



Supplementary Figure S1 | **Sample-size requirements for gene–environment interactions.** Sample-size requirements are shown for gene–environment (G×E) interactions in candidate-gene studies (**a**) and genome-wide association (GWA) studies (**b,c**). In **a** and **b**, the case–control design is assumed to have equal numbers of cases and unrelated controls (N), varying exposure prevalence (p) and varying minor-allele frequencies (q). The dashed line shows $p = q = 0.4$; the green line shows $p = 0.4, q = 0.1$ or $p = 0.1, q = 0.4$; and the light blue line shows $p = q = 0.1$. Panel **a** is for the two-sided significance level $\alpha = 0.05$, and panel **b** is for the two-sided significance level $\alpha = 10^{-7}$; the power for both is $1 - \beta = 80\%$. Both panels assume a log-additive genetic model with no main effects (calculated using the Quanto program¹). These panels show that very large sample sizes are required for detecting modestly sized interaction effects, particularly in GWA studies because of the very high significance levels required. Panel **c** shows the conditional power in a GWA study for detecting either a G×E interaction (dark blue line) or a genetic effect in an exposure subgroup at a significance level of $\alpha = 2$ (red line), given that the main effect (dashed line) is not significant. Each subpanel in **c** differs in the non-centrality parameter for the main effect and plots the conditional power as a function of the interaction effect for the fixed variables $p = 0.25, q = 0.1$. The plot demonstrates that there are plausible combinations of main effect and interaction sizes for which the latter would be discoverable even though the former is not. The region where the power to detect interactions in the absence of main effects is greatest is where the lines cross (positive in one stratum, negative in the other).

1. Gauderman, W. J. Sample size requirements for matched case–control studies of gene–environment interaction. *Stat. Med.* **21**, 35–50 (2002).