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

THE UNIVERSITY OF MELBOURNE

FACULTY OF SCIENCE

May 28, 2018

Dear Professor Oliver Stegle,

We wish to submit our manuscript 'DIABLO: an integrative approach for identifying key molecular drivers from multi-omic assays' as a research article for the Systems Biology category in Bioinformatics.

As you are well aware, computational solutions to integrate different types of biological data measured on the same specimens or samples are trailing behind data generation in the era of systems biology. 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 whilst identifying a discriminatory signature.

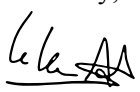
Although not published yet, DIABLO has already received interest from the Systems Biology community independent to our network to apply DIABLO (*e.g.* Mardinoglu et al. 2018, *Cell Metabolism* 27(3), https://doi.org/10.1016/j.cmet.2018.01.005; Tang et al. 2017, *Inflammatory Bowel Diseases* 23(9) https://doi.org/10.1097/MIB.0000000000001208). DIABLO was also a key method applied in our latest manuscript in revision in *Nature Communication* from Gill et al. 'Dynamic molecular changes during the first week of human life follow a robust developmental trajectory'. Those impactful studies cover a diverse spectrum of biological systems and types of biological data, including microbiome.

The main challenge in multi-omics data integration is the large heterogeneity and difference in scales between omics platforms. In addition, interpretability of the results is key for researchers to obtain insights into those complex biological processes. Our manuscript proposes a novel latent component-based method with a supervised framework that maximises the correlation structure between datasets whilst identifying the key molecular features that explain and reliably classify a phenotype of interest. The dimension reduction process using latent components enables intuitive visualisations of the samples and selected multi-omics signatures. We extensively benchmarked and demonstrated the ability of our method to select highly relevant, correlated and discriminative biomarkers in six multi-omics studies including two case studies in human breast cancer and asthma, and in a comprehensive simulation study. We integrated a wide range of omics datasets, from transcriptomics (mRNA, miRNA), epigenomics (CpGs), proteomics and cell-type frequencies. We also compared DIABLO against integrative methods recently proposed in this emerging field, such as the unsupervised method MOFA (https://doi.org/10.1101/217554, bioRxiv, April 2018).

DIABLO facilitates the integration of large and heterogeneous data sets to identify relevant biomarker candidates in a wide range of biological and clinical settings, which will be of significant interest to *Bioinformatics* readers who wish to capitalise on newly generated multi-omics data and push novel biological discoveries to an unprecedented level.

The method is implemented in the open source R package mixOmics, and our R scripts in R markdown format, along with detailed tutorials are available on our companion website http://www.mixOmics.org/mixDIABLO. We look forward to your reply.

Yours sincerely,



Dr. Kim-Anh LÊ CAO