GxEsum: Genotype-by-environment model based on GWAS summary statistics[a]

Jisu Shin[1,2], S Hong Lee[1,2]

[1]Australian Centre for Precision Health, University of South Australia Cancer Research Institute, University of South Australia, Adelaide, SA 5000, Australia

[2]UniSA: Allied Health and Human Performance, University of South Australia, Adelaide, SA 5000, Australia

Highlights

- GxEsum can estimate whole-genome GxE using GWAS summary statistics
- Its computational efficiency is thousands of times higher than existing methods
- The type I error rate is controlled and the power is reasonable.
- GxEsum can be used for a whole-genome GxE scan for large number of traits

Method Overview

The proposed method to estimate whole-genome GxE interaction based on GWAS summary statistics (i.e., GxEsum) is an extension of LDSC[b] approach. For SNP effects modulated by an environment, the expected chi-square statistics (χ_j^2) is

$$\mathrm{E}[\chi_j^2|\ell_j] = \frac{N\sigma_{g_1}^2}{M} * \ell_j + 1 + 2(\sigma_{g_1}^2 + \sigma_{\tau_1}^2)$$

Where N is the number of individuals, M is the number of SNPs, $\sigma_{g_1}^2$ is the variance due to GxE, $\sigma_{\tau_1}^2$ is the variance due to residual heterogeneity or scale effects caused by residual-environment interaction (RxE), ℓ_j is the LD score at the variant j.

Precision and Computational efficiency

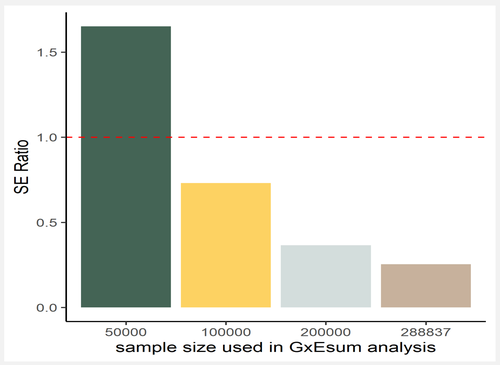

Figure 1. The ratio of standard error (SE) from GxEsum to that from RNM using UK Biobank data.
The SEs of GxE variance estimated from GxEsum with various sample sizes, and they were compared to that of the RNM with a sample size of 50,000.

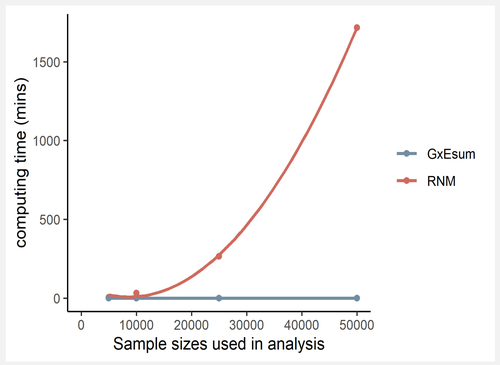

Figure 2. Computing time with various sample sizes used in GxEsum and RNM analyses.
As the sample size increases, the computing time of RNM (red) increases exponentially, while that of GxEsum (blue) is almost invariant (less than a minute).

Table 1. Estimates obtained from GxEsum analysis using real data.

Main trait	Environmental variable	Main genetic effect	The proportion of phenotypic variance explained by GxE effect	p-value for GxE
BMI	Age	0.216 (0.007)	0.004 (0.002)	1.86E-02
	Neuroticism score	0.216 (0.007)	0.007 (0.002)	**1.61E-05**
	Smoking	0.212 (0.007)	0.006 (0.002)	**3.18E-03**
	Physical activity	0.218 (0.007)	0.003 (0.001)	2.57E-02
	Alcohol consumption	0.216 (0.007)	0.003 (0.002)	5.98E-02
Hypertension	BMI	0.152 (0.008)	0.006 (0.002)	**2.09E-03**
	Waist-Hip ratio	0.154 (0.008)	0.005 (0.002)	3.21E-02
	Body fat percentage	0.151 (0.008)	0.008 (0.003)	2.66E-02

The p-value is from a Wald test for the estimated GxE variance not being different from zero. The estimates on the observed scale for the binary trait, hypertension, were transformed to those on the liability scale using Robertson transformation. Our results showed that the genetic effects of BMI were significantly modulated by neuroticism score and smoking, and the genetic effects of hypertension were modulated by BMI.

Table 2. Comparison between GxEsum and causality analyses

Exposures		Outcomes	p values for causal effects	p values from GxEsum
Morrison et al., 2020[c]				
Diastolic BP	→	Stroke	**1.10E-03**	9.56E-02
Systolic BP	→	Stroke	**4.10E-09**	8.30E-02
LDL cholesterol	→	Stroke	**4.60E-02**	8.86E-01
Smoking	→	Stroke	**2.30E-02**	6.03E-01
Diastolic BP	→	Type 2 diabetes (T2DM)	**2.00E-02**	3.51E-01
Systolic BP	→	Type 2 diabetes (T2DM)	**1.00E-02**	4.63E-01
BMI	→	Type 2 diabetes (T2DM)	**4.80E-03**	**2.01E-04**
Glucose	→	Type 2 diabetes (T2DM)	**1.30E-02**	1.00E+00
Davies et al., 2018[d]				
Left school Age after 15 years old	→	Hypertension	**1.9E-04**	5.18E-01
	→	Diabetes	**2.0E-03**	4.97E-01
	→	Stroke	**9.0E-03**	5.43E-01
	→	Heart Attack	**9.5E-04**	7.58E-01
	→	Ever smoked	**1.9E-15**	8.65E-02
	→	Currently Smoke	**1.7E-16**	7.86E-02
	→	Income over £18,000	**8.0E-15**	3.17E-01
	→	Income over £31,000	**4.1E-19**	**2.98E-02**
	→	Income over £52,000	**3.2E-20**	3.36E-01
	→	Income over £100,000	**2.5E-16**	**3.33E-04**
	→	Height	**3.6E-13**	1.73E-01
	→	BMI	**2.9E-10**	4.40E-01
	→	Diastolic BP	**1.0E-03**	3.42E-01
	→	Systolic BP	**1.2E-04**	2.11E-01
	→	Intelligence	**9.0E-15**	3.12E-01
	→	Alcohol Consumption	**1.3E-07**	8.65E-02
	→	Television viewing per day	**1.5E-16**	**2.27E-02**
	→	Moderate Exercise	**2.2E-06**	8.78E-01
	→	Vigorous Exercise	**2.0E-03**	3.36E-01

We compared p-value for causal association between covariates (exposures) and the main traits (outcomes), which were detected from previous causality analyses, with those from GxEsum analysis. We found that 4 out of 27 significant causal associations were explained by GxE interaction. Note that while the causality analyses (e.g. CAUSE) can detect the association between phenotypic means (of the main traits) and covariate values, GxEsum can estimate the heterogeneity of genetic variance across different lifestyle values. The bold fonts are for significant p-values, and the arrows indicate the direction of causal effects.

References

a. *Shin, J & Lee, SH 2020, 'GxEsum: genotype-by-environment interaction model based on summary statistics', bioRxiv, p. 2020.05.31.122549*
b. *Bulik-Sullivan, BK, Loh, PR, Finucane, HK, Ripke, S, Yang, J, Schizophrenia Working Group of the Psychiatric Genomics, C, Patterson, N, Daly, MJ, Price, AL & Neale, BM*
c. *Morrison, J., N. Knoblauch, J. H. Marcus, M. Stephens and X. He (2020). "Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics." Nature Genetics **52**(7): 740-747.*
d. *Davies, N. M., M. Dickson, G. Davey Smith, G. J. van den Berg and F. Windmeijer (2018). "The causal effects of education on health outcomes in the UK Biobank." Nature Human Behaviour **2**(2): 117-125.*