NANYANG TECHNOLOGICAL UNIVERSITY

School of Biological Sciences

Reg. No. 200604393R

Research Theme: Bioinformatics/Computational Biology/Bio-Data Science

Research Project Title: Missing protein prediction using networks

Principal Investigator/Supervisor: Wilson Goh

Co-supervisor/ Collaborator(s) (if any): Limsoon Wong

Project Description

a) Background:

Assaying protein identities/quantities in a sample, i.e. proteomics, paints an immediate picture of the molecular landscape underlying the sample. However, mass spectrometry-based proteomics suffers from incomplete proteome coverage issues (i.e. many proteins actually in a sample are not observable in a single screen), and consequent inconsistency issues (i.e. different screens on the same sample generate different protein sets). Thus efforts to extend proteome profiling to day-to-day clinical studies (i.e. practical use in clinics) are rendered ineffective.

The present project concerns this missing-protein problem, i.e. identifying and quantifying proteins that are present in a sample but are not detected in a proteome screen on that sample. The project aims to develop three levels of solutions (viz. a new protein-reporting rule, complex-based prediction of missing proteins, and complex-based inference of the abundance level of missing proteins). Successful resolution of missing proteins, as proposed by this project, will improve proteome coverage and consistency. This will be very useful for practical endeavours including clinical diagnosis, biomarker development, and drug target identification.

b) Proposed work:

The overall project comprises three thrusts. You may select 1 of these as the basis for the PhD project.

In Thrust #1, the conventional Human Proteome Project guideline known as the two-peptide rule is re-considered. The two-peptide rule defines a protein as detected in a sample when there are two non-nested detected peptides of length ≥ 9 amino acids that uniquely map to the protein. Thrust #1 investigates whether and how this two-peptide rule can be relaxed without compromising the reliability of a proteomic screen.

Thrust #2 is aimed at developing a novel ranking strategy for missing-protein recovery based on protein complexes. The strategy is to estimate the likelihood of candidate missing proteins as being present in a sample via a Bayesian inference on the likelihood of their parent complexes being present in the sample. This strategy enables inference of missing proteins even when there is only one or very few samples of the same phenotype in the same batch.

In Thrust #3, the correlation (observed in a set of samples) in the abundance of proteins in the same protein complex is postulated to be more likely genuine than that between proteins that are not in the same complex. Thus Thrust #3 is aimed at using only the former (i.e. the co-complex proteins) to impute the abundance of a missing-but-predicted-present protein in a sample. Moreover, Thrust #3 aims to use only complexes that are likely to be present in the sample; these complexes can be selected based on the likelihoods inferred as in Thrust #2 or based on domain knowledge.

60 Nanyang Drive, Singapore 637551
Tel: +65 6316 2800, Fax: +65 6791 3856
Website: http://w ww.sbs.ntu.edu.sg/Pages/Home.aspx

NANYANG TECHNOLOGICAL UNIVERSITY

School of Biological Sciences

Reg. No. 200604393R

Candidates with degrees in data science, data analytics, statistics, mathematics, engineering, computing are welcome to apply. Please link to https://gohwils.github.io/biodatascience/ to find out more.

Supervisor contact:

If you have questions regarding this project, please email the Principal Investigator: wilsongoh@ntu.edu.sg

SBS contact and how to apply:

Associate Chair-Biological Sciences (Graduate Studies) : AC-SBS-GS@ntu.edu.sg

Please apply at the following:

http://admissions.ntu.edu.sg/graduate/R-Programs/R-WhenYouApply/Pages/R-ApplyOnline.aspx

60 Nanyang Drive, Singapore 637551
Tel: +65 6316 2800, Fax: +65 6791 3856
Website: http://w ww.sbs.ntu.edu.sg/Pages/Home.aspx