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 17, 2018

Dear Profs Kimmel and Sczakiel,

We wish to enquire a pre-submission for our manuscript 'DIABLO: an integrative approach for identifying key molecular drivers from multi-omic assays' as a *Nucleic Acid Research* Methods special collection (Computational methods or Methods online) article.

In the era of systems biology, 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.

Although not published yet, DIABLO has been highlighted in our recent software article 'mixOmics: An R package for 'omics feature selection and multiple data integration' viewed > 11,800 times since publication in *PLOS Computational Biology* in November 2017 (https://doi.org/10.1371/journal.pcbi.1005752). DIABLO has been already used by researchers external to our network to integrate data (*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) and is also a key method used 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 human diseases and types of biological data, including microbiome. Therefore, we feel that our manuscript should be targeted to a journal such as *Nucleic Acid Research* with a scientifically diverse readership in both computational and biological disciplines.

The main challenge in multi-omics data integration is the large heterogeneity and difference in scales between omics platforms. Statistical integrative methods for biomarker discovery are still in their infancy and provide limited insight into complex biological processes. Existing methods adopt a naïve way of integrating such data and miss critical correlations between the different molecular levels. Novel integrative methods that explicitly model the correlation structure between datasets are needed to fully leverage on those expensive 'omics studies.

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 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 benchmarked DIABLO against integrative methods recently proposed in this emerging field, such as 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 the *Nucleic Acid Research* readers who wish to capitalise on newly generated multi-

omics data and push novel biological discoveries to an unprecedented level.

Our R scripts are available in R markdown format as supplementary material, the method is implemented in the open source R package mixOmics, with detailed tutorials on our companion website http://www.mixOmics.org/mixDIABLO. We enclose to this pre-submission enquiry the latest draft of our manuscript and we look forward to your reply.

Yours sincerely,

Dr. Kim-Anh LÊ CAO