High Dimensional Endophenotype Ranking in the Search for Major Depression Risk Genes

Supplemental Information

Quantitative Genetic Analyses

All analyses were conducted with SOLAR (1), which employs maximum likelihood variance decomposition methods to determine the relative importance of genetic and environmental influences by modeling the covariance among family members as a function of genetic proximity. For a single trait, the observed phenotypic covariance matrix of a pedigree is modeled as the sum of two components, the additive genetic variance that is structured by the observed coefficient of relationship matrix, and a random environmental variance that reflects the person-specific environment and is structured by an identity matrix. Additional variance components can be simply added. For example, allowing a matrix containing the pair-wise expected genome-wide probability that two individuals share 2 alleles identical-by-descent will accommodate a dominance genetic variance component. Similarly, a shared environmental variance component can be defined that is structured by a matrix reflecting household sharing at the time of measurement (or any other time for which such sharing is available). More complex interaction-based variance components such as those due to epistasis are also easily incorporated. Given a model for the phenotypic covariance amongst individuals and a model for mean effects (generally including an overall mean parameter and a set of covariate effects such as sex and age), the likelihood of the model is calculated assuming a multivariate normal distribution of the focal trait within pedigrees. For quantitative traits, we always use direct normalization transformation such as an inverse Gaussian transformation to minimize deviation from this assumption. Maximum likelihood parameter estimation (along with parameter standard errors) is performed using standard numerical optimization methods. For discrete traits, the likelihood is more complex involving high-dimensional integrals of the multivariate normal distribution. We accurately approximate this integration using the methods discussed in Williams

et al. (2). Tests of variance component parameters (such as heritability, the proportion of variance due to the additive genetic component) are performed using standard likelihood ratio tests in which the ln likelihood of the null model (in which the focal variance component is forced to be zero) is compared to that of the alternative model (in which the focal variance component is explicitly estimated from the data). Twice the difference in these ln likelihoods yields a likelihood ratio test statistic that is asymptotically distributed as a 50:50 mixture of a chi-square variate with one degree of freedom and a point mass at zero. These univariate quantitative genetic variance component models can be easily adapted for this multivariate case by replacing univariate variance components with covariance matrices that include terms parameterizing the genetic and environmental correlations between the two traits.

1. Almasy L, Dyer TD, Blangero J (1997): Bivariate quantitative trait linkage analysis: pleiotropy versus co-incident linkages. *Genet Epidemiol*. 14:953-958.

2. Williams JT, Van Eerdewegh P, Almasy L, Blangero J (1999): Joint multipoint linkage analysis of multivariate qualitative and quantitative traits. I. Likelihood formulation and simulation results. *Am J Hum Genet*. 65:1134-1147.