

Dr. Kim-Anh Lê Cao  
Snr Lecturer, Statistical Genomics  
School of Mathematics & Statistics  
Melbourne Integrative Genomics  
The University of Melbourne | VIC 3010  
T: +61 (0)3834 43971 | kimanh.lecao@unimelb.edu.au

March 19 2018

Dear Associate Editor of Nature Methods,

We wish to submit our manuscript “DIABLO: from multi-omics assays to biomarker discovery, an integrative approach” for consideration as a research article in your journal.

In the omics era, computational solutions to integrate different types of biological data measured on the same specimens or samples are trailing behind data generation. Our manuscript aims to fill this gap by proposing an efficient, flexible and easy-to-use computational framework to integrate multiple omics data generated from emerging high-throughput technologies.

The main challenge facing multi-omics data integration is the large heterogeneity and difference in scales between omics platforms. Statistical integrative methods for biomarker discovery are still at their infancy and provide limited insight into complex biological processes. They are built on existing multi-steps methods that either concatenate or combine the analyses from each data set separately, and do not model the correlation structure between the different molecular levels. This is highly problematic as important information can be missed, leading to incorrect conclusions. DIABLO maximises the correlation between data sets whilst identifying the key molecular features that explain and reliably classify a phenotype of interest. The dimension reduction process enables intuitive visualisations of the samples and selected multi-omics signatures. We benchmarked and demonstrated the ability of DIABLO to select relevant correlated and discriminative biomarkers in a comprehensive simulation studies and in six multi-omics studies including two case studies in human breast cancer and asthma. In each of those studies we integrated various omic data sets ranging from transcriptomics (mRNA, miRNA), epigenomics (CpGs), proteomics and cell-type frequencies.

DIABLO facilitates the integration of large and heterogeneous data sets to identify relevant biomarker candidates in a wide range of biological settings. The method will be of significant interest to the scientifically diverse readership of Nature methods to capitalise on multi omics data currently being generated and push novel biological discoveries of an unprecedented level.

We are fervent advocates of open data and open science. All analyses are available in R markdown format as supplementary material, and the method is implemented in the open source R package mixOmics, along with detailed tutorials on the companion website <http://www.mixOmics.org/mixDIABLO>.

The submitted manuscript has been approved by all authors and has not been submitted to any other journal. We look forward to your reply.

Yours sincerely,



Dr. Kim-Anh LÊ CAO