M2. RARE VARIANT DETECTION IN COMPLEX DISORDERS USING THE BIRTHDAY MODEL

Yael Berstein[1], Shane McCarthy[2], Melissa Kramer[1], Peter Zandi[3], Fernando Goes[3], James Potash[4], Richard McCombie[1]

[1]Cold Spring Harbor Laboratory, [2]Regeneron Pharmaceuticals, Inc, [3]Johns Hopkins University, [4]The University of Iowa Carver College of Medicine

Background Exome sequencing is a powerful technique for the identification of disease-causing genes. A number of disease genes with mendelian inheritance were already identified through this method. It nonetheless remains a challenge to leverage exome sequencing for the study of complex disorders, e.g. schizophrenia and bipolar disorder. The genetic and phenotypic heterogeneity of the data is a barrier to the detection of causative genes in complex disorders. For example, the aggregation of different rare variants associated with a given disease can make the identification of causal genes statistically challenging. Large sample sizes with >10,000 individuals were suggested as a mean to improve statistical power, although this may be sometimes unfeasible due to cost and logistics constrains. Therefore, new methods for detecting rare variants are imperative to identify causative genes of complex disorders.

Methods Here we propose a probabilistic method to predict causative rare variants. This model is based on general analysis of coincidences based on a popular probabilistic problem: the birthday problem. Analogically, we consider the probability of samples sharing a variant, as the chance of individuals sharing the same birthday.

Results We evaluated the performance of our method through simulations for identifying causal rare variants in complex disorders. We investigated the effect of the parameters of our model, providing guidelines for its use and interpretation of the results. We implemented this probabilistic method to published data on autism spectrum disorder, hypertriglyceridemia, schizophrenia, and also on a current case-control study on bipolar disorder. The top results based on our method were Sanger validated. Several genes in the top results were associated with psychiatric disorders in published studies.

Discussion Given that the core probability based on the birthday model is very sensitive to low recurrence, the method successfully detect rare variants, which generally do not provide enough signal in existing statistical tests. The simplicity of the model allows quick interpretation of genomic data, enabling users to select gene candidates for further biological validation of specific mutations.

Disclosure: Nothing to Disclose.