eAppendix
"Family-based gene-by-environment interaction studies: revelations and remedies"
Shi, Umbach and Weinberg 2011.

I. Simulation study of haplotype-based methods:

We simulated a null scenario under population stratification to evaluate several haplotype-based methods. Because some methods can only handle a small number of SNPs we simulated a 2-SNP, 3-haplotype scenario. We simulated a no-interaction null scenario with a dichotomous exposure and two equal-sized subpopulations; each subpopulation had all haplotypes in Hardy-Weinberg equilibrium (HWE) and the same baseline risk of disease. The exposure prevalences were 0.05 and 0.4 in the two subpopulations respectively. We assumed that the disease-causing mutation was not typed and resided in different haplotypes in the two subpopulations. The frequency of the mutation-carrying haplotype was 0.4 in both subpopulations. The haplotypes that were not associated with the risk allele occurred in the same relative frequencies as in HapMap. We set (R_1, R_2, I_1, I_2, R_e) = (2, 3, 1, 1, 2) in each subpopulation. We compared the following methods GEI-TRIMM [1], PCPH [2], Unphased [3] and Pseudocontrol [4]. Table A1 shows the Type I error rates of these tests.

eTable 1. Type I error rates for haplotype-based methods

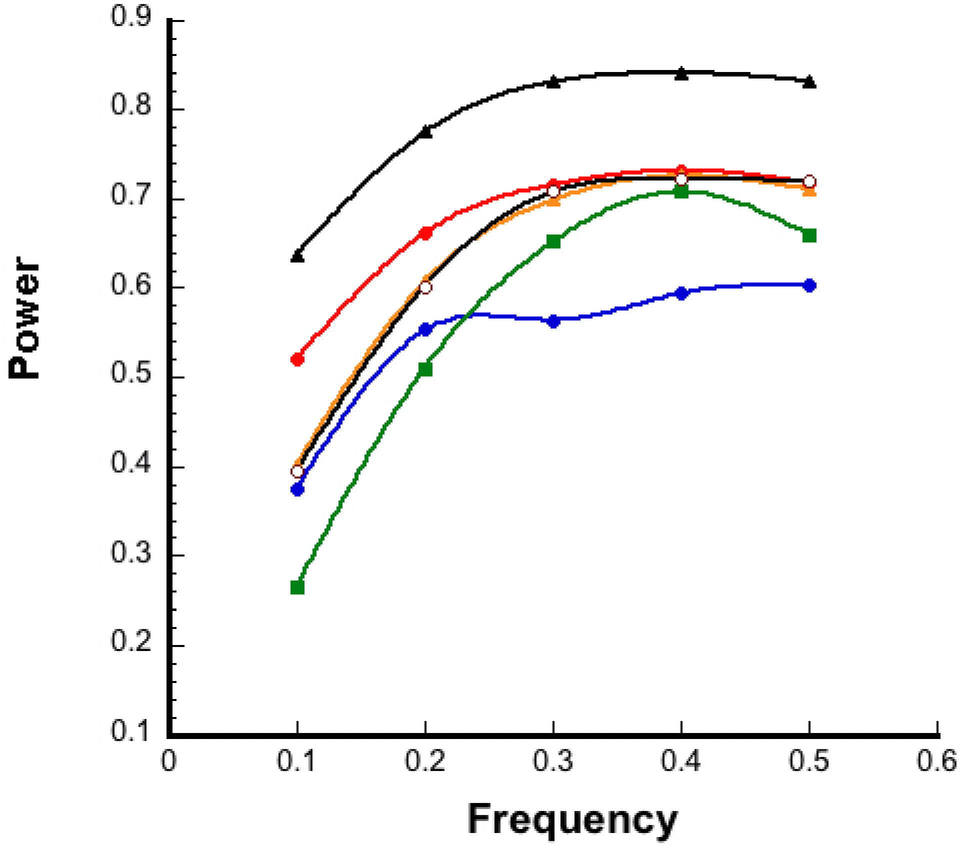
Method	Pseudocontrol	GEI-TRIMM	Unphased	PCPH
Type I Error Rate	0.717	0.850	0.782	0.891

II. Additional simulation studies to assess power

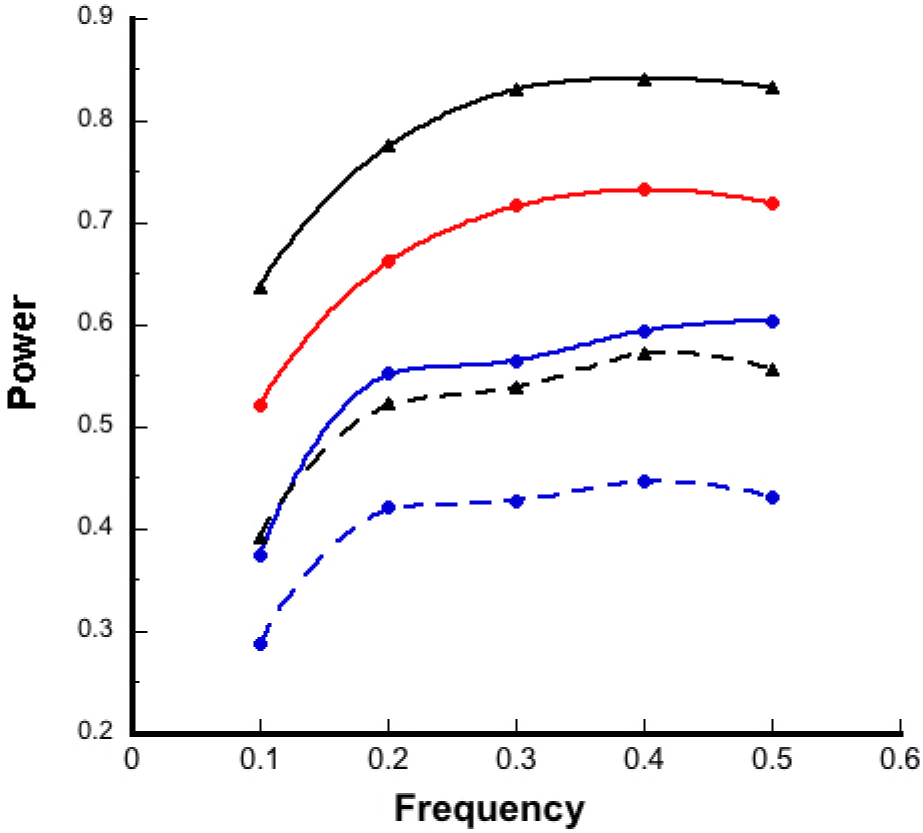
We used several additional simulation conditions to assess the performance of our tests. The models that we fit saturated the genetic main effects and enforced a log-additive GxE interaction.

eFigure 1. Power of GxE tests under a dominant genetic main effect scenario (R_1, R_2, I_1, I_2, R_e) set at (2, 2, 1.5, 2.25, 2) for a homogenous population: a) for unadjusted models; b) for models with (dashed line) and without (solid line) $G\overline{E}$-adjustment. The test and the associated symbols are: tetrad (black triangles), triad (orange triangles), case-sib (blue dots), Chatterjee's method (red dots), QPL (white dots) and FBAT-I (green square). After $G\overline{E}$-adjustment, the power of the tetrad design coincides with that of the Chatterjee's method; therefore the dashed line with black triangles represents both methods.

a)

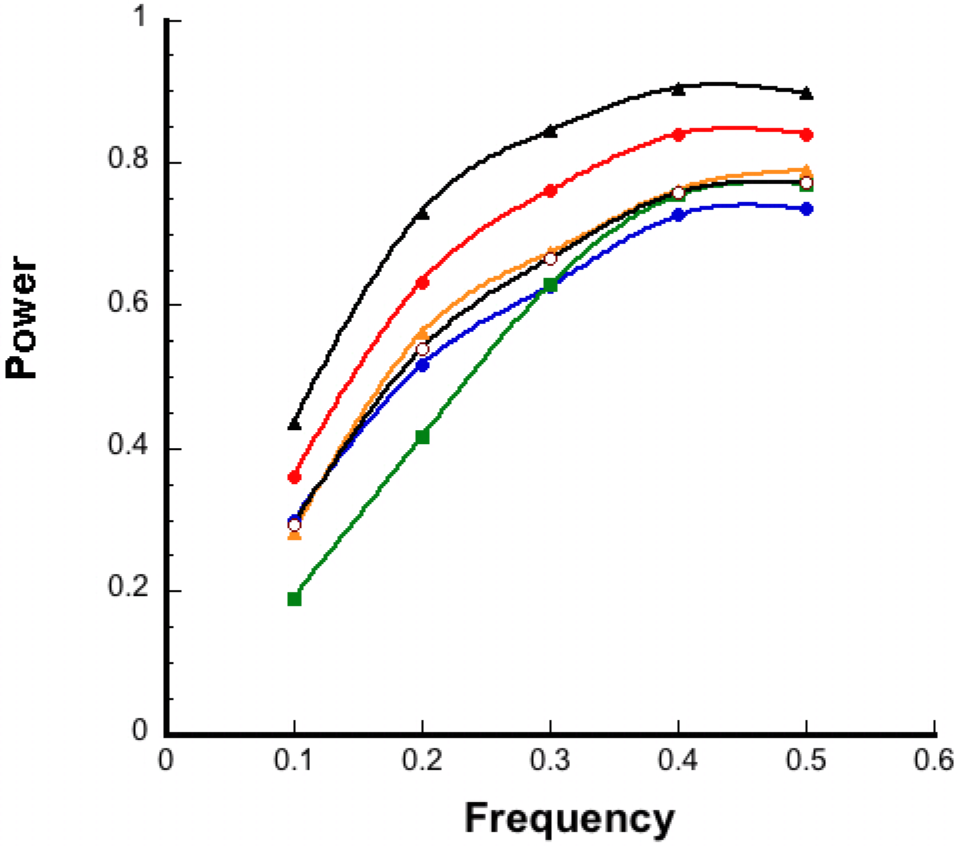


b)

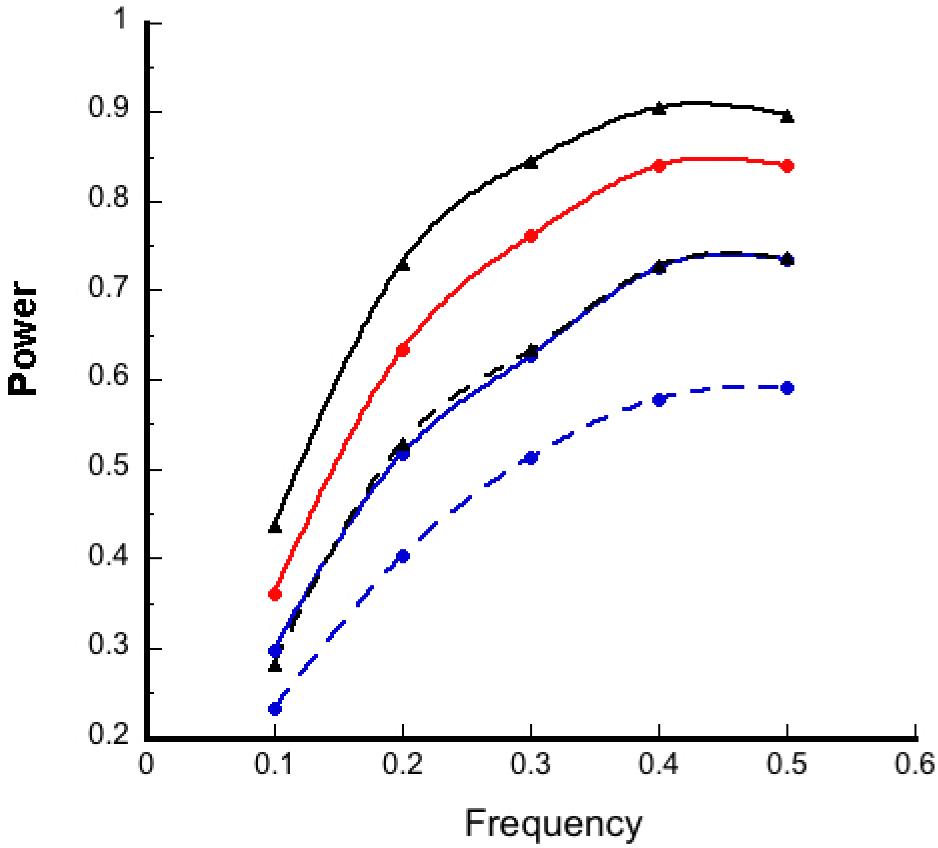


eFigure 2. Power of GxE tests under a pure interaction scenario (R_1, R_2, I_1, I_2, R_e) = (1, 1, 1.5, 2.25, 1) for a homogenous population: a) for unadjusted models; b) for models with (dashed line) and without (solid line) $G\overline{E}$-adjustment. The test and the associated symbols are: tetrad (black triangles), triad (orange triangles), case-sib (blue dots), Chatterjee's method (red dots), QPL (white dots) and FBAT-I (green square). After $G\overline{E}$-adjustment, the power of the tetrad design coincides with that of the Chatterjee's method; therefore the dashed line with black triangles represents both methods.

a)



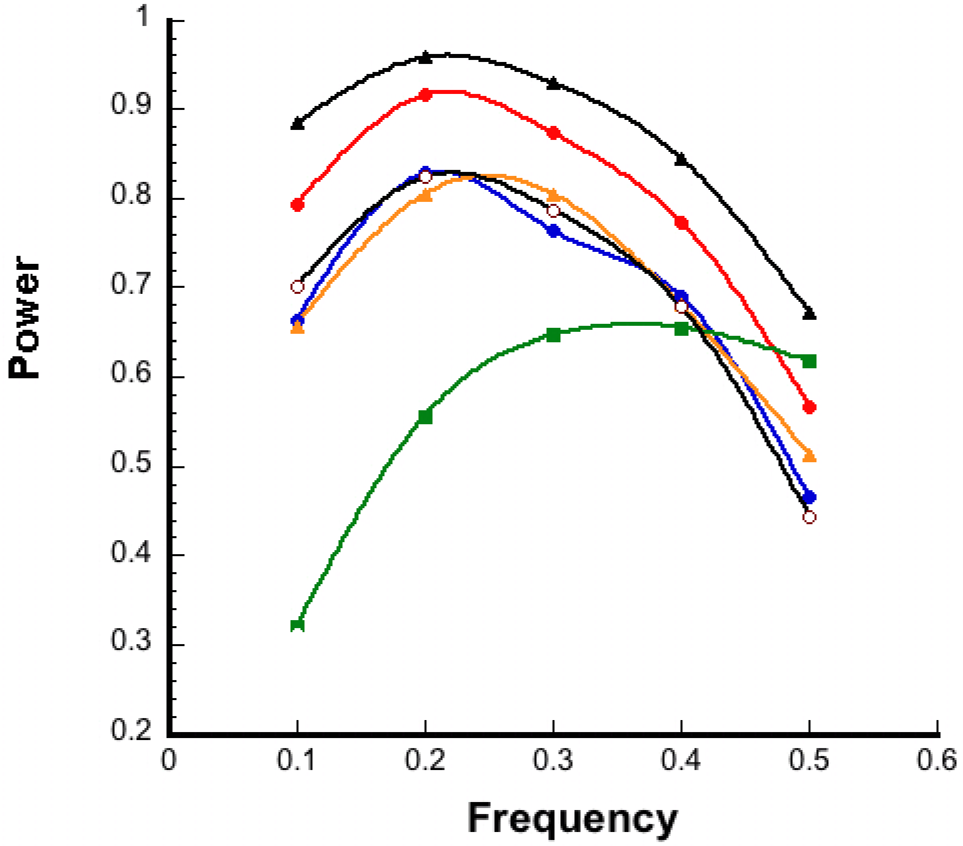
b)



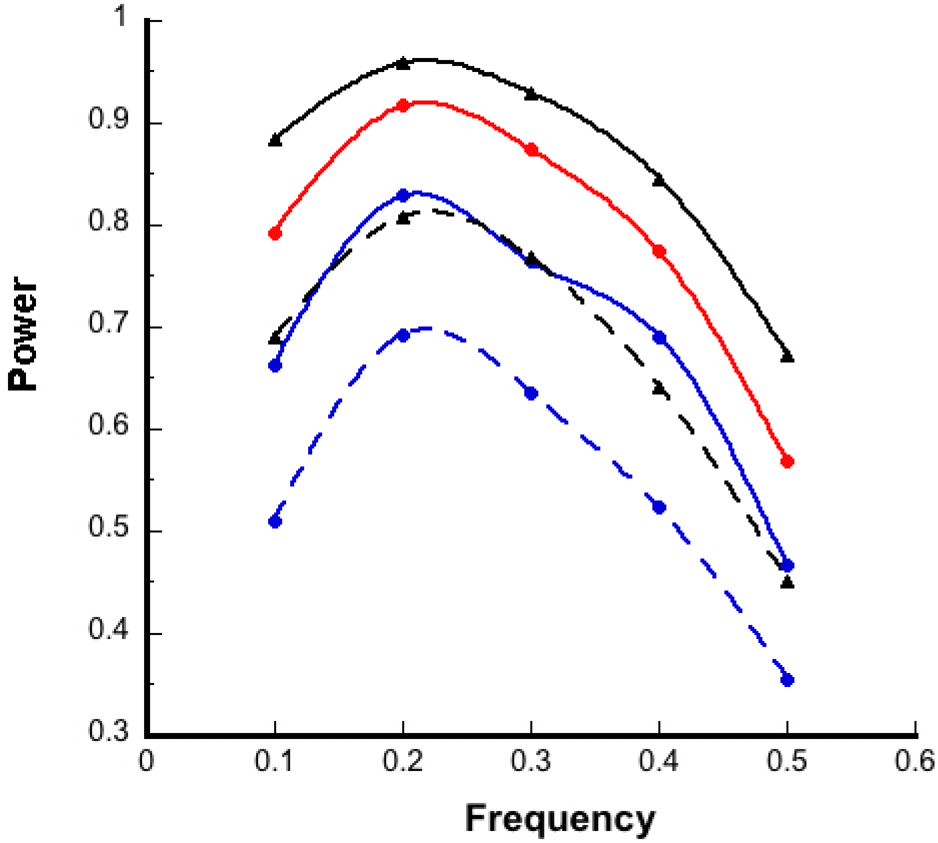
In the simulations above and those in the paper we simulated a log-additive interaction effect, which was the same as the model we used to fit the data. Next we show via simulation study that our models still perform reasonably well when the underlying interaction effect was not log-additive. We simulated scenarios where the interaction effect was dominant or recessive.

eFigure 3 Power of GxE tests under a dominant interaction effect scenario (R_1, R_2, I_1, I_2, R_e) = (1,1,2,2,1)) for a homogenous population when the data were fit with a log-additive interaction model: a) for unadjusted models; b) for models with (dashed line) and without (solid line) $G\overline{E}$-adjustment. The test and the associated symbols are: tetrad (black triangles), triad (orange triangles), case-sib (blue dots), Chatterjee's method (red dots), QPL (white dots) and FBAT-I (green square). After $G\overline{E}$-adjustment, the power of the tetrad design coincides with that of the Chatterjee's method; therefore the dashed line with black triangles represents both methods.

a)

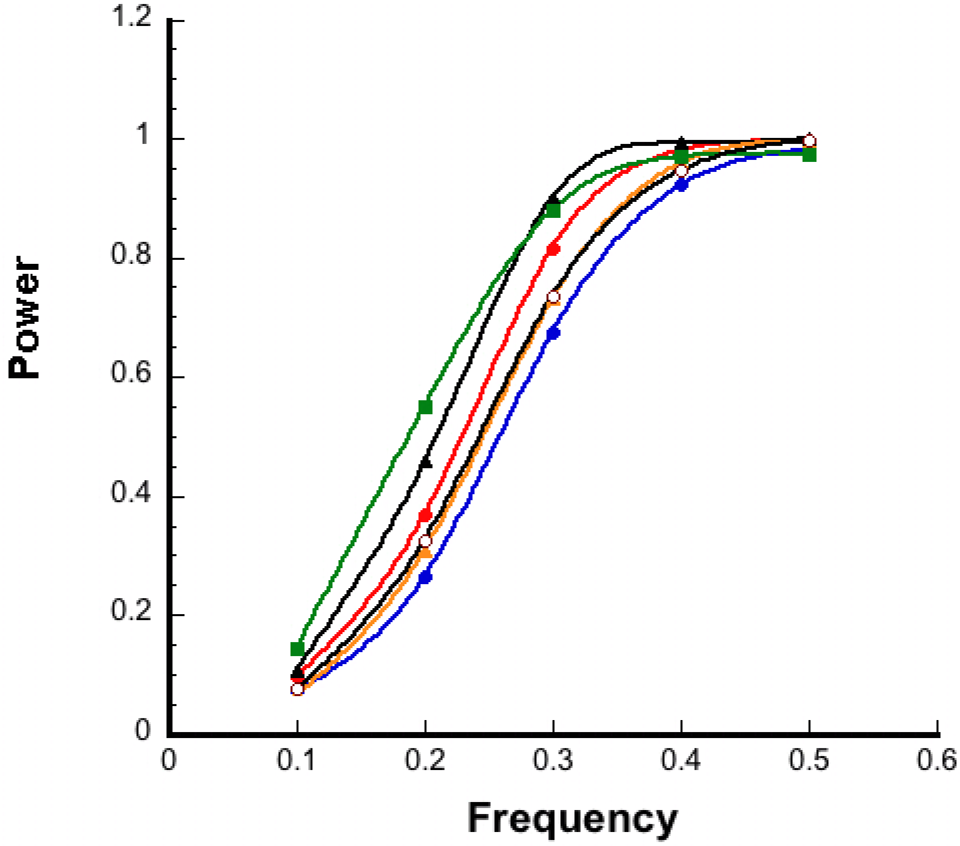


b)

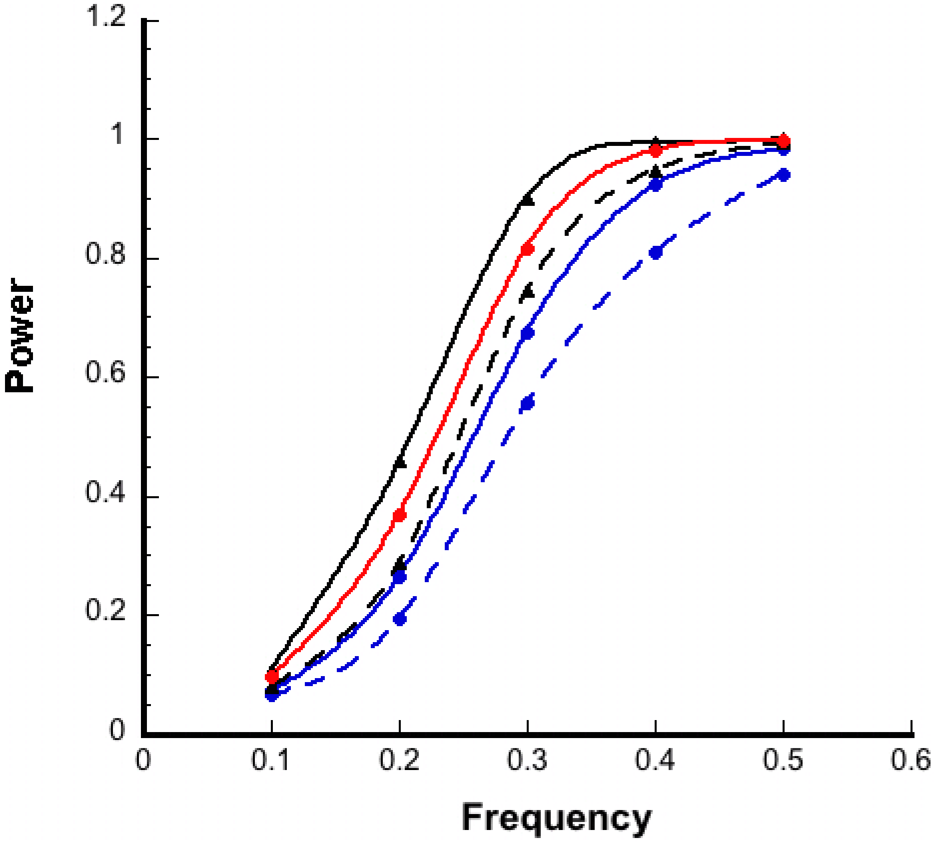


eFigure 4 Power of GxE tests under a recessive interaction effect scenario (R_1, R_2, I_1, I_2, R_e) = (1,1,1,3,1)) for a homogenous population when the data were fit with a log-additive interaction model: a) for unadjusted models; b) for models with (dashed line) and without (solid line) $G\overline{E}$-adjustment. The test and the associated symbols are: tetrad (black triangles), triad (orange triangles), case-sib (blue dots), Chatterjee's method (red dots), QPL (white dots) and FBAT-I (green square). After $G\overline{E}$-adjustment, the power of the tetrad design coincides with that of the Chatterjee's method; therefore the dashed line with black triangles represents both methods.

a)



b)



References

1. Shi, M., Umbach, D.M., and Weinberg, C.R. (2010). Testing Haplotype-Environment Interactions Using Case-Parent Triads. Hum Hered 70, 23-33.

2. Allen, A.S., and Satten, G.A. (2007). Inference on haplotype/disease association using parent-affected-child data: the projection conditional on parental haplotypes method. Genet Epidemiol 31, 211-223.

3. Dudbridge, F. (2008). Likelihood-based association analysis for nuclear families and unrelated subjects with missing genotype data. Hum Hered 66, 87-98.

4. Cordell, H.J., Barratt, B.J., and Clayton, D.G. (2004). Case/pseudocontrol analysis in genetic association studies: A unified framework for detection of genotype and haplotype associations, gene-gene and gene-environment interactions, and parent-of-origin effects. Genet Epidemiol 26, 167-185.