Li, Wei Vivian

683 Hoes Lane West Rm 210
Rutgers, The State University of New Jersey
Piscataway, NJ 08854

Telephone: 732-235-8574
Email: vivian.li@rutgers.edu
Homepage: vivianli.org

RESEARCH INTERESTS

My research focuses on statistical modeling, machine learning, and algorithm development, as well as their applications to high-throughput sequencing data in bioinformatics and clinical research. Major research topics of my group include:

· Developing statistical and machine learning models for
  · Denoising and handling sparsity in single-cell genomic data
  · Dimensionality reduction, integration, and clustering of gene expression data
  · Hypothesis testing of differential patterns in genomic data
· Using statistical and computational models to interpret and understand
  · Post-transcriptional RNA regulation
  · Gene regulatory networks

EDUCATION

University of California, Los Angeles (UCLA)   09/2014 - 06/2019
Department of Statistics
Ph.D. in Statistics
Dissertation: Statistical methods for bulk and single-cell RNA sequencing data
Advisor: Dr. Jingyi Jessica Li

Huazhong University of Science & Technology (HUST)   09/2010 - 06/2014
School of Mathematics and Statistics
B.S. in Statistics

EMPLOYMENT

Assistant Professor   08/2019 -
Department of Biostatistics and Epidemiology
School of Public Health
Rutgers, The State University of New Jersey

Teaching Assistant Consultant   10/2018 - 06/2019
Department of Statistics
University of California, Los Angeles

Graduate Student Researcher   09/2016 - 06/2018
Department of Statistics
University of California, Los Angeles

HONORS AND AWARDS

Busch Biomedical Grant Award, Rutgers	2020
The Caliburn Allocation Award of Research Computing, Rutgers	2020
The Most Outstanding Statistician Award, UCLA	2019
Physical Sciences Entrepreneurship & Innovation Award (co-inventor), UCLA	2018
Dissertation Year Fellowship, UCLA	2018
Pearl Cohen Poster Award, UCLA Biomedical & Life Science Innovation Day	2018
Diversity Scholarship, UseR 2016 Conference	2016
Doctoral Student Travel Grant, UCLA	2016
The Most Promising Computational Statistician Award, UCLA	2015
Distinguished Graduate of Class 2014, HUST	2014
National Fellowship, Ministry of Education, China	2013
Student Award of Excellence (awarded to 20 students per year), HUST	2013
National Fellowship, Ministry of Education, China	2012
National Fellowship, Ministry of Education, China	2011
Outstanding Student Award (awarded to 3 students in each department), HUST	2011

RESEARCH FUNDING

Ongoing:

NSF/IIS 2128307 (Zhang)   10/01/2021 - 09/30/2022
EAGER: Integration and analysis of high-dimensional dataset
Role: Co-PI

NIH/NIGMS R35 GM142702 (Li)   09/01/2021 - 06/30/2026
Novel statistical methods for multiscale analysis of single-cell transcriptomes
Role: PI

NIH/NIMH R21 MH126420 (Pang & Li)   07/01/2021 - 06/30/2023
Multimodal profiling of neurons in 3D human cortical organoids using patch-seq
Role: PI

Rutgers Busch Biomedical Grant (Li)   10/01/2020 - 09/30/2022
A novel approach for joint modeling of multi-patient single-cell gene expression data
Role: PI

NJ ACTS Mini-Methods Grant (Li)   06/04/2020 - 02/28/2022
Statistical methods for single-cell differential gene network analysis
Role: PI

Completed:

Rutgers School of Public Health Pilot Grant (Li)   11/01/2019 - 06/01/2021
Understanding gene co-expression patterns in single cells
Role: PI

PUBLICATIONS

___ corresponding authors   † members of Li group   * co-first authors

Preprint/Manuscript under review:

1. Qian K[†], Fu S[†], Li H, and **Li WV**. (2021). scINSIGHT for interpreting single-cell gene expression from biologically heterogeneous data. *Under review at Genome Biology*.

2. **Li WV**. (2021). Phitest for Analyzing the Heterogeneity of Single-cell Populations. *Under review at Bioinformatics*.

3. Colon A, Hirday R, Patel A, Poddar A, Tuberty-Vaughan E, Fu T, Ai X, **Li WV**, and Cai L. (2021). A computational pipeline for functional gene discovery. *Under review at Scientific Reports*.

4. **Li WV**, Tong X, and Li JJ. (2020). Bridging Cost-sensitive and Neyman-Pearson Paradigms for Asymmetric Binary Classification. *arXiv*.

5. Roque W, Cuevas-Mora K, Sales D, **Li WV**, Rosas IO, and Romero F. (2020). Transcriptomics, metabolomics and lipidomics of chronically injured alveolar epithelial cells reveals similar features of IPF lung epithelium. *bioRxiv*.

Refereed Article:

6. **Li WV** and Li Y[†]. (2021). scLink: Inferring Sparse Gene Co-expression Networks from Single-cell Expression Data. *Genomics, Proteomics & Bioinformatics*, in press.

7. Sheng J[†] and **Li WV**. (2021) Selecting gene features for unsupervised analysis of single-cell gene expression data. *Briefings in Bioinformatics*, bbab295.

8. **Li WV**, Zheng D, Wang R, and Tian B. (2021). MAAPER: model-based analysis of alternative polyadenylation using 3'end-linked reads. *Genome Biology*, 22, 222.

9. Zhang XS, Yin YS, Wang J, Battaglia T, Krautkramer K, **Li WV**, Li J, Brown M, Zhang M, Badri MH, Armstrong AJ, Strauch CM, Wang Z, Nemet I, Altomare N, Devlin JC, He L, Morton JT, Chalk JA, Needles K, Liao V, Mount J, Li H, Ruggles KV, Bonneau RA, Dominguez-Bello MG, BÃd'ckhed F, Hazen SL, Blaser MJ. (2021). Maternal cecal microbiota transfer rescues early-life antibiotic-induced enhancement of type 1 diabetes in mice. *Cell Host & Microbe*, 29(8), 1249-1265.

10. Jiang R, **Li WV**, and Li JJ. (2021). mbImpute: an accurate and robust imputation method for microbiome data. *Genome Biology*, 22(1), 192.

11. Sun T, Song D, **Li WV**, and Li JJ. (2021). scDesign2: an interpretable simulator that generates high-fidelity single-cell gene expression count data with gene correlations captured. *Genome Biology*, 22(1), 1-37.

12. Zheng Y, Chen Z, Han L, Han Y, Zou X, Zhou B, Hu R, Hao J, Bai S, Xiao H, **Li WV**, Bueker A, Ma Y, Xie G, Yang J, Chen S, Li H, Cao J, and Shen L. (2020). Immune suppressive landscape in the human esophageal squamous cell carcinoma microenvironment. *Nature Communications*, 11(1), 1-17.

13. Wang J, Deng F, Zeng F, Shanahan AJ, **Li WV**, and Zhang L. (2020). Predicting long-term multicategory cause of death in patients with prostate cancer: random forest versus multinomial model. *American Journal of Cancer Research*, 10(5), 1344.

14. Xu J, Cheng Y, Yuan X, **Li WV**, and Zhang L. (2020). Trends and prediction in daily incidence of novel coronavirus infection in China, Hubei Province and Wuhan City: an application of Farr's law. *American Journal of Translational Research*, 12(4), 1355.

15. **Li WV***, Li S*, Tong X, Deng L, Shi H, and Li JJ. (2019) AIDE: annotation-assisted isoform discovery with high precision. *Genome Research*, 29, 2056-2072.

16. **Li WV** and Li JJ. (2019). A statistical simulator scDesign for rational scRNA-seq experimental design. *Bioinformatics (ISMB/ECCB 2019)*, 35(14), i41-i50.

17. Ge X, Zhang H, Xie L, **Li WV**, Kwon SB, and Li JJ. (2019). EpiAlign: an alignment-based bioinformatic tool for comparing chromatin state sequences. *Nucleic Acids Research*, gkz287.

18. **Li WV** and Li JJ. (2018). An accurate and robust imputation method scImpute for single-cell RNA-seq data. *Nature Communications*, 9, 997.

19. **Li WV** and Li JJ. (2018). Modeling and analysis of RNA-seq data: a review from a statistical perspective. *Quantitative Biology*, 6(3), 195-209.

20. **Li WV**, Zhao A, Zhang S, and Li JJ. (2018). MSIQ: joint modeling of multiple RNA-seq samples for accurate isoform quantification. *Annals of Applied Statistics*, 12(1), 510-539.

21. **Li WV**, Chen Y, and Li JJ. (2017). TROM: A testing-based method for finding transcriptomic similarity of biological samples. *Statistics in Biosciences*, 9(1), 105-136.

22. **Li WV**, Razaee ZS, and Li JJ. (2016). Epigenome overlap measure (EPOM) for comparing tissue/cell types based on chromatin states. *BMC Genomics*, 17(1), S10.

    Dissertation:

23. **Li WV.** (2019). Statistical methods for bulk and single-cell RNA sequencing data. PhD dissertation, UCLA.

    Google Scholar: https://scholar.google.com/citations?user=HzjANkQAAAAJ

INVITED TALKS AND CONFERENCE PRESENTATIONS

Conference and Symposium Presentations:

1. Invited Talk, Advanced Genetic Sequencing Virtual Conference. 09/2021
   Model-based Analysis of Alternative Polyadenylation Using 3'end Reads

2. Invited Poster, 1st Annual NJ ACTS Science Symposium. 09/2021
   scLink: Inferring Sparse Gene Co-expression Networks from Single-cell Expression Data

3. Invited Talk, ICSA 2021 Applied Statistics Symposium. 09/2021
   MAAPER for Model-based Analysis of Alternative Polyadenylation Using 3'end Reads

4. Talk, RECOMB 2021. 09/2021
   scLink: Inferring Sparse Gene Co-expression Networks from Single-cell Expression Data

5. Invited Talk, Joint Statistical Meetings. 08/2021
   Comparing Cost-sensitive and Neyman-Pearson Frameworks for Asymmetric Binary Classification

6. Talk, 29th Conference on Intelligent Systems for Molecular Biology. 07/2021
   MAAPER for Model-based Analysis of Alternative Polyadenylation Using 3'end Reads

7. Invited Talk, Fifth International Conference on Recent Trends in Statistical 06/2021
   Theory and Applications.
   Selecting Gene Features for Unsupervised Analysis of Single-cell Gene Expression Data

8. Invited Talk, Sixth International Conference on Statistics for Twenty-first Century. 12/2020
   Statistical Challenges in Single-cell RNA Sequencing Data Analysis

9. Poster, RECOMB/ISCB 2020 Conference on Regulatory and Systems Genomics. 11/2020
   Inferring Gene Co-expression Networks from Single Cell Gene Expression Data

10. Talk, Human Cell Atlas Asia Meeting 2020. 10/2020
    Inferring Gene Co-expression Networks from Single Cell Gene Expression Data

11. Talk, 2020 Joint Statistical Meetings. 08/2020
    AIDE: Annotation-Assisted RNA Transcript Discovery with High Precision

12. Invited Talk, RCAF Forum: Challenges and Opportunities in 08/2020
    Computational Biology.
    Current Computational and Analytical Challenges in Transcriptomics

13. Poster, 28th Conference on Intelligent Systems for Molecular Biology 07/2020
    (ISMB 2020).
    Inferring Gene Co-expression Networks from Single Cell Gene Expression Data

14. Invited Talk, RCAF Webinar Series. 05/2020
    Statistical Challenges and Opportunities in Single-cell Studies

15. Invited Talk, ENAR 2020 Spring Meeting. 03/2020
    A Unified View of Asymmetric Binary Classification

16. Invited Talk, Inaugural Lange Symposium Software Workshop. 02/2020
    Improving Single-cell RNA-seq Data Analysis with scImpute and scDesign

17. Invited Talk, 2019 Joint Statistical Meetings. 07/2019
    A Unified View of Asymmetric Binary Classification

18. Invited Talk, ENAR 2019 Spring Meeting. 03/2019
    A Unified View of Asymmetric Binary Classification

19. Talk, 8th Annual Southern California Systems Biology Conference. 02/2019
    AIDE: Annotation-assisted Isoform Discovery with High Precision

20. Invited Talk, 2018 IEEE 8th International Conference on Computational Advances in Bio and Medical Sciences.   10/2018
    scImpute: Accurate and Robust Imputation for Single Cell RNA-seq Data

21. Talk and Poster, Inaugural Symposium on Multiscale Cell Fate.   10/2018
    scImpute for Imputation of ScRNA-seq Data

22. Talk, UCLA 4th Annual QCBio Retreat.   09/2018
    AIDE for Annotation-assisted mRNA Isoform Reconstruction and Quantification

23. Talk, 2018 Joint Statistical Meetings.   08/2018
    Joint Modeling of Multiple RNA-seq Samples for Accurate Isoform Quantification

24. Talk, 26th Conference on Intelligent Systems for Molecular Biology (ISMB 2018).   06/2018
    scImpute: Accurate and Robust Imputation for Single Cell RNA-seq Data

25. Poster, UCLA Biomedical & Life Science Innovation Day.   06/2018
    scImpute: Accurate Imputation for Single Cell RNA-seq Data

26. Invited Talk, DahShu Webinar.   04/2018
    An Accurate and Robust Imputation Method for Single-cell RNA-seq Data

27. Invited Talk, 3rd Annual NGS Data Analysis & Informatics Conference.   02/2018
    scImpute: Accurate and Robust Imputation for Single Cell RNA-seq Data

28. Poster, UCLA 3rd Annual QCBio Retreat.   09/2017
    scImpute: Accurate and Robust Imputation for Single Cell RNA-seq Data

29. Talk, Asia Pacific Bioinformatics Conference 2016.   01/2016
    Epigenome Overlap Measure (EPOM) for Comparing Tissue/cell Types based on Chromatin States

30. Talk, USC/UCLA Bioinformatics Joint Meeting.   11/2015
    A Testing-based Measure EPOM for Comparing Tissue/cell Types based on Chromatin States

31. Poster, UCLA 1st Annual QCBio Retreat.   09/2015
    A New Measure for Comparing Epigenetic Profiles in Tissue/cell Types

    Seminar Presentations:

32. NJ ACTS Biostatistics and Epidemiology Workshop, Rutgers University.   05/2021
    An Introduction to Analysis of RNA Sequencing Data at Multiple Scales

33. Department of Mathematics, Texas State University.   04/2021
    Statistical Methods for Analyzing Alternative Splicing from RNA Sequencing Data

34. Bioinformatics-Human Genetics Seminar, UCLA.   04/2021
    Model-based Analysis of Alternative Polyadenylation Using 3'end Reads

35. Institute for Quantitative Biomedicine, Rutgers University.   01/2021
    Statistical Methods for Understanding Post-transcriptional Regulation

36. Epigenetics User Group Meeting, Rutgers Department of Genetics   02/2020
    Improving Statistical Analysis of Bulk and Single-cell RNA-seq Data

37. Systems Biology/Bioinformatics Meeting, Rutgers Cancer Institute of New Jersey 10/2019
    Improving Statistical Analysis of Bulk and Single-cell RNA-seq Data

38. Department of Biostatistics and Epidemiology, Rutgers University.   02/2019
    Statistical Methods to Uncover Hidden Information from Large-scale Genomic Data

39. Department of Statistics, University of Wisconsin-Madison.   02/2019
    Statistical Methods to Uncover Hidden Information from Large-scale Genomic Data

40. Department of Public Health, University of California, Merced.   01/2019
    Statistical Methods to Uncover Hidden Information from Large-scale Genomic Data

41. Department of Statistics, Pennsylvania State University.   01/2019
    Statistical Methods to Uncover Hidden Information from Large-scale Genomic Data

42. Department of Statistics, University of Illinois at Urbana-Champaign.   12/2018
    Statistical Methods to Uncover Hidden Information from Large-scale Genomic Data

43. Jonsson Comprehensive Cancer Center Gene Regulation Intramural   03/2018
    Meeting, UCLA.
    scImpute: Accurate and Robust Imputation for Single Cell RNA-seq data

TEACHING EXPERIENCE

Instructor (Rutgers Graduate Courses):

BIST 0613: Biostatistics Theory I   Fall 2021

BIST 0720: Advanced Biostatistical Computing   Spring 2021

BIST 0725: Generalized Linear Models   Spring 2020

Student Instructor (UCLA Graduate Courses):

STATS 495A: Teaching College Statistics   Fall 2018

Teaching Assistant (UCLA Graduate Courses):

STATS 207: Statistical Learning with Sparsity   Spring 2018

STATS 203: Large Sample Theory, Including Resampling   Winter 2017

STATS 205: Hierarchical Linear Models   Winter 2017

STATS M254: Statistical Methods in Computational Biology   Spring 2016

STATS 200C: Large Sample Theory   Spring 2016

Teaching Assistant (UCLA Undergraduate Courses):

STATS 101A: Introduction to Data Analysis and Regression   Summer 2016

STATS 13: Introduction to Statistical Methods for the Health and Life Sciences Winter 2016

STATS 10: Introduction to Statistical Reasoning   Fall 2015

STATS 100B: Introduction to Mathematical Statistics   Summer 2015

Guest Lecturer:

RESH 0701: Doctoral Colloquium   11/2020
Rutgers School of Public Health
Career As a Statistician in the Bioinformatics Field

DSO 530: Applied Modern Statistical Learning Methods   04/2020
Marshall School of Business, University of South California
Introduction to Cloud Computing

DSO 530: Applied Modern Statistical Learning Methods   03/2019
Marshall School of Business, University of South California
Cloud Computing with the Amazon Web Services

BIOINFO 201: Bioinformatics Interdisciplinary Research Seminar   10/2018
Bioinformatics Interdepartmental Ph.D. Program, UCLA
Statistical Methods for Isoform Quantification of Multiple Samples

STATS M254: Statistical Methods in Computational Biology   06/2018
Department of Statistics, UCLA
Statistical Methods for Isoform Quantification of Multiple Samples

STATS M254: Statistical Methods in Computational Biology   05/2018
Department of Statistics, UCLA
Imputation Methods for Single-cell RNA-seq Data

DSO 530: Applied Modern Statistical Learning Methods   02/2018
Marshall School of Business, University of South California
Cloud Computing with the Amazon Web Services

DSO 530: Applied Modern Statistical Learning Methods   03/2017
Marshall School of Business, University of South California
Introduction to the Amazon Web Services (AWS)

STUDENT MENTORING & ADVISING

Thesis Advisor of PhD Students:

· Shiwei Fu (08/2021 -)   PhD student in Biostatistics at Rutgers

Research Mentor of Graduate Students:

· Qiren Wang (09/2021 -)   MS student in Data Science at Rutgers
· Zhihan Wang (09/2021 -)   MS student in Statistics at Rutgers
· Jiayi Liu (09/2021 -)   PhD student in Molecular Biosciences at Rutgers
· Jie Sheng (11/2020 - 07/2021)   MS student in Data Science at UW-Madison
· Dayuan Wang (08/2020 - 12/2020)   MS student in Biostatistics at Rutgers
· Shiwei Fu (06/2020 - 06/2021)   MS student in Statistics at UIUC
· Yanzeng Li (12/2019 - 09/2020)   MS student in Biostatistics at Rutgers
· Alireza Naghizadeh (06/2020 - 08/2020)   PhD student in Computer Science at Rutgers

Research Mentor of Undergraduate Students:

· Andrew Cheng (09/2020 -)   Undergraduate student in Computer Science at Rutgers

Academic Advisor of Graduate Students:

· Maotian Li (08/2021 -)   MS student in Biostatistics at Rutgers
· Rachel Lee (05/2021 -)   MS student in Biostatistics at Rutgers
· Phoebe Chu (08/2020 -)   MS student in Biostatistics at Rutgers
· Changhan Lin (08/2020 -)   MS student in Biostatistics at Rutgers
· Adriel Staana (08/2020 -)   MS student in Biostatistics at Rutgers
· Daniel Yen (08/2020 -)   MS student in Biostatistics at Rutgers
· Zifan Guo (01/2020 -)   MS student in Biostatistics at Rutgers

Thesis Committee Member:

· Alireza Nohadani (2020 - 2020)   PhD student in Computer Science at Rutgers
· Tianyi Sun (2019 -)   PhD student in Statistics at UCLA
· Aysegul Guvenek (2019 - 2020)   PhD student in Biomedical Science at Rutgers

Qualifying Exam Committee Member:

· Wenxuan Xiong (2021)   PhD student in Biostatistics at Rutgers
· Peng Zhang (2021)   PhD student in Biostatistics at Rutgers

PROFESSIONAL SERVICE

Reviewer for Grant Applications:

NIH BDMA (Biodata Management and Analysis) Study Section   2021
Rutgers Busch Biomedical Grants Program   2021

Reviewer for Conferences:

2022 Pacific Symposium on Biocomputing

Reviewer for Scientific Journals:

Advanced science (×1)
Applied Sciences (×1)
Bioinformatics (×12)
BioMed Research International (×1)
Biometrics (×3)
BMC Bioinformatics (×1)
BMC Genomics (×1)
Cells (×1)
Cell Systems (×2)
Chapman & Hall/CRC Press (×1)
Clinical and Translational Medicine (×1)
Frontiers in Genetics (×4)
Gene Report (×1)
Genes (×2)
Genome Biology (×2)
Genomics, Proteomics & Bioinformatics (×1)
IEEE Access (×2)
Journal of Molecular Cell Biology (×1)
Laboratory Investigation (×2)
NAR Cancer (×2)

NAR Genomics and Bioinformatics (×1)
Nature Biotechnology (×1)
Nature Communications (×3)
Nature Methods (×2)
Nucleic Acids Research (×5)
PeerJ (×1)
Plos Computational Biology (×4)
Plos One (×1)
Proteomics (×1)
Quantitative Biology (×1)

Committee Member:

Selection Committee, ASA Biopharmaceutical Section Nonclinical Biostatistics Conference Scholarship	2021
Bylaws and Elections Committee, Rutgers School of Public Health	2020 -
Bioinformatics and Microbiome Research Committee, Rutgers University	2020

PROFESSIONAL AFFILIATIONS

Member:

Rutgers Cancer Institute of New Jersey	2019 -
Eastern North American Region International Biometric Society	2019 -
American Statistical Association	2019 -
International Chinese Statistical Association	2019 -

Student Member:

International Society for Computational Biology	2017 - 2019
International Chinese Statistical Association	2015 - 2019
American Statistical Association	2014 - 2019

SOFTWARE

R packages:

- **scDesign** for rational scRNA-seq experimental design
  https://github.com/Vivianstats/scDesign

- **scLink** for constructing single-cell gene co-expression networks
  https://github.com/Vivianstats/scLink

- **scImpute** for accurate imputation of scRNA-seq data
  https://github.com/Vivianstats/scImpute

- **MAAPER** for analyzing alternative polyadenylation using 3′ end-linked reads
  https://github.com/Vivianstats/MAAPER

- **AIDE** for annotation-assisted transcript reconstruction and quantification
  https://github.com/Vivianstats/AIDE

- **MSIQ** for quantifying expression of RNA transcripts from multiple RNA-seq samples
  https://github.com/Vivianstats/MSIQ

- **TUBE** for bridging cost-sensitive and Neyman-Pearson paradigms for asymmetric binary classification
  https://github.com/Vivianstats/TUBE
- **TROM** for comparing transcriptomes of biological samples from the same or different species
  https://cran.r-project.org/web/packages/TROM/index.html

Last updated: October 21, 2021