

Dongyuan Song

Department of Genetics and Genome Sciences
400 Farmington Avenue
Farmington, CT 06030-7064

Phone: 860-679-2259
Email: dosong@uchc.edu
Website: songdongyuan1994.github.io

PROFESSIONAL EXPERIENCE

University of Connecticut, Health Center (UConn Health)	11/2024 - Present
Assistant Professor, Department of Genetics and Genome Sciences	Farmington, CT

EDUCATION

University of California, Los Angeles (UCLA)	09/2019 - 09/2024
Ph.D. in <i>Bioinformatics</i> , Interdepartmental Ph.D. Program	Los Angeles, CA
Advisor: Dr. Jingyi Jessica Li, Department of Statistics and Data Science	

Harvard University	09/2022 - 09/2023
Visiting Ph.D. student, Department of Statistics	Boston, MA

Harvard T.H. Chan School of Public Health	09/2017 - 05/2019
M.S. in <i>Computational Biology</i> , Department of Biostatistics	Boston, MA
Advisor: Dr. Rafael Irizarry, Dana-Farber Cancer Institute	

Fudan University	09/2013 - 06/2017
B.S. in <i>Biological Science</i> , School of Life Sciences	Shanghai, China

RESEARCH INTERESTS

My research focuses on developing novel computational methods for analyzing single-cell and spatial omics data. Major research topics include:

- Generative models of single-cell and spatial multi-omics
- *In silico* control data generation for reducing false biological discoveries
- Differential expression tests for inferred clusters or pseudotime
- Gene selection and cell subsampling of large-scale single-cell data

PUBLICATIONS

author* equal contribution author co-corresponding author **Peer reviewed**

1. Kian Hong Kock, Patrick K Kimes, Stephen S Gisselbrecht, Sachi Inukai, Sabrina K Phanor, James L Anderson, Gayatri L Ramakrishnan, Colin H Lipper, **Dongyuan Song**, Jesse V Kurland, Julia M Rogers, Raehoon Jeong, Stephen C Blacklow, Rafael A Irizarry, and Martha L Bulyk. DNA binding analysis of rare variants in homeodomains reveals novel homeodomain specificity-determining residues. *Nature Communications*, 15(1):3110, 2024
2. **Dongyuan Song**, Qingyang Wang, Guanao Yan, Tianyang Liu, Tianyi Sun, and Jingyi Jessica Li. scDesign3 generates realistic in silico data for multimodal single-cell and spatial omics. *Nature Biotechnology*, 42(2):247–252, 2024 (Impact factor: 33.1)

3. Guanao Yan, **Dongyuan Song**, and Jingyi Jessica Li. scReadSim: a single-cell RNA-seq and ATAC-seq read simulator. *Nature Communications*, 14(1):7482, 2023
4. Elvis Han Cui*, **Dongyuan Song***, Weng Kee Wong, and Jingyi Jessica Li. Single-cell generalized trend model (scGTM): a flexible and interpretable model of gene expression trend along cell pseudotime. *Bioinformatics*, 38(16):3927–3934, 2022
5. **Dongyuan Song***, Nan Miles Xi*, Jingyi Jessica Li, and Lin Wang. scSampler: fast diversity-preserving subsampling of large-scale single-cell transcriptomic data. *Bioinformatics*, 38(11):3126–3127, 2022
6. Tianyi Sun, **Dongyuan Song**, Wei Vivian Li, and Jingyi Jessica Li. Simulating single-cell gene expression count data with preserved gene correlations by scDesign2. *Journal of Computational Biology*, 29(1):23–26, 2022
7. Ruochen Jiang, Tianyi Sun, **Dongyuan Song**, and Jingyi Jessica Li. Statistics or biology: the zero-inflation controversy about scRNA-seq data. *Genome Biology*, 23(31), 2022
8. **Dongyuan Song***, Kexin Li*, Zachary Hemminger, Roy Wollman, and Jingyi Jessica Li. scPNMF: sparse gene encoding of single cells to facilitate gene selection for targeted gene profiling. *Bioinformatics*, 37(Supplement_1):i358–i366, 2021
9. Xinzhou Ge, Yiling Elaine Chen, **Dongyuan Song**, MeiLu McDermott, Kyla Woyshner, Antigoni Manousopoulou, Ning Wang, Wei Li, Leo D Wang, and Jingyi Jessica Li. Clipper: p-value-free FDR control on high-throughput data from two conditions. *Genome Biology*, 22(288), 2021
10. Tianyi Sun, **Dongyuan Song**, Wei Vivian Li, and Jingyi Jessica Li. scDesign2: a transparent simulator that generates high-fidelity single-cell gene expression count data with gene correlations captured. *Genome Biology*, 22(163), 2021
11. **Dongyuan Song** and Jingyi Jessica Li. PseudotimeDE: inference of differential gene expression along cell pseudotime with well-calibrated p-values from single-cell RNA sequencing data. *Genome Biology*, 22(124), 2021
12. Elizabeth Christina Miller, Kenji T Hayashi, **Dongyuan Song**, and John J Wiens. Explaining the ocean’s richest biodiversity hotspot and global patterns of fish diversity. *Proceedings of the Royal Society B*, 285(1888):20181314, 2018
13. **Dongyuan Song***, Zhe Wang*, Zhuo-Jun Song, Cheng-Chuan Zhou, Peng-Hao Xu, Jie Yang, Ji Yang, and Bao-Rong Lu. Increased novel single nucleotide polymorphisms in weedy rice populations associated with the change of farming styles: Implications in adaptive mutation and evolution. *Journal of Systematics and Evolution*, 55(2):149–157, 2017

Under review

14. Zachary Hemminger, Gabriela Sanchez-Tam, Haley De Ocampo, Aihui Wang, Thomas Underwood, Fangming Xie, Qiying Zhao, **Dongyuan Song**, Jingyi Jessica Li, Hongwei Dong, et al. Spatial single-cell mapping of transcriptional differences across genetic backgrounds in mouse brains. *Under review at Nature*, 2024
15. **Dongyuan Song***, Kexin Li*, Xinzhou Ge, Christy Lee, and Jingyi Jessica Li. ClusterDE: a post-clustering differential expression (DE) method robust to false-positive inflation caused by double dipping. *bioRxiv*, 2023

16. Zhijian Li, Zain Patel, **Dongyuan Song**, Jingyi Jessica Li Guanao Yan, and Luca Pinello. Benchmarking the efficacy and scalability of computational methods for the identification of spatially variable genes and peaks. *Under review at Genome Biology*, 2024
17. Qingyang Wang, Zhiqian Zhai, **Dongyuan Song**, and Jingyi Jessica Li. Review of computational methods for estimating cell potency from single-cell rna-seq data, with a detailed analysis of discrepancies between method description and code implementation. *In revision at Nature Communications*, 2023
18. Bicna Song, Dingyu Liu, Weiwei Dai, Natalie McMyn, Qingyang Wang, Dapeng Yang, Adam Krejci, Anatoly Vasilyev, Nicole Untermoser, Anke Loregger, et al. Decoding heterogenous single-cell perturbation responses. *Accepted by Nature Cell Biology*, 2023

AWARDS

Dissertation Year Fellowship (\$38,000), UCLA	2023
James P. Taylor Foundation + CSHL Biology of Genomes Scholarship (\$1,600)	2023
Summer Mentored Research Fellowship (\$6,000), UCLA	2021
QCBio Retreat Best Poster Award, UCLA	2021
Outstanding Graduate Student (Top 5%), Fudan University	2017
National Life Science Innovation Competition First Prize, China	2017
Member of National Top Talent Undergraduate Training Program, Fudan University	2017
DuPont First-class Scholarship (Top 3%), Fudan University	2016
First-Class Scholarship Awarded by the NTTUTP (Top 10%), Fudan University	2016

TEACHING

Graduate level

BIOINFO M223 <i>Statistical Methods in Computational Biology</i> : Teaching Assistant	Winter 2022
STATS 205 <i>Hierarchical Linear Models</i> : Teaching Assistant	Fall 2021

Undergraduate level

STATS 100B <i>Introduction to Mathematical Statistics</i> : Teaching Assistant	Winter 2022
STATS 19 <i>Fiat Lux Seminar</i> : Guest Lecturer	Winter 2021

PRESENTATIONS

Invited talks

1. Department of Genome Sciences, University of Virginia 08/2024
Improving statistical rigor in single-cell and spatial omics
2. Department of Genetics and Genome Sciences, UConn Health 06/2024
Improving statistical rigor in single-cell and spatial omics
3. Michael Smith Lab, University of British Columbia 03/2024
Improving statistical rigor in single-cell and spatial omics
4. Eastern North American Region of International Biometric Society (ENAR 2024) 03/2024
scDesign3 generates realistic in silico data for multimodal single-cell and spatial omics

5. Department of Biostatistics, Yale University 03/2024
Improving statistical rigor in single-cell and spatial omics
6. Department of Biostatistics & Bioinformatics, Emory University 02/2024
Improving statistical rigor in single-cell and spatial omics
7. Cancer Institute of New Jersey & Rutgers University 02/2024
Improving statistical rigor in single-cell and spatial omics
8. Department of Biostatistics & Bioinformatics, Moffitt Cancer Center 01/2024
Improving statistical rigor in single-cell and spatial omics
9. Chicago Computational Biology Symposium 12/2023
Improving statistical rigor in single-cell and spatial omics
10. Department of Epidemiology & Biostatistics, University of Georgia 12/2023
Improving statistical rigor in single-cell and spatial omics
11. CMSE department seminar at Michigan State University 10/2023
Improving statistical rigor in single-cell and spatial omics
12. BU-Tsinghua-Keio Workshop 2023: Probability and Statistics 06/2023
In silico data generation and statistical model inference for single-cell and spatial omics
13. Chan Zuckerberg Initiative (CZI) Single-Cell Monthly Webinar 11/2022
Fast diversity-preserving subsampling of large-scale single-cell transcriptomic data
14. 26th Conference on Intelligent Systems for Molecular Biology and the 20th European Conference on Computational Biology (ISMB/ECCB 2021) 07/2021
scPNMF: sparse gene encoding of single cells to facilitate gene selection for targeted gene profiling

Seminar talks

1. UCLA Institute for Quantitative and Computational Biosciences (QCBio) 03/2023
ClusterDE: a post-clustering differentially expressed (DE) gene identification method robust to false-positive inflation caused by double-dipping
2. STAT 300, Department of Statistics, Harvard University 11/2022
scDesign3: an all-in-one statistical framework that generates realistic single-cell omics data and infers cell heterogeneity structure
3. UCLA Institute for QCBio 01/2022
scDesign3: an all-in-one statistical framework that generates realistic single-cell omics data and infers cell heterogeneity structure
4. UCLA Institute for QCBio 10/2020
PseudotimeDE: inference of differential gene expression along cell pseudotime with valid p-values from single-cell RNA sequencing data

Poster presentations

1. Biology of Genomes Meeting, Cold Spring Harbor Laboratory (CSHL) 05/2023
A unified framework of realistic in silico data generation and statistical model inference for single-cell and spatial omics

2. UCLA Jonsson Comprehensive Cancer Center Annual Retreat 05/2022
PseudotimeDE: inference of differential gene expression along cell pseudotime with well-calibrated p-values from single-cell RNA sequencing data
3. Human Cell Atlas (HCA) General Meeting 06/2021
scPNMF: sparse gene encoding of single cells to facilitate gene selection for targeted gene profiling
4. Biology of Genomes Meeting, Cold Spring Harbor Laboratory (CSHL) 05/2021
PseudotimeDE: inference of differential gene expression along cell pseudotime with well-calibrated p-values from single-cell RNA sequencing data

MENTORING

Yuheng Lai, Undergraduate student at Renmin University <i>Fast identification of differentially expressed genes along pseudotime</i>	03/2023-Present
Siqi Chen, Visiting PhD student from Central South University <i>In silico control data generation for spatial domain detection</i>	07/2023-04/2024
Shiyu Ma, Undergraduate student at UCLA <i>Construction of cell-type hierarchy by machine learning on scRNA-seq data</i>	02/2022-05/2023
Lehan Zou, Undergraduate student at UCLA <i>Development of scGTM R package</i>	05/2022-12/2022
Tianyang Liu, Master of Applied Statistics at UCLA <i>Differential expression test along cell pseudotime by quantile non-parametric additive models</i>	02/2021-04/2022
Huy Nguyen, Undergraduate student at UCLA <i>Differential expression test along cell pseudotime by quantile non-parametric additive models</i>	02/2021-03/2022

PROFESSIONAL SERVICE

Reviewer for Scientific Journals (# papers in parentheses):

Nature Communications (1), Science Advances (1), Genome Biology (1), Genome Medicine (1), Cell Systems (1), Bioinformatics (2), Communications Biology (2), BMC Bioinformatics (2), Journal of Computational Biology (1), GigaScience (1), Scientific Reports (3), STAR Protocols (2), IEEE/ACM Transactions on Computational Biology and Bioinformatics (1), Frontiers in Molecular Biosciences (1), Annals of Applied Statistics (1)

PROFESSIONAL AFFILIATIONS

International Society for Computational Biology (ISCB)	2021-2023
American Statistical Association (ASA)	2022-2024
International Biometric Society (IBS)	2024

SOFTWARE PACKAGES

1. **ClusterDE**: a post-clustering DE method for controlling FDR regardless of clustering quality.
R package: <https://github.com/SONGDONGYUAN1994/ClusterDE>

2. **PseudotimeDE**: a DE method that accounts for the uncertainty in pseudotime inference.
R package: <https://github.com/SONGDONGYUAN1994/PseudotimeDE>
3. **scDesign3**: a realistic simulator for multimodal single-cell and spatial omics.
R package: <https://github.com/SONGDONGYUAN1994/scDesign3>
4. **scPNMF**: a dimensionality reduction method to facilitate gene selection
R package: <https://github.com/JSB-UCLA/scPNMF>
5. **scsampler**: diversity-preserving subsampling of large-scale single-cell transcriptomic data
Python package: <https://github.com/SONGDONGYUAN1994/scsampler>

SKILLS

Computer Languages	R, Python, shell script
Tools	Git/GitHub, L ^A T _E X