GBSBC

Genetics, Bioinformatics, & Systems Biology Colloquium

Genetics, Bioinformatics, & Systems Biology Colloquium

presents

Yufeng Shen, PhD

Columbia Genome Center

Thursday Feb. 22nd | **12PM** | **Zoom** | **ID: 892 9868 6869 Password: genome!**

Predicting Genetic Effect of Missense Variants by Machine Learning

Accurate prediction of genetic effect of missense variants is critically important in analysis of human diseases and conditions. Commonly used computational prediction methods does not capture the quantitative impact on fitness in populations. We developed MisFit to estimate missense fitness effect using biobank-scale human population genome data. MisFit jointly models the effect at molecular level (d) and population level (selection coefficient, s), assuming that in the same gene, missense variants with similar d have similar s. MisFit is a probabilistic graphical model that integrates deep neural network components and population genetics models efficiently with inductive bias based on biological causality of variant effect. We trained it by maximizing probability of observed allele counts in 236,017 European individuals. We show that s is informative in predicting frequency across ancestries and consistent with the fraction of de novo mutations given s. Finally, we show the utility of MisFit in prioritizing missense variants in individuals with neurodevelopmental disorders.

SPONSORED BY:

Institute for Genomic Medicine | UC San Diego Bioinformatics and Systems Biology | CCMI Cancer Cell Map Initiative | UC San Diego School of Medicine

FOR MORE INFORMATION PLEASE VISIT GENOMIC.WEEBLY.COM