

Haoyun Lei

E-mail: leovam171@gmail.com

Phone: +1(412)969-3798

Linkedin: [linkedin.com/in/haoyunlei/](https://www.linkedin.com/in/haoyunlei/)

Website: <https://leovam.github.io/>

EDUCATION

Ph.D. in Computational Biology

May 2021

Joint Carnegie Mellon-University of Pittsburgh Ph.D. Program in Computational Biology

(expected)

Computational Biology Department, School of Computer Science

Carnegie Mellon University (CMU), Pittsburgh, PA, USA

B.S. in Biological Science

June 2012

College of Life Science and Technology

Huazhong University of Science and Technology (HUST), Wuhan, China

TECHNICAL SKILLS

Python, Java, R, Matlab, Shell, Linux, L^AT_EX, Data Analysis, Machine Learning, Algorithm Design,

Combinatorial Optimization

EXPERIENCE

Research Assistant

Aug 2016 – present

Computational Biology Department, School of Computer Science,

Carnegie Mellon University, Pittsburgh, PA, USA

- Design algorithms that integrate multiple genomic data to accurately infer intratumor heterogeneity
 - Create a mixed membership model to for clonal deconvolution to infer the copy number of segments in chromosome of a brain cancer called Glioblastoma
 - Develop an efficient coordinate descent algorithm to balance the deconvolution quality among different genomic data such as bulk sequencing data, single-cell sequencing data and fluorescent *in situ* hybridization data
 - Design a mixed integer linear programming model that incorporates minimum evolution phylogenetic tree that explains the tumor progression
 - Compared to other methods, ours only need small size of samples, and are robust to the noise
- Derive phylogenies from transcriptome data to explain the tumor evolution and potential order of perturbed pathways in Breast Cancer Brain Metastases
 - Preprocess the transcriptome to get the most important bulk gene expression data and map them to Gene Modules and Pathways with external knowledge from different database
 - Develop a novel method called Neural Network Deconvolution using backpropagation to detect the major cell clones with corresponding fractions in bulk samples
 - The pathways found are consistent with previous hypothesis and may provide new markers and targets for drugs
- Participate in Commonwealth Universal Research Enhancement (CURE) project where a lot of researchers develop better methods for integrating, analyzing and modeling large volumes of diverse data on cancer patients
 - Apply bioinformatics tools on TCGA data
 - Detect and extract features of genes in breast and lung cancers with machine learning approaches

Visiting Research Scholar (wet lab experience)

Sep 2012 – Nov 2015

Life Sciences Institute,

University of Michigan, Ann Arbor, MI, USA

TEACHING EXPERIENCE

Algorithm and Advanced Data Structure
Laboratory Methods for Computational Biologists

Aug 2019 – present
Aug 2018 – April 2019

COURSE AND COURSE PROJECTS

Algorithm and Advanced Data Structure

Computational Methods for Biological Modeling and Simulation

- Project: Designed a Markov Chain Monte Carlo (MCMC) model to predict recurrent rate and survival rate of different complications of diabetes

Machine Learning

- Project: Applied and improved supervised learning models such as SVM and Neural Network to find the patterns of well-stored images of footprint, and then matched the new-incoming images. The project contained imaging process, PCA-analysis, parameters tuning and objective functions design

Computational Genomics

- Project: Improved different classifiers such as Logistic Regression, Naïve Bayes Classifier and SVM Classifier to detect important features that could benefit the novel protein-coding genes that were barely reported previously

Cellular and Systems Modeling

- Project: Designed a hybrid model to address the problems that the existing Boolean network model loses dynamic mechanics details of the system and that the existing continuous ODE models can only capture small-scale network module properties and hard to apply to complicated networks

Introduction to Bioinformatics Programming in Python

- Project: Applied Python scientific computing packages as Numpy and Scipy to deal with meta-genomic data, applied machine-learning packages such scikit-learn to do the features extraction and cluster, also deployed visualization package such as matplotlib to plot the results

Structure Biology

BIBLIOGRAPHY

Peer-reviewed articles

Tao, Y., **Lei, H.**, Lee, A., Ma, J., and Schwartz, R. (2019). Phylogenies Derived from Matched Transcriptome Reveal the Evolution of Cell Populations and Temporal Order of Perturbed Pathways in Breast Cancer Brain Metastases (submitted to ISMCO'19)

Lei, H., Lyu, B., Gertz, E., Schäffer, A., Shi, X., Wu, K., Li, G., Xu, L., Hou, Y., Dean, M., and Schwartz, R. (2018). Tumor Copy Number Deconvolution Integrating Bulk and Single-Cell Sequencing Data (RECOMB 2019, accepted by *Journal of Computational Biology* as special issue)

Abstracts

Lei, H., Lyu, B., Gertz, E. M., Schäffer, A. A., & Schwartz, R. (2018, October). Tumor Copy Number Data Deconvolution Integrating Bulk and Single-cell Sequencing Data. In *2018 IEEE 8th International Conference on Computational Advances in Bio and Medical Sciences (ICCABS)* (pp. 1-1). IEEE.

Lei, H., Roman, T., Eaton, J., and Schwartz, R. (2018, July). New directions in deconvolving genomics mixtures of copy number variation data. SIAM Conference on Discrete Mathematics, Denver, CO.

Lei, H., Roman, T., Eaton, J., and Schwartz, R. (2018, April). Deconvolution of tumor copy number data using bulk and single-cell sequencing data. Conference on Intelligent System for Molecular Biology, Chicago, IL.