LINK TO ORIGINAL ARTICLE

Empirical tests for compositional epistasis

Tyler J. VanderWeele

In her Review article (Detecting gene–gene interactions that underlie human diseases. *Nature Rev. Genet.* **10**, 392–404 (2009))[1] and elsewhere[2], Cordell has argued that statistical tests for interactions are of limited use for elucidating epistasis in a more biological sense of the term — that is, when the effect of a particular genetic variant is masked by a variant at another locus[3,4] ('compositional epistasis'[5]). Suppose two genetic factors, X_1 and X_2, are binary indicators for genotypes at loci *A* and *B*, respectively, so that $X_1 = 0$ denotes genotype *a/a* or *a/A* and $X_1 = 1$ denotes genotype *A/A* and, similarly, $X_2 = 0$ denotes genotype *b/b* or *b/B* and $X_2 = 1$ denotes genotype *B/B*. Suppose *D* is a dichotomous trait. Epistasis in the sense of masking[2,3] would be present if there were individuals for whom locus *A* had no effect without the *B/B* genotype at locus *B*, as in TABLE 1.

A statistical model that accommodates statistical interaction might be formulated as:

$$P(D = 1|X_1 = x_1, X_2 = x_2) = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_1 x_2$$

in which *P* denotes the probability and α_{0-3} are the model parameters.

Cordell[1,2] points out that statistical interaction tests (for example, $\alpha_3 > 0$) do not generally allow for conclusions about epistasis in the more biological sense of masking.

There are, however, relationships between empirical data patterns and compositional epistasis, as in TABLE 1, that have not been previously noted and that can be used to derive non-standard interaction tests to test empirically for such epistasis[6]. Provided that the effects of the genetic factors are not confounded by stratification or admixture (or if appropriate controls have been made for these[7–9]), if $\alpha_3 > 2\alpha_0$ there must be some individuals with the phenotype response pattern of TABLE 1 — that is, individuals in which compositional epistasis is present[6]. If one of the genetic factors has an effect that is in the same direction for all individuals (that is, not causative for some individuals and preventive for others), then $\alpha_3 > \alpha_0$ implies compositional epistasis[6]. Only when both factors have effects that are in the same direction for all individuals does the standard interaction test $\alpha_3 > 0$ imply compositional epistasis[6]. In most cases, there will be insufficient prior knowledge to make these 'monotonicity' assumptions, and one would have to test $\alpha_3 > 2\alpha_0$ to detect epistasis in the sense of masking.

Suppose instead a log-linear model is used:

$$\log_e\{P(D = 1|X_1 = x_1, X_2 = x_2)\} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2.$$

If $\beta_1 \geq 0$ and $\beta_2 \geq 0$, then $\beta_3 > \log_e(3)$ implies the existence of individuals with the phenotype response pattern of TABLE 1 (that is, compositional epistasis)[6]. If it can be assumed that the effect of at least one of the factors is monotonic, as described above, $\beta_3 > \log_e(2)$ with $\beta_1 \geq 0$ and $\beta_2 \geq 0$ suffices for compositional epistasis; if it can be assumed the effects of both factors are monotonic, $\beta_3 > 0$ suffices.

These empirical tests for compositional epistasis presuppose that the phenotype probabilities $P(D = 1|X_1 = x_1, X_2 = x_2)$ reflect the true effects of the genes; a similar approach to that described above could be used in conjunction with techniques to control for confounding by stratification or admixture[7–9]. The conditions for compositional epistasis are related to but stronger than the notion of 'synergism' in Rothman's sufficient-cause framework[6,10–12].

Table 1 | **Compositional epistasis**

		Genotype at locus *B*	
		b/b* or *b/B	***B/B***
Genotype at locus *A*	***a/a* or *a/A***	0	0
	A/A	0	1

An example of potential phenotypes for a particular individual that might result from different genotypes at two loci that exhibit 'compositional epistasis'[*5] in the sense of masking[1–3].

Compositional epistasis is arguably a more biological notion of epistasis than is 'statistical epistasis', but even compositional epistasis does not necessarily imply the physical molecular interaction of one protein with another ('functional epistasis'[5]). The remarks here apply also to settings in which the genetic factors are considered to have three relevant levels. See a fuller report[6] for additional discussion.

Tyler J. VanderWeele is at the Departments of Epidemiology and Biostatistics, Harvard School of Public Health, 677 Huntington Avenue, Boston, Massachusetts 02115, USA.
e-mail: tvanderw@hsph.harvard.edu

doi:10.1038/nrg2579-c1
Published online 18 January 2010

1. Cordell, H. J. Detecting gene–gene interactions that underlie human diseases. *Nature Rev. Genet.* **10**, 392–404 (2009).
2. Cordell, H. J. Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. *Hum. Mol. Genet.* **11**, 2463–2468 (2002).
3. Bateson, W. *Mendel's Principles of Heredity* (Cambridge Univ. Press, 1909).
4. Cordell, H. J. & Clayton, D. G. Genetic epidemiology 3 — genetic association studies. *Lancet* **366**, 1121–1131 (2005).
5. Phillips, P. C. Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Rev. Genet.* **9**, 855–867 (2008).
6. VanderWeele, T. J. Epistatic interactions. *Stat. Appl. Genet. Mol. Biol.* **9**, Article 1 (2010).
7. Pritchard, J. K. *et al.* Association mapping in structured populations. *Am. J. Hum. Genet.* **67**, 170–181 (2000).
8. Hoggart, C. J. *et al.* Control of confounding of genetic associations in stratified populations. *Am. J. Hum. Genet.* **72**, 1492–1504 (2003).
9. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genet.* **38**, 904–909 (2006).
10. Rothman, K. J. Causes. *Am. J. Epidemiol.* **104**, 587–592 (1976).
11. VanderWeele, T. J. & Robins, J. M. Empirical and counterfactual conditions for sufficient cause interactions. *Biometrika* **95**, 49–61 (2008).
12. VanderWeele, T. J. Sufficient cause interactions and statistical interactions. *Epidemiol.* **20**, 6–13 (2009).