**Bayesian integrated analyze analysis of multiple types of rare variants to infer risk genes for schizophrenia**

*Hoang Nguyen, Douglas Ruderfer, Menachem Fromer, Pamela Sklar, Shaun Purcell, Xin He, Patrick Sullivan, Eli Stahl*

Integrating rare variation from family and case/control studies has successfully, implicated specific genes contributing to risk of autism spectrum disorder (ASD). In schizophrenia, however, while sets of genes have been implicated through study of rare variation, very few individual risk genes have been identified. Here, we apply a hierarchical Bayesian modeling of rare variation in schizophrenia and describe the proportion of risk genes and distribution of risk variant effect sizes across multiple variant annotation categories.

Briefly, we employed the same model used previously in ASD studies (De Rubeis, Silvia, et al., 2014; He, Xin, et al. , 2013). However, to simplify the complexity of the model, an approximation for the case-control model in which case variants are conditional on total counts is used. In addition, instead of using only one class of de novo mutation as in the previous studies, all classes of de novo mutations and case-control variants are used to infer genetic parameters. These parameters are estimated using a Markov Chain Monte Carlo method.

We applied this method to 1,024 trios and 4,954 cases/6,239 controls. We defined four variant annotation categories: disruptive (nonsense, frameshift, essensial splice site mutations) and missense damaging *de novos* (predicting damaging by all seven algorithms: SIFT, Polyphen2\_HDIV, Polyphen2\_HVAR, LRT, Mutation Taster, Mutation Assessor and PROVEAN), private disruptive (not seen in ExAC) and missense damaging case/control singletons.

We estimated that 5.8% of approximate 20,000 estimated genes are risk genes (95% credible interval 2-10%), with mean effect sizes (95% CIs) of 17.46 (7.45- 29.4) for disruptive *de novos*, 2.34 (1-4.9) for missense damaging *de novos*, 2.06 (1.00-3.72) for disruptive case/control private singletons, and 1.82 (1.00-3.01) for missense damaging case/control private singletons. The model also yields Bayes factors for each gene, and we estimate FDRs for these. Our analysis identified only one gene with FDR<0.05, which is the previously identified gene *SETD1A*. We further analyzed the top 100 genes, with FDR<=0.594, for enrichment in several candidate gene sets. Significant results are observed in gene sets previously implicated in schizophrenia (including in a subset of these data); targets of the fragile X mental retardation protein (FMRP, p = 6.7x10-5), haplo-insufficient (p = 6.7x10-5), the RNA binding proteins Rbfox1/2/3 (p = 6.7x10-5), constrained (p = 6.7x10-5 ), *de novo mutations* in ASD (p = 6.7x10-5), synaptic (p = 1.2x10-4), *de novo* mutations in congenital heart disease (p = 5x10-4), *de novo* mutations in epileptic encephalopathies: (p = 9x10-4). Overall, our results replicate previous studies for known gene sets as well as the single gene SETD1A indicating the robustness of the approach. We anticipate this approach will improve our power to detect schizophrenia risk genes as more data is included.