

Interpreting Multi-Omics Data through Advancing Statistical Analysis, Parallel Computing, and Information Integration

Wenchao Zhang¹, Shizhong Xu², Patrick Zhao³

¹ Noble Research Institute, 2510, Sam Noble Parkway, Ardmore, OK, 73401, USA.
wezhang@noble.org

² Department of Botany and Plant Sciences, University of California, Riverside, CA, 92521, USA. shizhong.xu@ucr.edu

³ Noble Research Institute, 2510, Sam Noble Parkway, Ardmore, OK, 73401, USA
pzhao@noble.org

Abstract

We extended the concept of classical genome-wide association studies (GWAS) into a broader ‘ome’-wide association. Briefly, we proposed a novel LMM and developed an online tool named PATOWAS (<https://bioinfo.noble.org/PATOWAS/>), by which we can address not only GWAS, but also TWAS (Transcriptome wide association studies), and MWAS (Metabolome wide association studies) in one unified platform. To explain more the phenotypic variation and address the missing heritability, we also proposed another new LMM and developed another association tool named PEPIS (https://bioinfo.noble.org/PolyGenic_QTL/), by which the polygenic effect and epistasis from marker pairs can be accounted for. Using these tools, two-dimensional (2D) GWAS that accounts for epistatic genetic effects can be mapped, which are complementary to the one-dimensional (1D) GWAS mapping, and can provide more genetic information.

Statistically solving the LMM which involves two main procedures: (1) calculating the required kinship matrices that involve large scale matrix operations and (2) statistical testing of the huge amount of p values for all marker/marker pair variants, which are computational demanding. To efficiently calculate kinship matrices, we developed several GPU pipelines (KMC1D: <https://bioinfo.noble.org/KMC1D/>, KMC2D: <https://bioinfo.noble.org/KMC2D/>) for parallel calculating kinship matrix, which can achieve several hundreds of acceleration rate in one of our GPU test servers. To efficiently calculate p -value, we employ hundreds of CPU nodes in our distributed Linux system in both PEPIS and PATOWAS, which can effectively reduce the whole analysis time 100-200x for a typical association analysis.

In short, we proposed several novel statistical LMMs, parallel computing algorithms, and developed several association tools. These methods and tools can enable us to analyze the huge multi-omics data in one platform, providing researchers a systems biology view into the analyzed traits.

Keywords: LMM; Omics; Statistical testing; Kinship Matrix; Parallel Computing.