

International Conference on

COMPUTATIONAL BIOLOGY AND BIOINFORMATICS

September 05-06, 2018 Tokyo, Japan

A robust statistical method to integrate genotype and gene expression data identifies novel associated loci

Sarmistha Das and Indranil Mukhopadhyay

Indian Statistical Institute, India

Genome wide association studies identify many SNPs that are associated with disease traits. However, single marker test might miss SNPs with moderate effect. Moreover, gene expression contains information about the deregulation of genes when compared between cases and controls. Due to multiple testing or other issues some signals may remain unidentified especially when the sample size is not too large. Moreover, RNA being unstable than DNA, high cost is involved in RNA analysis leading to smaller sample for expression data than genotype data. No standard statistical procedure is available that integrates data from various sources to decode biologically sound interpretation on heritable traits. This motivates us to propose a novel method that tests for multi-loci association in the existing scenario. Based on a two-stage regression method our method essentially concatenates genome-wide expression data and disease-associated SNP data, when sample size for expression data is much smaller than genotype data. We integrated the information contained in both data sources into a latent variable based model. Our simple yet powerful multi-loci association test integrates two databases that broadcasts more of the deep-seated features comprehensively in a single test, which might be lost when datasets are considered in singularity. We also developed asymptotic distribution of our test statistic for fast calculation of p-value for real data set. Extensive simulation confirms that our method is powerful and robust to many genetic models. We have received promising result and identified few novel markers at genome-wide level even with a small gene expression dataset related to psoriasis.

sarmisthadascu@gmail.com