The new one: using De novo + Transmitted + Case-control

**Integrated 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. 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, essential splice site mutations) and missense damaging *de novos* (predicting damaging by seven algorithms), disruptive and missense damaging case/control singletons. We estimated that 8.4% of approximate 20,000 estimated genes are risk genes (95% credible interval 3.5-16%), with mean effect sizes (95% CIs) of 14.21 (5.04- 25.65) for disruptive *de novos*, 1.99 (1-3.99) for missense damaging *de novos*, 1.79 (1-2.94) for disruptive case/control singletons, and 1.56 (1-2.46) for missense damaging case/control singletons. Our analysis identified only three gene with FDR<0.1, SETD1A, TAF13 (FDR<0.05) and RB1CC1. We further analyzed the top 100 genes, with FDR<=0.496, 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): FMRP, Rbfox1/2/3, constrained, *de novo mutations* in ASD (all p values less than 7.8x10-4), and synaptome (p = 1.3x10-3). 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.

==============We only have 2500 words, please don’t add more words, should we add Giulio Genovesein the authors’ list?

Below is the current one on ASHG website (using De novo + Private (Transmitted + Case/Control)

**Bayesian integrated 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. 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, essential splice site mutations) and missense damaging *de novos* (predicting damaging by seven algorithms), 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 (1-3.71) for disruptive case/control private singletons, and 1.82 (1-3.01) for missense damaging case/control private singletons. 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): FMRP, haplo-insufficient, Rbfox1/2/3, constrained, *de novo mutations* in ASD (all p values = 6.7x10-5) and synaptic (p = 1.2x10-4). 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.