Rethinking single-locus analysis in rare variant association tests

Rationale for analyzing rare variants

Rare variants (RVs) have been shown to play a significant etiological role in complex disorders. Due to extreme rarity of mutations, association of RVs tends to be evaluated at aggregate level instead of at single-locus level. Identifying individual risk RVs is critical for understanding disease mechanisms and variant functions, but it is a challenging task because the information content of single RVs is little and consequently significances of causal variants cannot easily stand out from those of non-causal ones. In this presentation, I will discuss two proposed strategies of single RV analysis to detect individual risk RVs.

Strategies for single RV analysis

The first strategy is to perform genome-wide single RV analysis using a false-negative control procedure for multiple testing adjustment, with an aim to select a large proportion of risk RVs with high probability. Given the observation that majority of the causal RVs can only be identified along with some non-causal ones, we propose the adaptive false negative control (AFNC) procedure, use ANFC to exclude variants that can be confidently dispatched as non-causal, and select a modest number of potentially causal RVs in genome-wide RV analysis.

In the second strategy, we propose a structure-supervised test that uses protein tertiary structure to guide local variant collapsing and assesses single RV association. Protein tertiary structures contain constructive information on how variants interact and function together, and enable information borrowing from variants that are close in structural space but are apart in sequence location.