The American Journal of Human Genetics, Volume *111*

Supplemental information

Many roads to a gene-environment interaction

Kenneth E. Westerman and Tamar Sofer

Supplemental Table S1: Phenomenon-specific statistical approaches.

Phenomenon	Assessment strategy	Relevant statistical model and/or software tools
Phenomenon 1 (functional)	Perform standard product term regression-based test; rule out other phenomena using associated statistical tests and domain knowledge	• Some standard GWAS software programs (e.g., PLINK[1]) • Interaction-specific GWAS programs (e.g., GEM[2]) • Regression using standard statistical analysis programs
Phenomenon 2 (nonlinear mediator)	Test for nonlinearity of the M-Y relationship (if M is known and measured)	• If known and measured, test for nonlinearity of the mediator (e.g., using likelihood ratio test of spline vs. simple linear model[3])
	Test for nonlinearity of the E-Y relationship	• E.g., likelihood ratio test comparing nested linear and non-linear models
Phenomenon 3 (G-E correlation plus nonlinearity)	Test for nonlinearity of the E-Y relationship	• E.g., likelihood ratio test comparing nested linear and non-linear models
Phenomenon 4 (Heterogeneous variability)	Test for dispersion versus artifactual vQTL from mean-variance relationship	• Dispersion test from Young et al.[4]
	Test for specific GxE versus general dispersion	• Scaling model test from Domingue et al.[5]
Phenomenon 5 (Heterogeneous exposure measurement)	Domain knowledge	• Difficult to test statistically

Supplemental Note: Simulation setup.

1. Simulate G (N = 10,000) with MAF = 0.25: $G \sim binomial(2, 0.25)$
2. Simulate E, M (where applicable), and Y as follows:

Phenomenon	Exposure (E)	Mediator (M)	Outcome (Y)
Phenomenon 1: Functional	$N(5,1)$		$N(E + G * E, 1)$
Phenomenon 2: Nonlinear mediator	$N(0,1)$	$N(G + E, 0.1)$	$N(\sqrt[3]{M + 4}, 0.1)$
Phenomenon 3: G-E correlation with nonlinearity	$N(G,1)$		$N(\sqrt[3]{E + 4}, 0.1)$
Phenomenon 4: Heterogeneous variability	$N(5,1)$		$Y^* \sim N(E_{std}, 1)$ $Y \sim N(G * Y^*, 1)$
Phenomenon 5: Heterogeneous measurement	$E \sim N(5,1)$ $E_m \sim N(E, e^{0.5*G})$		$Y \sim N(E, 1)$

E_{std} denotes a standardized (mean zero and unit variance) version of E. E_m denotes the measured exposure, which is then used for the interaction test.

3. Test for statistical interaction using the following regression model: $Y \sim G + E + G * E$

Regression results for these scenarios are presented in the following table:

Phenomenon	Regression term	Effect estimate	P-value
Phenomenon 1: Functional	G	0.103	0.224
	E	1.021	<0.001
	G:E	0.98	<0.001
Phenomenon 2: Nonlinear mediator	G	0.12	<0.001
	E	0.137	<0.001
	G:E	-0.017	<0.001
Phenomenon 3: G-E correlation with nonlinearity	G	0.008	0.001
	E	0.136	<0.001
	G:E	-0.014	<0.001
Phenomenon 4: Heterogeneous variability	G	-4.934	<0.001
	E	0.004	0.822
	G:E	0.981	<0.001
Phenomenon 5: Heterogeneous measurement	G	1	<0.001
	E_measured	0.485	<0.001
	G:E_measured	-0.197	<0.001

References

1. Chang, C.C., Chow, C.C., Tellier, L.C.A.M., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience *4*, 7. 10.1186/s13742-015-0047-8.
2. Westerman, K.E., Pham, D.T., Hong, L., Chen, Y., Sevilla-González, M., Sung, Y.J., Sun, Y. V, Morrison, A.C., Chen, H., and Manning, A.K. (2021). GEM: scalable and flexible gene–environment interaction analysis in millions of samples. Bioinformatics. 10.1093/bioinformatics/btab223.
3. Harrell, F.E. (2015). General Aspects of Fitting Regression Models. In 10.1007/978-3-319-19425-7_2.
4. Young, A.I., Wauthier, F.L., and Donnelly, P. (2018). Identifying loci affecting trait variability and detecting interactions in genome-wide association studies. Nat Genet *50*, 1608–1614. 10.1038/s41588-018-0225-6.
5. Domingue, B.W., Kanopka, K., Mallard, T.T., Trejo, S., and Tucker-Drob, E.M. (2022). Modeling Interaction and Dispersion Effects in the Analysis of Gene-by-Environment Interaction. Behav Genet *52*. 10.1007/s10519-021-10090-8.