as well. The range (maximum–minimum) of WA is up to 2, i.e. the number of minor alleles at each SNP. (For a set of values x_i s $\in \{0, 1, 2\}$ with nonnegative weights w_i s, the weighted average of this set is $\frac{\sum w_i x_i}{\sum w_i}$, ranging from 0 to 2.) To reflect the number of phenotype-increasing alleles from multiple loci, WA is multiplied by the number of available SNPs, as shown in Equation (5) [18].

Given the t^{th} P -value threshold ($t = 1, \dots, 10$), we calculate $GRS_{Mi,t}$ for all the n subjects, fit the following GLM, and test $H_0 : \phi_{GE} = 0$ versus $H_1 : \phi_{GE} \neq 0$:

$$g[E(Y_i)] = \phi_0 + \phi_G GRS_{Mi,t} + \phi_E E_i + \phi_{GE} GRS_{Mi,t} \cdot E_i + \phi'_X X_i, i = 1, \dots, n. \quad (6)$$

Because we consider 10 P -value thresholds, 10 GLMs are fitted and $H_0 : \phi_{GE} = 0$ is tested 10 times.

In Equation (6), $GRS_{Mi,t}$ is in the same scale as the number of phenotype-increasing alleles [18], and therefore the regression coefficient ϕ_{GE} can be explained as follows. For continuous traits, each additional trait-increasing allele is associated with ϕ_{GE} change in trait for subjects with $E_i = 1$ than for subjects with $E_i = 0$. For binary traits, each additional disease susceptibility allele is associated with an OR of $\exp(\phi_{GE})$ for subjects with $E_i = 1$ than for subjects with $E_i = 0$.

GRS based on SNP × E interactions

Hüls et al. have proposed the ‘GRS-interaction-training approach’ [28]. This is the 1st study presenting GRS with weights from the SNP × E interaction term itself. This approach was originally designed for pathway-oriented G × E studies. Borrowing the concept of ‘GRS-interaction-training approach’ [28], we construct a GRS according to the weights from the SNP × E interaction term itself. Different from the ‘GRS-interaction-training approach’ [28], we cannot fit a multivariate elastic net regression [36] on the typical large number of SNPs in GWAS. Instead, we estimate the SNP × E interaction effects by respective GLMs, as shown in the following Equation (7).

We first randomly split the whole sample into a training subset and a testing subset and an even split (1:1) is expected to yield the greatest power for GRS-I [22, 28]. Suppose the sample sizes of the training subset and the testing subset are n_1 and n_2 , ($n_1 \approx n_2$), respectively. We used the training subset to regress the phenotype on each SNP, E and SNP × E. The GLM for the l^{th} SNP ($l = 1, \dots, L$) is as follows:

$$g[E(Y_i)] = \beta_0 + \beta_{G_l} G_{il} + \beta_E E_i + \beta_{GE_l} G_{il} E_i + \beta'_X X_i, i = 1, \dots, n_1. \quad (7)$$

The pre-scaled GRS-I of the i^{th} subject in the testing subset is as follows:

$$GRS_{ii,t}^{pre} = \sum_{l=1}^L \hat{\beta}_{GE_l} G_{il} I(P_{GE_l} < P_t), i = 1, \dots, n_2, \quad (8)$$

where $\hat{\beta}_{GE_l}$ has been estimated by the GLM in Equation (7) and P_{GE_l} is the P -value of testing $H_0 : \beta_{GE_l} = 0$ versus $H_1 : \beta_{GE_l} \neq 0$ using the training subset.

For example, if the minor allele at the l^{th} SNP has a synergistic effect with E on the phenotype, $\hat{\beta}_{GE_l}$ estimated by the GLM in

Equation (7) will be positive. Subjects with more minor alleles at this SNP will get an increase in their $GRS_{i,t}^{pre}$. In contrast, if the minor allele at the i^{th} SNP has an antagonistic effect with E on the phenotype, $\hat{\beta}_{GE_i}$ will be negative and subjects with more minor alleles at this SNP will get a decrease in their $GRS_{i,t}^{pre}$. Therefore, summing the information of the L SNPs, a larger $GRS_{i,t}^{pre}$ indicates that the genetic makeup has a larger synergistic effect with E on the phenotype. $GRS_{i,t}^{pre}$ is then rescaled to calibrate the number of alleles exhibiting synergistic effects with E on the phenotype [18]:

$$GRS_{i,t} = \frac{GRS_{i,t}^{pre} \times \text{number of available SNPs}}{\text{sum of } |\hat{\beta}_{GE_i}| \text{ of available SNPs}}. \quad (9)$$

Given the t^{th} P-value threshold ($t = 1, \dots, 10$), we calculate $GRS_{i,t}$ for the n_2 subjects in the testing subset, fit the following GLM and test $H_0 : \psi_{GE} = 0$ versus $H_1 : \psi_{GE} \neq 0$:

$$g[E(Y_i)] = \psi_0 + \psi_G GRS_{i,t} + \psi_E E_i + \psi_{GE} GRS_{i,t} \cdot E_i + \psi_X X_i, i = 1, \dots, n_2. \quad (10)$$

Because we consider 10 P-value thresholds, 10 GLMs are fitted and $H_0 : \psi_{GE} = 0$ is tested 10 times.

In Equation (10), $GRS_{i,t}$ is in the same scale as the number of alleles exhibiting synergistic effects with E [18] and therefore the regression coefficient ψ_{GE} can be explained as follows. For continuous traits, each additional allele exhibiting synergistic effects with E is associated with ψ_{GE} change in trait for subjects with $E_i = 1$ than for subjects with $E_i = 0$. For binary traits, each additional allele exhibiting synergistic effects with E is associated with an OR of $\exp(\psi_{GE})$ for subjects with $E_i = 1$ than for subjects with $E_i = 0$.

The data-splitting strategy is required to preserve the type I error rates of GRS-I [22, 28]. However, we do not need to split the whole sample into two subsets when performing GRS-M, because the SNPs are screened according to their marginal associations with the phenotype (rather than the strength of SNP \times E). Corollary 1 proposed by Dai et al. [43] has justified the validity of using marginal associations as the screening test. Moreover, the following results of empirical type I error rates also verify that the data-splitting strategy is not required for GRS-M.

Numerical experiments and simulations

TWB GWAS data were used for our simulations to consider real human LD patterns. The TWB aims at building a database that integrates the genomic data and lifestyles of residents aged 30–70 years in Taiwan [65]. Our study included 16 555 unrelated community-based volunteers, among which 8213 were males and 8342 were females. This study was approved by the Research Ethics Committee of National Taiwan University Hospital (NTUH-REC no. 201612188RINA).

We removed SNPs with genotyping rates <95%, with Hardy–Weinberg test $P < 5.7 \times 10^{-7}$ [52], or with minor allele frequencies <1%. In total, 601 531 autosomal SNPs remained after removing SNPs that could not pass these quality control tests. These 601 531 autosomal SNPs that passed the quality control stage were used to construct the principal components (PCs) to adjust for population stratification.

In order to eliminate a large degree of redundancy in SNPs and compare our ADABF with the GRS approaches, we removed SNPs in high LD [42] according to the pruning stage described

in the ‘Methods’ section. After this pruning stage, 143 574 SNPs remained.

Then, we obtained a subset of SNPs that passed the screening stage by regressing diastolic blood pressure (DBP) on each of the 143 574 SNPs while adjusting for sex, age, BMI and the 1st seven PCs (the reason of considering the 1st seven PCs can be seen in the section of ‘Application to TWB data’). There were 7652 SNPs with larger marginal effects on DBP ($P < 0.05$). The genotypes of the 16 555 subjects at these 7652 SNPs were used as our simulation materials. Moreover, without losing generality, we used smoking status as our E. Among the 16 555 subjects, 4104 subjects (~24.8%) smoked for over 6 months, whereas 12 429 subjects did not. A total of 22 subjects did not respond to this question. We created a binary environmental exposure E, which was coded as 1 if the subject smoked for over 6 months and as 0 otherwise.

Type I error rates

We assessed the type I error rates by assuming a disease with a prevalence of 5% and generating binary traits based on the following model:

$$\log \frac{\Pr(Y = 1)}{1 - \Pr(Y = 1)} = \log \frac{0.05}{1 - 0.05} = -2.94. \quad (11)$$

The continuous traits were simulated by the following model:

$$Y = e, \quad (12)$$

where e was the random error term following the standard normal distribution.

Power (in the absence of SNP main effects)

We evaluated the power performance of ADABF and GRS by randomly selecting D SNPs and letting them interact with E, where $D = 20$ or 50. The following three situations were considered: (1) 20 SNP \times E with smaller effect sizes, (2) 20 SNP \times E with larger effect sizes and (3) 50 SNP \times E with smaller effect sizes.

The binary traits were simulated according to the following model:

$$\log \frac{\Pr(Y = 1)}{1 - \Pr(Y = 1)} = -2.94 + \sum_{d=1}^D \beta_{GE_d} G_d E. \quad (13)$$

We let 50% of β_{GE_d} s be positive and 50% of β_{GE_d} s be negative. $|\beta_{GE_d}|$ s ($d = 1, \dots, D$) were uniformly drawn from the intervals $[\log(1.2), \log(1.4)]$ and $[\log(1.4), \log(1.6)]$ for smaller effect sizes and larger effect sizes, respectively.

Our setting of D (20 or 50) and the size of β_{GE_d} [from $\log(1.2)$ to $\log(1.6)$] in Equation (13) is reasonable for a GWAS. Take a binary trait—hypertension (HYP) in TWB as an example, which is defined as DBP > 80 mm Hg or systolic BP (SBP) > 130 mm Hg [66]. Totally 7474 SNPs that passed the pruning and screening stages were analyzed according to Equation (2), in which age, gender, BMI and the 1st seven PCs were adjusted. We considered two binary Es—alcohol drinking (described in the following ‘Application to TWB data’) and smoking. In total, 127 $|\hat{\beta}_{GE}|$ s > $\log(1.2)$ when E = drinking and 22 $|\hat{\beta}_{GE}|$ s > $\log(1.4)$ when E = smoking; 1064 $|\hat{\beta}_{GE}|$ s > $\log(1.2)$ when E = drinking and 405 $|\hat{\beta}_{GE}|$ s > $\log(1.2)$ when E = smoking. Approximately 50% of these $\hat{\beta}_{GE}$ s were

positive and 50% of $\hat{\beta}_{GEs}$ were negative. Histograms of $\hat{\beta}_{GEs}$ can be seen from the top row of Figure 1.

Continuous traits were generated by the following model:

$$Y = \sum_{d=1}^D \beta_{GE_d} G_d E + e, \quad (14)$$

where e was the random error term following the standard normal distribution. We let 50% of β_{GE_d} s be positive and 50% of β_{GE_d} s be negative. $|\beta_{GE_d}|$ s ($d = 1, \dots, D$, where $D = 20$ or 50) were uniformly drawn from the intervals $[0.05, 0.07]$ and $[0.07, 0.09]$ for smaller effect sizes and larger effect sizes, respectively. Likewise, the abovementioned situations (1–3) were accordingly considered in the simulations of continuous traits.

Our assumption of D and the size of β_{GE_d} (from 0.05 to 0.09) in Equation (14) is also reasonable for continuous traits. Take DBP in TWB as an example. We first standardized DBP by $DBP' = (DBP - \overline{DBP}) / SD(DBP)$, where \overline{DBP} and $SD(DBP)$ were the mean and the SD of DBP, respectively. Through this standardization, DBP' was in the similar scale with Y in Equation (14), where e was simulated from the standard normal distribution. In total, 7652 SNPs that passed the screening stage were analyzed according to Equation (2), in which age, gender, BMI and the 1st seven PCs were adjusted. A total of 1315 $|\hat{\beta}_{GE}|$ s > 0.07 when $E =$ drinking and 451 $|\hat{\beta}_{GE}|$ s > 0.07 when $E =$ smoking. Approximately 50% of these $\hat{\beta}_{GEs}$ were positive and 50% of $\hat{\beta}_{GEs}$ were negative. Histograms of $\hat{\beta}_{GEs}$ are presented in the middle row of Figure 1.

Similarly, we obtained the standardized SBP' and analyzed 7508 SNPs that passed the screening stage. In total, 1167 $|\hat{\beta}_{GE}|$ s > 0.07 when $E =$ drinking and 353 $|\hat{\beta}_{GE}|$ s > 0.07 when $E =$ smoking. Among these stronger $\hat{\beta}_{GEs}$, ~50% of them were negative. Histograms of $\hat{\beta}_{GEs}$ can be found from the bottom row of Figure 1. Therefore, in our simulation, the number of D (20 or 50) and the size of β_{GE} are reasonable and modest.

Power (in the presence of SNP main effects)

We then evaluated the power performance of the polygenic approaches in the presence of SNP main effects. The binary traits were simulated according to the following model:

$$\log \frac{\Pr(Y = 1)}{1 - \Pr(Y = 1)} = -2.94 + \sum_{d=1}^D \beta_{G_d} G_d + \sum_{d=0.5D+1}^{1.5D} \beta_{GE_d} G_d E, \quad (15)$$

$|\beta_{G_d}|$ s ($d = 1, \dots, D$) and $|\beta_{GE_d}|$ s ($d = 0.5D + 1, \dots, 1.5D$, where $D = 20$ or 50) were uniformly drawn from the interval $[\log(1.2), \log(1.4)]$ for smaller effect sizes, and were uniformly drawn from $[\log(1.4), \log(1.6)]$ for larger effect sizes.

The continuous traits were simulated according to the following model:

$$Y = \sum_{d=1}^D \beta_{G_d} G_d + \sum_{d=0.5D+1}^{1.5D} \beta_{GE_d} G_d E + e, \quad (16)$$

where e was the random error term following the standard normal distribution. $|\beta_{G_d}|$ s ($d = 1, \dots, D$) and $|\beta_{GE_d}|$ s ($d = 0.5D + 1, \dots, 1.5D$, where $D = 20$ or 50) were uniformly drawn from $[0.05, 0.07]$ for smaller effect sizes and were uniformly drawn from $[0.07, 0.09]$ for larger effect sizes.

As expressed by Equations (15) and (16), we assume that SNPs $1 \sim 0.5D$ present only main effects, SNPs $(0.5D + 1) \sim D$ exhibit both main effects and SNP × E interactions, and SNPs $(D + 1) \sim 1.5D$ exhibit only SNP × E interactions on traits. According to our observation from real data analyses (Figure 1), we let 50% of β_{GE_d} s be positive and 50% of β_{GE_d} s be negative. Moreover, we found minor alleles could be trait increasing or trait decreasing in real data analyses and therefore 50% of β_{G_d} s were assumed to be positive and 50% of β_{G_d} s were assumed to be negative. Among the SNPs that exhibit both main effects and SNP × E interactions, we let $\beta_{G_d} \cdot \beta_{GE_d} > 0$ for ~50% SNPs and $\beta_{G_d} \cdot \beta_{GE_d} < 0$ for the remaining ~50% SNPs, where $d = 0.5D + 1, \dots, D$.

Results

Type I error rates

In GWAS, a stringent genome-wide significance level (5×10^{-8}) is typically used due to multiple hypothesis correction. The polygenic approaches investigated here combine SNPs across the genome in one test, and therefore, no multiple hypothesis correction is required. Table 1 presents the empirical type I error rates under a nominal significance level of 0.05 or 0.01, based on 10 000 replications of the binary traits and continuous traits separately. All the tests preserved the type I error rates. This simulation result confirms that the data-splitting strategy is not required for GRS-M or ADABF.

For GRS, in addition to the type I error rates under 10 P-value thresholds, respectively, we also present its type I error rate while considering the 10 P-value thresholds simultaneously and then correcting for multiple testing. In Table 1, M^* and I^* are GRS-M and GRS-I corrected for multiple testing, respectively. According to the Bonferroni correction (BON), the P-value of M^* (or I^*) is 10 times the minimum P-value of the 10 GRS-M (or GRS-I) tests. An M^* (or I^*) test is claimed to be significant if its P-value < 0.05 or 0.01 (the nominal significance level). The type I error rates of M^* and I^* are smaller than the nominal significance levels because of the conservative nature of the BON.

Power

Besides ADABF, GRS-M and GRS-I, we also evaluated the power performance of single marker analysis while controlling the family-wise error rate (FWER) at 5% using the BON, or controlling the FDR at 5% using the Benjamini–Hochberg approach (BH) [67]. Although BON and BH are not polygenic tests, they are also evaluated here because of their popularity. The power of BON or BH was calculated as the proportion of replications in which at least one SNP × E was identified.

Figure 2 presents the empirical power given the nominal significance level of 0.05, where the power of each scenario was calculated by 1000 replications. Our ADABF is more powerful than other methods in the absence of SNP main effects (Figure 2A and C). On the other hand, the presence of SNP main effects can considerably increase the power of GRS-M (Figure 2B and D). It constructs a GRS by aggregating the information of SNPs with stronger marginal effects. This approach becomes very powerful when some phenotype-associated SNPs also exhibit interactions with E.

Regarding the performance of pinpointing true SNP × E, we compared the sensitivity and PPV of our ADABF with BON and BH. Sensitivity is defined as the total number of true findings over the total number of SNP × E in the 1000 simulation

Table 1. Empirical type I error rates in the simulation study

Traits	Nominal significance levels	ADABF	P-value threshold (P_t)											
			0.0001	0.00025	0.0005	0.001	0.0025	0.005	0.01	0.025	0.05	0.1	M* or I*	
Binary	0.05	0.0466	0.0494	0.0510	0.0497	0.0524	0.0514	0.0539	0.0439	0.0485	0.0446	0.0439	0.0468	0.0363
	0.01	0.0086	GRS-I	0.0534	0.0560	0.0557	0.0491	0.0439	0.0446	0.0514	0.0488	0.0514	0.0507	0.0351
			GRS-M	0.0135	0.0116	0.0087	0.0098	0.0088	0.0090	0.0076	0.0100	0.0083	0.0083	0.0059
Continuous	0.05	0.0508	0.0085	0.0123	0.0115	0.0108	0.0107	0.0112	0.0088	0.0085	0.0134	0.0107	0.0076	
	0.01	0.0105	GRS-M	0.0470	0.0480	0.0499	0.0434	0.0507	0.0495	0.0488	0.0522	0.0480	0.0484	0.0357
			GRS-I	0.0486	0.0509	0.0523	0.0570	0.0488	0.0501	0.0509	0.0490	0.0515	0.0476	0.0386
			GRS-M	0.0077	0.0105	0.0079	0.0087	0.0093	0.0100	0.0100	0.0094	0.0111	0.0077	
			GRS-I	0.0087	0.0127	0.0110	0.0126	0.0108	0.0095	0.0114	0.0104	0.0101	0.0119	0.0075

Empirical type I error rates of ADABF, the GRS based on marginal associations (GRS-M), and the GRS based on interaction effects (GRS-I) are shown. Each GRS is evaluated at ten P-value thresholds: 0.0001, 0.00025, 0.0005, 0.001, 0.0025, 0.005, 0.01, 0.025, 0.05, and 0.1. Each entry represents the proportion of P-values smaller than the corresponding nominal significance level based on 10,000 simulation replications. M* and I* are GRS-M and GRS-I corrected for multiple testing, respectively. The P-value of M* (or I*) is ten times the minimum P-value of the ten GRS-M (or GRS-I) tests. An M* (or I*) test is claimed to be significant if its P-value < 0.05 or 0.01 (the nominal significance level).

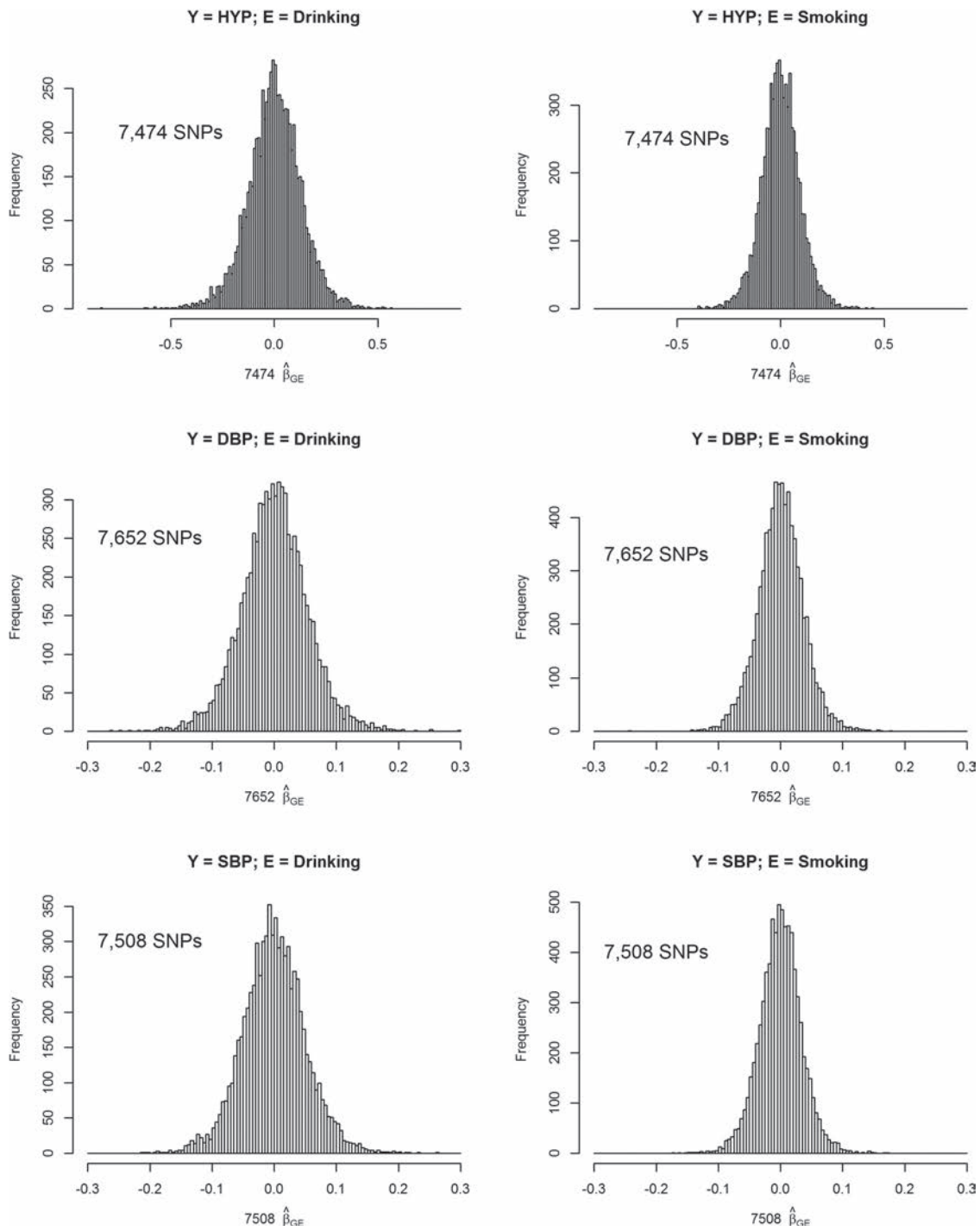


Figure 1. Histograms of $\hat{\beta}_{GE}$ s for SNPs that passed the pruning and screening stages. In total, 7474, 7652 and 7508 SNPs passed the pruning and screening stages for analyses of HYP (a binary trait), DBP and SBP, respectively. These SNPs were analyzed according to Equation (2), in which age, gender, BMI and the 1st seven PCs were adjusted. Two binary Es including alcohol drinking (left column) and smoking (right column) were considered. Here we show the histograms of $\hat{\beta}_{GE}$ s from the GLM in Equation (2).

replications, i.e. 20000 or 50000 (recall $D = 20$ or 50 in Equations 13–16). PPV is defined as the total number of true findings over the total number of findings in the 1000 simulation replications. BON is the most conservative method, and therefore, it has the lowest sensitivity and the highest PPV (Figure 3). ADABF and BH performed similarly.

$PPV = 1 - FDP$, where FDP is the false discovery proportion. As shown in Figure 3, the FDPs are generally larger than the desired level of FDR, 5%. This is mostly because the FDR procedures assume that the statistics are unbiased [68]. Because fitting a multivariate regression on all SNPs is computationally infeasible, ADABF, BON and BH are all based on

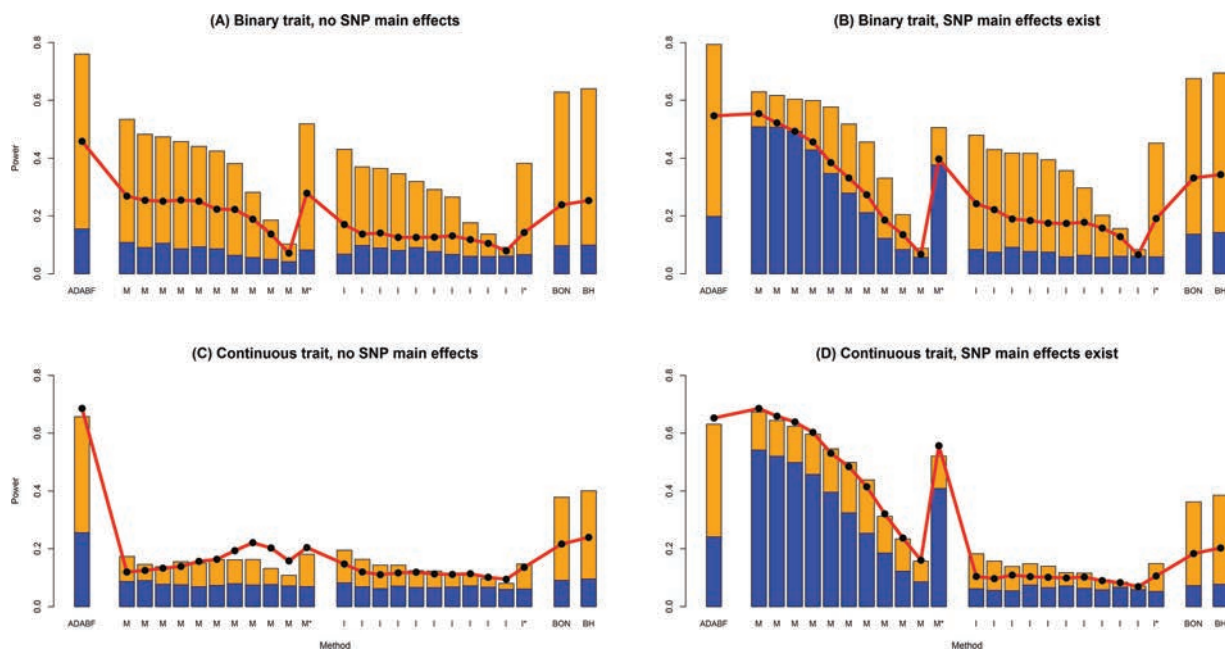


Figure 2. Empirical power given a nominal significance level of 0.05. The empirical power of ADABF, GRS-M (M) and GRS-I (I). GRS-M and GRS-I were evaluated at 10 P-value thresholds (from the left bar to the right bar): 0.0001, 0.00025, 0.0005, 0.001, 0.0025, 0.005, 0.01, 0.025, 0.05 and 0.1. M* and I* are GRS-M and GRS-I corrected for multiple testing, respectively. The P-value of M* (or I*) is 10 times the minimum P-value of the 10 GRS-M (or GRS-I) tests. An M* (or I*) test is claimed to be significant if its P-value < 0.05 (the nominal significance level). BON is single marker analysis while controlling the FWER at 5% using the BON; BH is single marker analysis while controlling the FDR at 5% using the BH. The height of the blue (yellow) bars marks the empirical power of each test given 20 SNP × E with smaller (larger) effect sizes. The red line with black points marks the empirical power given 50 SNP × E with smaller effect sizes.

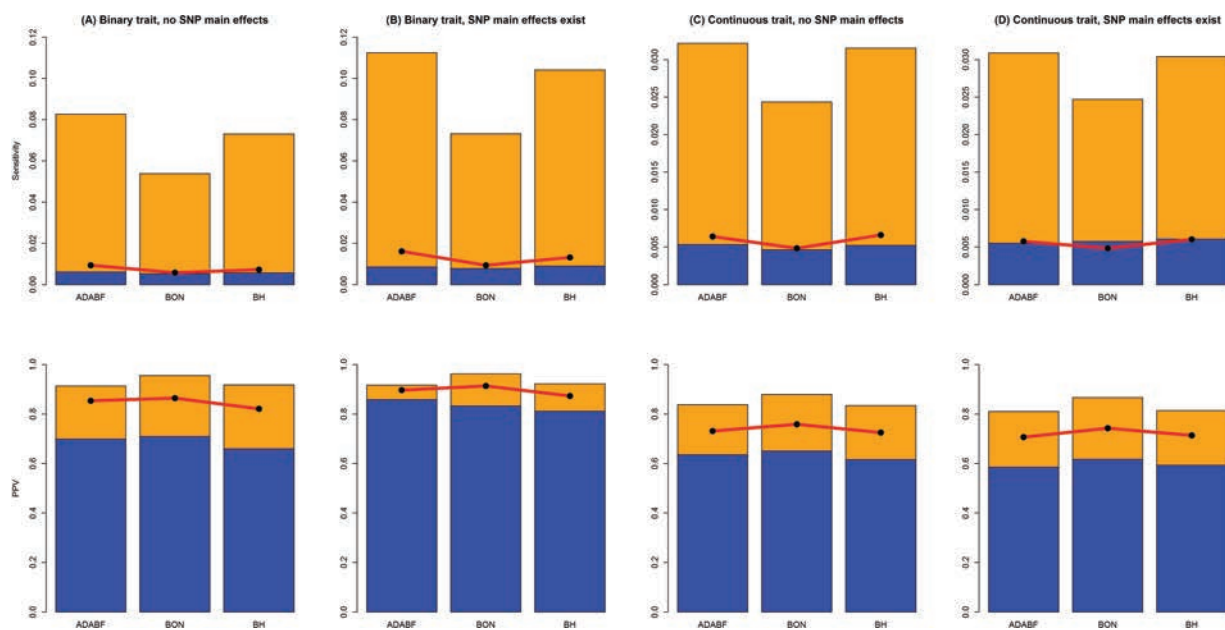


Figure 3. Sensitivity and PPV The sensitivity (the top row) is defined as the total number of true findings over the total number of SNP × E in the 1000 simulation replications, i.e. 20 000 or 50 000 (recall $D = 20$ or 50 in Equations 13–16). PPV (the bottom row) is defined as the total number of true findings over the total number of findings in the 1000 simulation replications. BON is single marker analysis while controlling the FWER at 5% using the BON; BH is single marker analysis while controlling the FDR at 5% using the BH. The height of the blue (yellow) bars marks the sensitivity/PPV of each method given 20 SNP × E with smaller (larger) effect sizes. The red line with black points marks the sensitivity/PPV given 50 SNP × E with smaller effect sizes.

regression models that consider one SNP at a time (Equation 2). The statistics estimated from Equation 2 (β_{GE_i} s) are biased because complex traits are usually influenced by multiple

genetic variants, environmental exposures and the interplay between them (e.g. Equations 13–16). Therefore, the FDPs are larger than 5% for all the three methods (i.e. ADABF, BON and

BH). The high FDP irrespective of the procedure used to correct for multiple testing has also been observed in regression-based analyses for environment-wide association study [68].

Application to TWB data

We then applied these approaches to TWB data to explore SNP \times alcohol and SNP \times smoking interactions on BP. Approximately 85% of the TWB subjects were of the Han Chinese ancestry and \sim 14.5% of the subjects belonged to a 3rd group that is genetically distinct from neighboring Southeast Asians [65].

In the TWB data, 'drinking' is defined as a weekly intake of greater than 150 cc of alcohol for at least 6 months. Among the 16 555 subjects, 1764 subjects (\sim 10.6%) answered 'yes' to alcohol drinking, whereas 14 779 subjects answered 'no'. A total of 12 subjects did not respond to this question and were regarded as missing values. Smoking has been described in the section of 'Numerical experiments and simulations'. Both alcohol drinking and smoking are binary Es. To obtain more reliable BP [69, 70], two measurements were taken with a 5 min rest interval and the average of the two measurements was recorded.

Regarding single marker analysis, for each of the 601 531 autosomal SNPs, we regressed DBP, SBP or HYP (yes versus no) on SNP, the E (i.e. alcohol drinking or smoking), SNP \times E, while adjusting for age, gender, BMI and the 1st seven PCs. After adjusting for the 1st seven PCs, PLINK reported very low genomic inflation factors that were based on the median of the Chi-square statistics, i.e. $\lambda_{GC} = 1.00$ for both the DBP and SBP analyses and $\lambda_{GC} = 1.02$ for HYP analysis. No significant SNP \times E were found from the 601 531 autosomal SNPs, by controlling the FWER at 5% with the BON or by controlling the FDR at 5% with the BH [67].

We then performed the GRS tests using the subset of SNPs that were nearly independent [22], i.e. the 143 574 SNPs that passed the pruning stage by the PLINK command shown in the section of 'Methods'. Considering the 10 P -value thresholds used for the GRS, the Bonferroni-corrected significance level to control the FWER at 5% should be $0.05/10 = 0.005$. The pink horizontal line in Figure 4A marks the significance threshold, i.e. $-\log_{10}(0.005) \approx 2.3$. GRS-M identified SNP \times alcohol interactions on DBP and SBP and SNP \times smoking interactions on DBP (Figure 4A).

Figure 4C shows the regression coefficient $\hat{\psi}_{GE}$ in Equation (6) at 10 P -value thresholds (P_t). For example, when $P_t = 5 \times 10^{-4}$, GRS_Ms of SBP, DBP and HYP analyses are computed by 91, 102 and 88 SNPs, respectively. Each additional SBP-increasing allele (DBP-increasing allele) is associated with \sim 0.20 (\sim 0.10) mm Hg higher SBP (DBP) in drinkers than in nondrinkers. Each additional HYP susceptibility allele is associated with an OR of exp(0.015) = 1.015 in drinkers than in nondrinkers. Moreover, each additional SBP-increasing allele (DBP-increasing allele) is associated with \sim 0.07 (\sim 0.07) mm Hg higher SBP (DBP) in smokers than in nonsmokers.

Finally, as described in the screening stage of the 'Methods' section, SNPs passing the screening at the desired significance level ($P < .05$) are then analyzed using ADABF. In the analysis of marginal associations, DBP, SBP and HYP were separately regressed on each of the 143 574 SNPs while adjusting for age, gender, BMI and the 1st seven PCs. Linear regression models were used for analyses of DBP and SBP, whereas logistic regression models were fitted for HYP. In total, 7652, 7508 and 7474 SNPs

passed the screening test for DBP, SBP and HYP, respectively. Table 2 shows the analysis results based on 10^5 resampling replicates in the ADABF approach. A P -value < 0.05 or 0.01 is sufficient to reject H_0 of no polygenic SNP \times E interactions [57]. Like GRS-M (Figure 4A), ADABF identified SNP \times alcohol interactions on DBP and SBP and SNP \times smoking interactions on DBP, but ADABF provided more significant P -values than GRS-M. Additionally, ADABF identified SNP \times alcohol interactions on HYP.

GRS-I did not provide any significant results under the Bonferroni-corrected significance level of $0.05/10 = 0.005$ (Figure 4B). The pink horizontal line in Figure 4B marks the significance threshold, i.e. $-\log_{10}(0.005) \approx 2.3$. This result was consistent with the above finding in our simulations. That is, GRS-I is the least powerful approach due to its data-splitting strategy.

Figure 4D shows the regression coefficient $\hat{\psi}_{GE}$ in Equation (10) at 10 P -value thresholds (P_t). Although GRS-I tests are not significant at all the 10 P_t s, we still explain the meaning of $\hat{\psi}_{GE}$ here. For example, when $P_t = 5 \times 10^{-4}$, GRS_Is of SBP (E = drinking and smoking) and DBP (E = drinking and smoking) analyses are computed by 89, 84, 146 and 113 SNPs, respectively. Each additional allele exhibiting synergistic effects with drinking is associated with \sim 0.05 (\sim 0.05) mm Hg higher SBP (DBP) in drinkers than in nondrinkers. Moreover, each additional allele exhibiting synergistic effects with smoking is associated with \sim 0.02 (\sim 0.04) mm Hg higher SBP (DBP) in smokers than in nonsmokers.

We also identified an rs10811568 \times alcohol interaction on DBP (resampling FDR = 1.2%), rs62065089 \times alcohol interaction on SBP (resampling FDR = .4%) and rs79990035 \times smoking interaction on DBP (resampling FDR = 1.1%) through the resampling procedure in ADABF. Table 3 lists the detailed information on these three SNPs. Figure 5 presents plots of these interaction effects.

These three SNPs can also be identified by BON/BH when controlling the FWER/FDR at 5% given the \sim 7600 SNPs that passed the pruning and screening stages. Because we have removed SNPs in high LD, these \sim 7600 SNPs are nearly independent of each other and therefore the resampling FDR and BH lead to the same results. Despite this, performing ADABF is still worthwhile. As shown by Table 2, SNP \times alcohol interactions on HYP can only be detected by the ADABF polygenic test, although no individual SNP \times alcohol interactions can be identified from the subsequent resampling FDR. Nothing can be found if we bypass the ADABF polygenic test and directly use BON/BH. This result is consistent with the power gain of ADABF compared with BON/BH (Figure 2).

For a polygenic test, a P -value < 0.05 or 0.01 is sufficient to reject H_0 of no polygenic effects [57]. Therefore, resampling 10^5 replicates is sufficient for a real data analysis. ADABF takes \sim 2.5 h to analyze SNP \times E on a continuous phenotype based on a Linux platform with a Dell Intel Xeon E5-2690 2.9 GHz processor and 8 GB of memory. Approximately 3.4 h are required for the analysis of a binary phenotype, because fitting a logistic regression takes more computation time than fitting a linear regression.

Discussion

Genetic effects can differ between subjects depending on their lifestyle factors or environmental exposure because of $G \times E$ [5]. Therefore, the identification of $G \times E$ is important for investigating new mechanisms in disease [71]. Given a specific sample size,

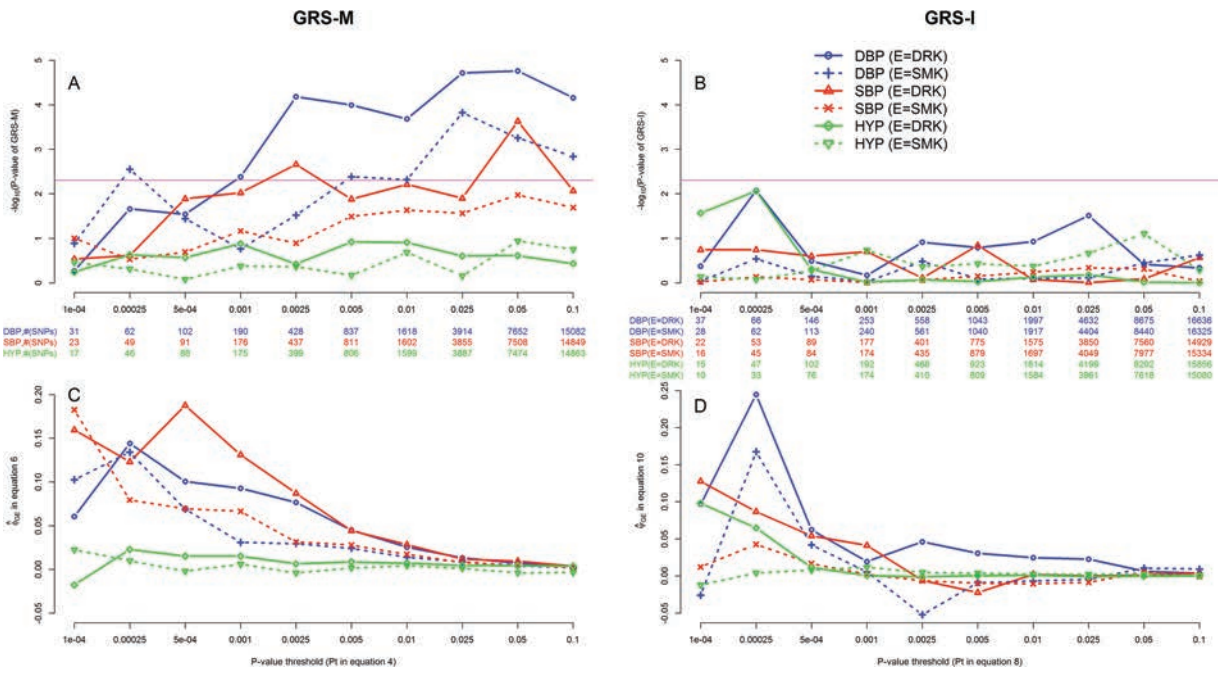


Figure 4. TWB analysis results using the GRS-M and GRS-I tests. The left and right columns show the GRS-M and GRS-I results, respectively. The black x-axes list the 10 P-value thresholds, i.e. in Equations (4) or (8). The blue (for DBP analysis), red (for SBP analysis) and green (for HYP analysis) x-axes list the number of SNPs used to construct GRS_M or GRS_I . The y-axes of plots (A) and (B) are $-\log_{10}(P\text{-value of GRS-M})$ and $-\log_{10}(P\text{-value of GRS-I})$, respectively. Considering the 10 P-value thresholds used for the GRS, the Bonferroni-corrected significance level to control the FWER at 5% is $^{.05}/10 = .005$. The pink horizontal lines in plots (A) and (B) mark $-\log_{10}(0.005) = 2.3$. Moreover, ϕ_{GE} and ψ_{GE} are estimated from Equations (6) and (10), respectively, and are shown in the y-axes of plots (C) and (D).

Table 2. TWB analysis results using the ADABF, BON, and BH approaches

	ADABF ¹	BON ²	BH ³
SNP_xalcohol on DBP (based on 7,652 SNPs)			
P-value	< 0.00001	—	—
SNP found to have interaction with alcohol consumption	rs10811568 (Resampling FDR = 1.2%)	rs10811568	rs10811568
SNP_xalcohol on SBP (based on 7,508 SNPs)			
P-value	< 0.00001	—	—
SNP found to have interaction with alcohol consumption	rs62065089 (Resampling FDR = 0.4%)	rs62065089	rs62065089
SNP_xalcohol on HYP (based on 7,474 SNPs)			
P-value	0.00098	—	—
SNP found to have interaction with alcohol consumption	—	—	—
SNP_xsmoking on DBP (based on 7,652 SNPs)			
P-value	0.00059	—	—
SNP found to have interaction with smoking	rs79990035 (Resampling FDR = 1.1%)	rs79990035	rs79990035
SNP_xsmoking on SBP (based on 7,508 SNPs)			
P-value	0.1573	—	—
SNP found to have interaction with smoking	—	—	—
SNP_xsmoking on HYP (based on 7,474 SNPs)			
P-value	0.0592	—	—
SNP found to have interaction with smoking	—	—	—

¹The P-value of ADABF and the resampling FDR were based on 10^5 resampling replicates. In SNP_xalcohol interaction analysis on DBP or SBP, the observed interaction signal was more significant than that of all the 10^5 resampling replicates. Therefore, the P-values were represented as “<0.00001”. A P-value < 0.05 or 0.01 is sufficient to reject H_0 of no polygenic SNPE interactions [57]. No more resampling replicates are required to obtain a more precise P-value. P-values < 0.05 are highlighted.

²BON is single marker analysis while controlling the FWER at 5% using the Bonferroni correction.

³BH is single marker analysis while controlling the FDR at 5% using the Benjamini-Hochberg approach.

the power to detect $G \times E$ is much lower than the power to detect genetic main effects [47, 72]. The exploration of $G \times E$ from GWAS data is even more challenging due to the harsh multiple-testing penalty. There is a great need to discover a powerful polygenic approach that can identify $G \times E$ and further pinpoint SNPs that interact with E.

Most complex diseases are polygenic (influenced by many small genetic effects), including obesity [73, 74], HYP [75], schizophrenia and bipolar disorder [76]. The GRS that aggregates multiple genetic variants into a score is widely used for testing and prediction [77, 78]. A majority of the $G \times E$ findings were discovered by the GRS approach [3, 5, 79–83]. For example,

Table 3. The three SNPs that interact with alcohol consumption or smoking

SNP	Chromosome	Position (base pair)	Mapped gene	Minor allele	Major allele	MAF	E	Phenotype	SNP × E interaction test ¹				
									$\widehat{\beta}_{GE}$	s.e. ($\widehat{\beta}_{GE}$)	Wald statistic	P-value	Bayesfactor
rs10811568	9	21543444	MIR31HG	G	A	0.235	Alcohol	DBP ²	1.9753	0.4125	4.788	1.7×10^{-6}	11,875
rs62065089	17	63843058	CEP112	A	C	0.122	Alcohol	SBP ³	3.1164	0.6240	4.994	6.0×10^{-7}	31,233
rs79990035	2	54418123	ACY2	T	C	0.053	Smoking	DBP ⁴	1.4872	0.5403	2.753	0.00592	8
								SBP ⁵	4.0673	0.8169	4.979	6.5×10^{-7}	28,933
								SBP	-2.6929	0.5601	-4.808	1.5×10^{-6}	12,827
								SBP	-2.4950	0.8469	-2.946	0.00322	14

¹DBP or SBP was regressed on each of the SNPs, the environmental factor (i.e., alcohol consumption or smoking), and SNP × E, while adjusting for age, gender, BMI, and the first seven principal components.

²The rs10811568 × alcohol interaction was found in DBP analysis (the highlighted row).

³Because the marginal association of SNP rs10811568 with SBP was not significant ($P > 0.05$), this SNP was not in the screened subset of 7,508 SNPs. Therefore, it was not pinpointed from the SNP × alcohol interaction analysis on SBP.

⁴The rs62065089 × alcohol interaction was found in SBP analysis (the highlighted row).

⁵The rs79990035 × smoking interaction was found in DBP analysis (the highlighted row).

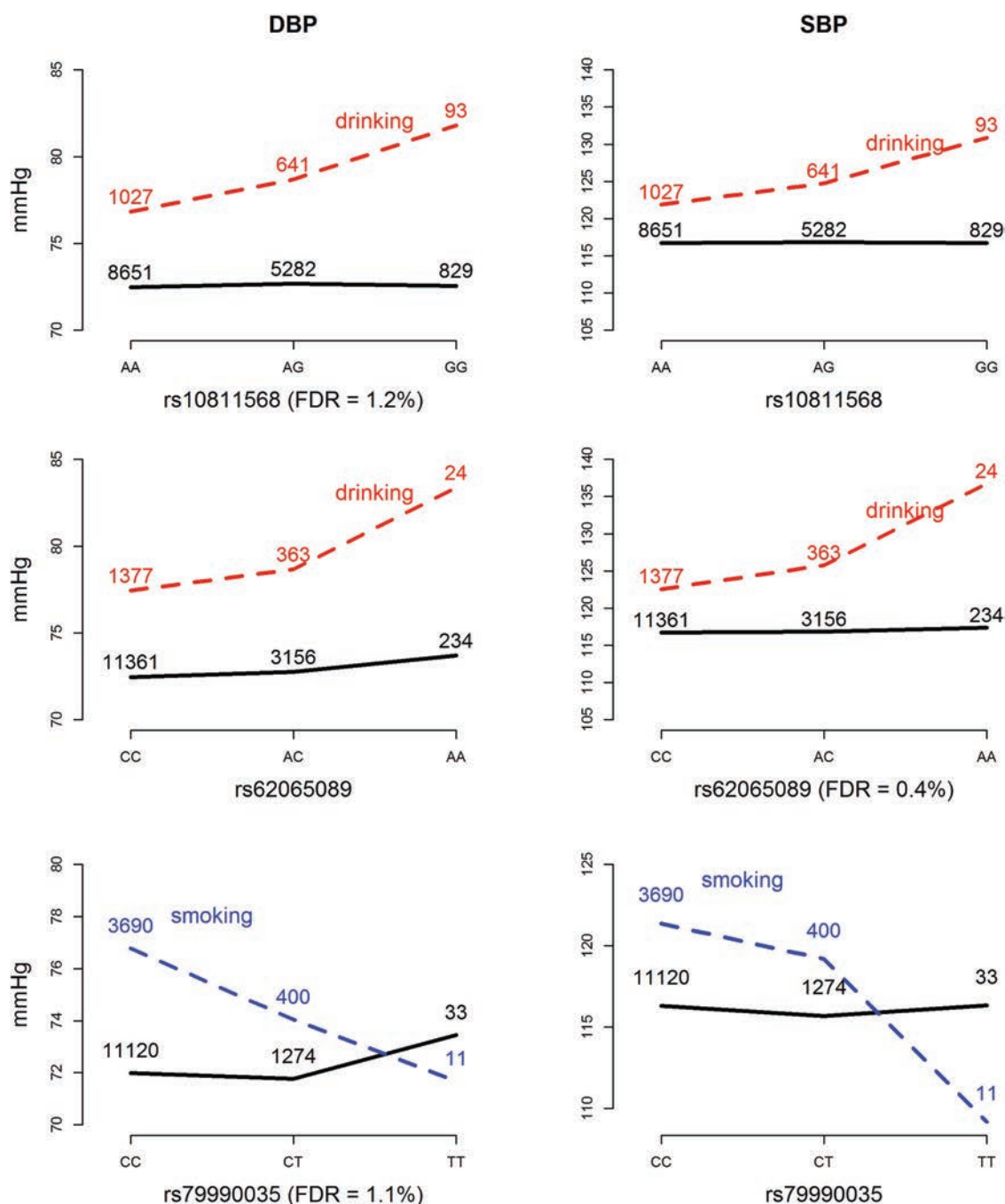


Figure 5. Plots of SNP \times alcohol or SNP \times smoking interaction effects on DBP and SBP. Controlling the resampling FDR at 5%, we found an rs10811568 \times alcohol interaction on DBP, rs62065089 \times alcohol interaction on SBP and rs79990035 \times smoking interaction on DBP. As shown in these plots, these identified interaction patterns in DBP/SBP are similar to those in SBP/DBP. The black curves depict the mean of DBP/SBP among the nondrinkers/nonsmokers, whereas the red/blue dashed curves depict the mean among the drinkers/smokers. The number shown on each point represents the sample size of that category.

recently, Rask-Andersen *et al.* [5] constructed a GRS composed of 94 independent BMI-associated SNPs that were reported by a previous GWAS [32], and they found interactions between this GRS and several Es. However, an appropriate external GWAS may not be available for other phenotypes or other ethnicity.

When external information is unavailable, the weights for a GRS have to be determined internally. Because the 'best' P-value threshold for the 'optimal' subset of SNPs is unknown, many studies constructed a panel of GRSs under various P-value

thresholds [21, 78, 84, 85]. Therefore, the significance of a GRS test has to be corrected by the number of P-value thresholds evaluated [57]. (Specifying a P-value threshold to select SNPs is not an issue in the 'GRS-marginal-internal approach' and 'GRS-interaction-training approach'. Hüls *et al.* select SNPs according to a multivariate elastic net regression [28, 35].)

We here compare our ADABF with the GRS-M and GRS-I tests. Regarding the power to detect polygenic-environment interactions, ADABF is the most powerful test in the absence

of SNP main effects (Figure 2A and C). When SNP main effects exist, GRS-M can outperform ADABF and can become the best test (Figure 2B and D). GRS-I is the least powerful approach due to its data-splitting strategy.

In most applications of BFs [52], specifying a different prior variance (W) will change the magnitude of a BF and can lead to a different inference. For example, a BF between 10 and 100 is regarded as a 'strong' evidence against H_0 , whereas a BF larger than 100 is deemed as a 'decisive' evidence against H_0 [86]. However, ADABF does not fully rely on the absolute magnitudes of BFs. Instead, ADABF compares the BFs from the observed data with those from the resampling replicates, under the same prior. Therefore, this method is robust to the setting of the prior variance (W) [40]. When W was set at $.1^2 = 0.01$ or $0.3^2 = 0.09$, the ADABF results were very close to those obtained from $W = 0.2^2 = 0.04$ (Supplementary Table S1 and 2).

When multiple SNPs interact with E but their effect sizes are small, we may obtain a significant ADABF test result; however, no individual $SNP \times E$ can be identified by controlling the resampling FDR at 5%. $SNP \times$ alcohol on HYP shown in Table 2 is an example. In this situation, if we bypass ADABF and directly use BH to identify $SNP \times E$, we will find nothing, and $G \times E$ can be missed (comparing the power of ADABF and BH in Figure 2).

Gene \times alcohol and gene \times smoking interactions on BP have been found in Caucasians [49, 87] and Japanese [88, 89]. Our results support these $G \times E$ on BP in Han Chinese as well. Although ADABF is a powerful polygenic test for detecting $G \times E$, the identification of individual $SNP \times E$ is still very challenging. If the ADABF test is significant (P -value < 0.05 or 0.01 , no multiple hypothesis correction is required), we conclude that $G \times E$ interactions exist and the E can modify the genetic effect on the phenotype. However, each $SNP \times E$ test may not be able to reach a sufficient power that can withstand the multiple hypothesis correction. In this situation, GRS-M can help to identify whether synergistic or antagonistic interactions exist between risk alleles and E , according to the sign of $\hat{\phi}_{GE}$ in Equation (6). For example, the positive $\hat{\phi}_{GE}$ s (Figure 4C) indicate that BP-increasing alleles elevate more BP in drinkers (smokers) than in nondrinkers (nonsmokers).

We found the rs79990035 \times smoking interaction on DBP (resampling FDR = 1.1%). The SNP rs79990035 is located in the *acylphosphatase 2 (ACYP2)* gene. Cheng *et al.* recently also found an ACYP2 \times smoking interaction related to susceptibility to ischemic stroke (IS) in a Han Chinese population [90]. High BP levels are usually observed in acute IS patients [91–93]. Therefore, for Han Chinese, the ACYP2 \times smoking interaction on BP warrants further investigation.

Some $G \times E$ have been discovered by constructing a GRS based on external GWAS results [3–5, 17–19, 29–31]. However, appropriate external information is not always available. This work compares polygenic approaches to identify $G \times E$ in the context of GWAS, when external information is unavailable. Using ADABF or GRS-M, many hidden $G \times E$ might be explored. Moreover, GRS-M can help to identify whether synergistic or antagonistic interactions exist between risk alleles and E .

Key Points

- The scientific community usually constructs a GRS and tests the interaction between this score and an E . However, until now, little has been known about the

performance of the GRS for detecting $G \times E$. Moreover, appropriate external weights required for a GRS are not always available.

- We explored a powerful polygenic approach for detecting $G \times E$ when external weights are not available, by comparing our 'ADABF' with the GRS-M and GRS-I effects.
- ADABF is the most powerful method in the absence of SNP main effects, whereas GRS-M is generally the best test when SNP main effects exist. GRS-I is the least powerful test due to its data-splitting strategy. This work provides guidance to choose a polygenic approach to detect $G \times E$ when external information is unavailable.

Supplementary Data

Supplementary data are available online at <https://academic.oup.com/bib>.

Acknowledgements

The authors would like to thank the four anonymous reviewers for their insightful and constructive comments and Ya-Chin Lee for assisting with the acquisition of TWB data.

Funding

Ministry of Science and Technology of Taiwan (grant nos. 107-2314-B-002-195-MY3 and 106-2314-B-002-040 to W.-Y.L.), MST grant (MST 102-2314-B-002-117-MY3 to P.-H.K.) and National Taiwan University Hospital (UN106-050 to Shyr-Chyr Chen and P.-H.K.).

References

1. Cadoret RJ, Cain CA, Crowe RR. Evidence for gene-environment interaction in the development of adolescent antisocial behavior. *Behav Genet* 1983;13:301–10.
2. Manuck SB, McCaffery JM. Gene-environment interaction. *Annu Rev Psychol* 2014;65:41–70.
3. Ahmad S, Rukh G, Varga TV, *et al.* Gene \times physical activity interactions in obesity: combined analysis of 111 421 individuals of European ancestry. *PLoS Genet* 2013;9:e1003607.
4. Li S, Zhao JH, Luan J, *et al.* Physical activity attenuates the genetic predisposition to obesity in 20 000 men and women from EPIC-Norfolk prospective population study. *Plos Med* 2010;7:e1000332.
5. Rask-Andersen M, Karlsson T, Ek WE, *et al.* Gene-environment interaction study for BMI reveals interactions between genetic factors and physical activity, alcohol consumption and socioeconomic status. *PLoS Genet* 2017;13:e1006977.
6. Ottman R. Gene-environment interaction: definitions and study designs. *Prev Med* 1996;25:764–70.
7. Aschard H. A perspective on interaction effects in genetic association studies. *Genet Epidemiol* 2016;40:678–88.
8. Winham SJ, Biernacka JM. Gene-environment interactions in genome-wide association studies: current approaches and new directions. *J Child Psychol Psychiatry* 2013; 54:1120–34.

9. Aschard H, Chen J, Cornelis MC, et al. Inclusion of gene–gene and gene–environment interactions unlikely to dramatically improve risk prediction for complex diseases. *Am J Hum Genet* 2012;**90**:962–72.
10. Campa D, Kaaks R, Le Marchand L, et al. Interactions between genetic variants and breast cancer risk factors in the breast and prostate cancer cohort consortium. *J Natl Cancer Inst* 2011;**103**:1252–63.
11. Hutter CM, Mechanic LE, Chatterjee N, et al. Gene–environment interactions in cancer epidemiology: a National Cancer Institute Think Tank report. *Genet Epidemiol* 2013;**37**:643–57.
12. Frost HR, Shen L, Saykin AJ, et al. Identifying significant gene–environment interactions using a combination of screening testing and hierarchical false discovery rate control. *Genet Epidemiol* 2016;**40**:544–57.
13. Chen H, Meigs JB, Dupuis J. Incorporating gene–environment interaction in testing for association with rare genetic variants. *Hum Hered* 2014;**78**:81–90.
14. Jiao S, Hsu L, Bezieau S, et al. SBERIA: set-based gene–environment interaction test for rare and common variants in complex diseases. *Genet Epidemiol* 2013;**37**:452–64.
15. Lin XY, Lee S, Christiani DC, Lin X. Test for interactions between a genetic marker set and environment in generalized linear models. *Biostatistics* 2013;**14**:667–81.
16. Lin XY, Lee S, Wu MC, et al. Test for rare variants by environment interactions in sequencing association studies. *Biometrics* 2016;**72**:156–64.
17. Nakamura S, Narimatsu H, Sato H, et al. Gene–environment interactions in obesity: implication for future applications in preventive medicine. *J Hum Genet* 2016;**61**:317–22.
18. Tyrrell J, Wood AR, Ames RM, et al. Gene–obesogenic environment interactions in the UK Biobank study. *Int J Epidemiol* 2017;**46**:559–75.
19. Mullins N, Power RA, Fisher HL, et al. Polygenic interactions with environmental adversity in the aetiology of major depressive disorder. *Psychol Med* 2016;**46**:759–70.
20. International Schizophrenia Consortium, Purcell SM, Wray NR, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009;**460**:748–52.
21. Wang SH, Hsiao PC, Yeh LL, et al. Polygenic risk for schizophrenia and neurocognitive performance in patients with schizophrenia. *Genes Brain Behav* 2018;**17**:49–55.
22. Dudbridge F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet* 2013;**9**:e1003348.
23. Frayling TM, Timpson NJ, Weedon MN, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 2007;**316**:889–94.
24. Scuteri A, Sanna S, Chen WM, et al. Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. *PLoS Genet* 2007;**3**:1200–10.
25. Loos RJF, Lindgren CM, Li SX, et al. Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nat Genet* 2008;**40**:768–75.
26. Willer CJ, Speliotes EK, Loos RJF, et al. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat Genet* 2009;**41**:25–34.
27. Thorleifsson G, Walters GB, Gudbjartsson DF, et al. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat Genet* 2009;**41**:18–24.
28. Hüls A, Kramer U, Carlsten C, et al. Comparison of weighting approaches for genetic risk scores in gene–environment interaction studies. *BMC Genet* 2017;**18**:115.
29. Karlson EW, Ding B, Keenan BT, et al. Association of environmental and genetic factors and gene–environment interactions with risk of developing rheumatoid arthritis. *Arthritis Care Res* 2013;**65**:1147–56.
30. Joshi AD, Lindstrom S, Husing A, et al. Additive interactions between susceptibility single-nucleotide polymorphisms identified in genome-wide association studies and breast cancer risk factors in the breast and prostate cancer cohort consortium. *Am J Epidemiol* 2014;**180**:1018–27.
31. Maas P, Barrdahl M, Joshi AD, et al. Breast cancer risk from modifiable and nonmodifiable risk factors among white women in the United States. *JAMA Oncol* 2016;**2**:1295–302.
32. Locke AE, Kahali B, Berndt SI, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* 2015;**518**:197–206.
33. Sullivan PF, Daly MJ, Ripke S, et al. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry* 2013;**18**:497–511.
34. Peyrot WJ, Milaneschi Y, Abdellaoui A, et al. Effect of polygenic risk scores on depression in childhood trauma. *Br J Psychiatry* 2014;**205**:113–19.
35. Hüls A, Ickstadt K, Schikowski T, et al. Detection of gene–environment interactions in the presence of linkage disequilibrium and noise by using genetic risk scores with internal weights from elastic net regression. *BMC Genet* 2017;**18**:55.
36. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Statist Soc B* 2005;**67**:301–20.
37. Li HX, Beeghly-Fadiel A, Wen WQ, et al. Gene–environment interactions for breast cancer risk among Chinese women: a report from the Shanghai Breast Cancer Genetics Study. *Am J Epidemiol* 2013;**177**:161–70.
38. Garcia-Closas M, Gunsoy NB, Chatterjee N. Combined associations of genetic and environmental risk factors: implications for prevention of breast cancer. *J Natl Cancer Inst* 2014;**106**:dju305.
39. Garcia-Closas M, Rothman N, Figueroa JD, et al. Common genetic polymorphisms modify the effect of smoking on absolute risk of bladder cancer. *Can Res* 2013;**73**:2211–20.
40. Lin WY, Chen WJ, Liu CM, et al. Adaptive combination of Bayes factors as a powerful method for the joint analysis of rare and common variants. *Sci Rep* 2017;**7**:13858.
41. International Multiple Sclerosis Genetics C, Bush WS, Sawcer SJ, et al. Evidence for polygenic susceptibility to multiple sclerosis—the shape of things to come. *Am J Hum Genet* 2010;**86**:621–5.
42. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;**81**:559–75.
43. Dai JY, Kooperberg C, Leblanc M, et al. Two-stage testing procedures with independent filtering for genome-wide gene–environment interaction. *Biometrika* 2012;**99**:929–44.
44. Hsu L, Jiao S, Dai JY, et al. Powerful cocktail methods for detecting genome-wide gene–environment interaction. *Genet Epidemiol* 2012;**36**:183–94.
45. Kooperberg C, Leblanc M. Increasing the power of identifying gene x gene interactions in genome-wide association studies. *Genet Epidemiol* 2008;**32**:255–63.

46. Murcray CE, Lewinger JP, Gauderman WJ. Gene–environment interaction in genome-wide association studies. *Am J Epidemiol* 2009;**169**:219–26.
47. Murcray CE, Lewinger JP, Conti DV, et al. Sample size requirements to detect gene–environment interactions in genome-wide association studies. *Genet Epidemiol* 2011;**35**:201–10.
48. Wu TT, Chen YF, Hastie T, et al. Genome-wide association analysis by lasso penalized logistic regression. *Bioinformatics* 2009;**25**:714–21.
49. Simino J, Sung YJ, Kume R, et al. Gene–alcohol interactions identify several novel blood pressure loci including a promising locus near SLC16A9. *Front Genet* 2013;**4**:277.
50. Rudolph A, Chang-Claude J, Schmidt MK. Gene–environment interaction and risk of breast cancer. *Br J Cancer* 2016;**114**:125–33.
51. Sung YJ, Winkler TW, de Las Fuentes L, et al. A large-scale multi-ancestry genome-wide study accounting for smoking behavior identifies multiple significant loci for blood pressure. *Am J Hum Genet* 2018;**102**:375–400.
52. WTCCC. Genome-wide association study of 14 000 cases of seven common diseases and 3000 shared controls. *Nature* 2007;**447**:661–78.
53. Wang Y, Sha N, Fang Y. Analysis of genome-wide association data by large-scale Bayesian logistic regression. *BMC Proc* 2009;**3**(Suppl 7):S16.
54. Wakefield J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am J Hum Genet* 2007;**81**:208–27.
55. Wakefield J. Bayes factors for genome-wide association studies: comparison with P-values. *Genet Epidemiol* 2009;**33**:79–86.
56. Yang J, Ferreira T, Morris AP, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* 2012;**44**:369, S361–75, 363.
57. Pan W, Chen YM, Wei P. Testing for polygenic effects in genome-wide association studies. *Genet Epidemiol* 2015;**39**:306–16.
58. Lin WY, Lee WC. Incorporating prior knowledge to facilitate discoveries in a genome-wide association study on age-related macular degeneration. *BMC Res Notes* 2010;**3**:26.
59. Xie Y, Pan W, Khodursky AB. A note on using permutation-based false discovery rate estimates to compare different analysis methods for microarray data. *Bioinformatics* 2005;**21**:4280–8.
60. Stephens M, Balding DJ. Bayesian statistical methods for genetic association studies. *Nat Rev Genet* 2009;**10**:681–90.
61. Smith JA, Ware EB, Middha P, et al. Current applications of genetic risk scores to cardiovascular outcomes and subclinical phenotypes. *Curr Epidemiol Rep* 2015;**2**:180–90.
62. Goldstein BA, Yang L, Salfati E, et al. Contemporary considerations for constructing a genetic risk score: an empirical approach. *Genet Epidemiol* 2015;**39**:439–45.
63. Lall K, Magi R, Morris A, et al. Personalized risk prediction for type 2 diabetes: the potential of genetic risk scores. *Genet Med* 2017;**19**:322–9.
64. Shi JX, Park JH, Duan JB, et al. Winner's curse correction and variable thresholding improve performance of polygenic risk modeling based on genome-wide association study summary-level data. *PLoS Genet* 2016;**12**:e1006493.
65. Chen CH, Yang JH, Chiang CWK, et al. Population structure of Han Chinese in the modern Taiwanese population based on 10 000 participants in the Taiwan Biobank project. *Hum Mol Genet* 2016;**25**:5321–31.
66. Hüls A, Ickstadt K, Schikowski T, et al. Erratum to: detection of gene–environment interactions in the presence of linkage disequilibrium and noise by using genetic risk scores with internal weights from elastic net regression. *BMC Genet* 2017;**18**:73.
67. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B* 1995;**57**:289–300.
68. Agier L, Portengen L, Chadeau-Hyam M, et al. A systematic comparison of linear regression-based statistical methods to assess exposome–health associations. *Environ Health Perspect* 2016;**124**:1848–56.
69. Jamieson MJ, Webster J, Phillips S, et al. The measurement of blood pressure: sitting or supine, once or twice? *J Hypertens* 1990;**8**:635–40.
70. Husemoen LLN, Fenger M, Friedrich N, et al. The association of ADH and ALDH gene variants with alcohol drinking habits and cardiovascular disease risk factors. *Alcohol Clin Exp Res* 2008;**32**:1984–91.
71. Albert PS, Ratnasinghe D, Tangrea J, Wacholder S. Limitations of the case-only design for identifying gene–environment interactions. *Am J Epidemiol* 2001;**154**:687–93.
72. Mukherjee B, Ahn J, Gruber SB, et al. Testing gene–environment interaction in large-scale case-control association studies: possible choices and comparisons. *Am J Epidemiol* 2012;**175**:177–90.
73. Hinney A, Hebebrand J. Polygenic obesity in humans. *Obes Facts* 2008;**1**:35–42.
74. Hinney A, Vogel CI, Hebebrand J. From monogenic to polygenic obesity: recent advances. *Eur Child Adolesc Psychiatry* 2010;**19**:297–310.
75. Deng AY. Genetic basis of polygenic hypertension. *Hum Mol Genet* 2007;**16**Spec No. 2):R195–202.
76. International Schizophrenia C, Purcell SM, Wray NR, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009;**460**:748–52.
77. Lewis CM, Vassos E. Prospects for using risk scores in polygenic medicine. *Genome Med* 2017;**9**:96.
78. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. *Bioinformatics* 2015;**31**:1466–8.
79. Fu ZM, Shrubsole MJ, Li GL, et al. Interaction of cigarette smoking and carcinogen-metabolizing polymorphisms in the risk of colorectal polyps. *Carcinogenesis* 2013;**34**:779–86.
80. Langenberg C, Sharp SJ, Franks PW, et al. Gene–lifestyle interaction and type 2 diabetes: the EPIC interact case-cohort study. *PLoS Med* 2014;**11**:e1001647.
81. Pollin TI, Isakova T, Jablonski KA, et al. Genetic modulation of lipid profiles following lifestyle modification or metformin treatment: the Diabetes Prevention Program. *PLoS Genet* 2012;**8**:e1002895.
82. Qi L, Cornelis MC, Zhang CL, et al. Genetic predisposition, Western dietary pattern, and the risk of type 2 diabetes in men. *Am J Clin Nutr* 2009;**89**:1453–8.
83. Qi QB, Chu AY, Kang JH, et al. Sugar-sweetened beverages and genetic risk of obesity. *N Engl J Med* 2012;**367**:1387–96.
84. Kapoor M, Chou YL, Edenberg HJ, et al. Genome-wide polygenic scores for age at onset of alcohol dependence and association with alcohol-related measures. *Transl Psychiatry* 2016;**6**:e761.
85. Nyholt DR. SECA: SNP effect concordance analysis using genome-wide association summary results. *Bioinformatics* 2014;**30**:2086–8.

86. Kass RE, Raftery AE. Bayes factors. *J Am Stat Assoc* 1995;**90**:773–95.
87. Sung YJ, de las Fuentes L, Schwander KL, et al. Gene-smoking interactions identify several novel blood pressure loci in the Framingham Heart Study. *Am J Hypertens* 2015;**28**: 343–54.
88. Chen Y, Nakura J, Jin JJ, et al. Association of the GNAS1 gene variant with hypertension is dependent on alcohol consumption. *Hypertens Res* 2003;**26**:439–44.
89. Abe M, Nakura J, Yamamoto M, et al. Association of GNAS1 gene variant with hypertension depending on smoking status. *Hypertension* 2002;**40**:261–5.
90. Cheng Q, Li YK, Lu F, et al. Interactions between ACYP2 genetic polymorphisms and environment factors with susceptibility to ischemic stroke in a Han Chinese population. *Oncotarget* 2017;**8**:97913–9.
91. Chen ZM, Hui JM, Liu LS, et al. CAST: randomised placebo-controlled trial of early aspirin use in 20 000 patients with acute ischaemic stroke. *Lancet* 1997; **349**:1641–9.
92. Sandercock P, Collins R, Counsell C, et al. The international stroke trial (IST): A randomised trial of aspirin, subcutaneous heparin, both, or neither among 19 435 patients with acute ischaemic stroke. *Lancet* 1997;**349**:1569–81.
93. Sharma VK. Elevated blood pressure in acute ischemic stroke—treat or leave? *Cerebrovasc Dis* 2016;**41**:101–2.