

I. IDENTIFICATION DATA

Thesis name:	Evaluating Directional and Association Methods on Single-cell RNA Sequencing Data
Author's name:	Eliška Dvořáková
Type of thesis :	master
Faculty/Institute:	Faculty of Electrical Engineering (FEE)
Department:	Computer Science
Thesis supervisor:	Joe Song
Supervisor's department:	Department of Computer Science & Molecular Biology Graduate Program, New Mexico State University, Las Cruces, USA

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment	challenging
<i>Evaluation of thesis difficulty of assignment.</i>	
Single-cell RNA sequencing (scRNA-seq) data are notoriously challenging to analyze due to technical errors introduced from sample preparation to sequencing. Recent literature suggests a wide range of variations in performance for normalization procedures and network inference algorithms when applied on scRNA-seq data. This current status limits the promise of this new technology in uncovering biological patterns at the single-cell resolution. This thesis assignment to Ms. Dvořáková on evaluating and improving scRNA-seq data analysis methods is thus challenging.	

Satisfaction of assignment	fulfilled
<i>Assess that handed thesis meets assignment. Present points of assignment that fell short or were extended. Try to assess importance, impact or cause of each shortcoming.</i>	
The thesis studied several association methods to show that a recent method called functional index is relatively effective on detecting directional dependency on scRNA-seq data which often have highly skewed marginal distributions due to the technical phenomenon of dropout introduced during the capture of RNA molecules in sample preparation. Most satisfactorily, a new method for library size normalization based on randomized up-down sampling has been designed to prevent correlation artifacts introduced by popular library size normalization methods. Both simulation studies and real scRNA-seq data were used to evaluate the proposed methods. The collected evidence is thus convincing in advancing the field. Thus, the thesis assignment is considered fulfilled.	

Activity and independence when creating final thesis	A - excellent.
<i>Assess that student had positive approach, time limits were met, conception was regularly consulted and was well prepared for consultations. Assess student's ability to work independently.</i>	
Ms. Dvořáková demonstrated outstanding intellectuality, professionalism, and independence during the thesis work. She reliably presented weekly progress in person and via online Zoom meetings. She developed all R code that generated the results in her thesis. She communicated well in English.	

Technical level	B - very good.
<i>Assess level of thesis specialty, use of knowledge gained by study and by expert literature, use of sources and data gained by experience.</i>	
The thesis work is highly technical. The part on evaluation of directional association methods is already peer-reviewed and published in a recent international conference on bioinformatics in December 2019. The part on library size normalization has been evaluated on latest multiomics data collected on single cells of leukemia and healthy blood cells published in 2019. The proposed method in the second part is promising for another publication in bioinformatics.	

Formal and language level, scope of thesis	B - very good.
<i>Assess correctness of usage of formal notation. Assess typographical and language arrangement of thesis.</i>	
Methods are presented in satisfactory formal notation.	

Selection of sources, citation correctness**A - excellent.**

Present your opinion to student's activity when obtaining and using study materials for thesis creation. Characterize selection of sources. Assess that student used all relevant sources. Verify that all used elements are correctly distinguished from own results and thoughts. Assess that citation ethics has not been breached and that all bibliographic citations are complete and in accordance with citation convention and standards.

Ms. Dvořáková demonstrated excellent skills in documenting sources and citations that support the thesis work.

Additional commentary and evaluation

Present your opinion to achieved primary goals of thesis, e.g. level of theoretical results, level and functionality of technical or software conception, publication performance, experimental dexterity etc.

I started working with Ms. Dvořáková during my visit to Czech Technical University in Spring 2019, when I met the student weekly in person to discuss progress. For the past one and half year, she has become drawn to the field, while improving proficiency in technical details involving computer science, statistics, and biology. I learned that she met great personal challenges due to the COVID-19 pandemic. In spite of such conditions, the accomplishments of the student have been extraordinary.

III. OVERALL EVALUATION, QUESTIONS FOR DEFENSE, CLASSIFICATION SUGGESTION

Summarize thesis aspects that swayed your final evaluation.

The thesis presented methods for normalization and directional association on single-cell RNA sequencing data. It advocates a method called functional index for directional association due to its robustness to marginal distributions skewed to zero widely observed on single-cell RNA sequencing data. Most importantly, a novel approach for library size normalization based on randomized up-down sampling is proposed. Both simulation and real data studies suggest that the method may help reduce spurious correlation introduced by other existing library size normalization methods. The methods are novel for single-cell biology and the results hint at potential positive impacts of the methods on advancing biological sciences.

I evaluate handed thesis with classification grade **A - excellent**.

Date: **24.8.2020**

Signature: