

# Xiaoru Dong

(217) 778-5683 · [xdong1@ufl.edu](mailto:xdong1@ufl.edu)  
228 Buckman Drive · Gainesville, FL 32603  
[www.linkedin.com/in/xiaoru-dong/](https://www.linkedin.com/in/xiaoru-dong/) · <https://github.com/xiaorudong/>

## EDUCATION

<b>University of Florida</b> <b>Ph.D.</b> , Biostatistics Advisor: Rhonda Bacher, Ph.D. Dissertation: Improving Analysis and Modelling of Dynamic, High-Dimensional Single-Cell Data	Gainesville, FL May 2024
<b>University of Illinois at Urbana-Champaign</b> <b>B.S.</b> , Statistics	Champaign, IL Dec 2018

## RESEARCH INTERESTS

My research interests center on the intersection of biology and statistics, focusing on developing statistical and computational tools for genomics and biomedical research. I specialize in analyzing complex datasets, especially human single-cell data from next-generation sequencing technologies using advanced statistical techniques. My aim is to drive novel biological discoveries and enhance our understanding of genomic processes through innovative and data-driven approaches.

## EXPERIENCE

<b>University of Florida</b> <i>Graduate Research Assistant</i>	Gainesville, FL Aug 2019- May 2024
<ul style="list-style-type: none"><li>&gt; Collaborate with Dr. Todd Brusko and Dr. Philip Efron and their teams at the University of Florida College of Medicine on data analysis projects in immunology, focusing on type 1 diabetes and sepsis.</li><li>&gt; Conduct advanced analysis using existing software packages and developing custom scripts on single-cell RNA-seq (e.g. 10X) and bulk RNA-seq data, identifying critical disease-associated factors and biomarkers.</li><li>&gt; Apply machine learning methods, such as random forest and lasso, to analyze high-dimensional omics datasets, including genomic, epigenomic, cytokine, and immunophenotyping data.</li><li>&gt; Contribute to study design, data analysis, preparation of statistical sections in grants for applications and the publication of findings in scientific journals.</li><li>&gt; Develop novel computational methods for scRNA-seq data analysis using R, including the creation of innovative algorithms for data-driven scRNA-seq trajectory evaluation aimed at identifying the most accurate trajectory.</li><li>&gt; Provide mentorship and support in R programming to lab members, aiding in data analysis tasks and troubleshooting.</li></ul>	

<b>University of Florida</b> <i>Graduate Teaching Assistant</i>	Gainesville, FL Aug 2019-May 2023
<ul style="list-style-type: none"><li>&gt; Mentored undergraduate and graduate students across various departments in five classes including Public Health Computing, Biostatistical Methods, and Statistical Analysis of Genetic Data, enhancing engagement and performance in both online and in-person classes (20+ students per class).</li><li>&gt; Provided weekly office hours for coding guidance in R and SAS, clarifying complex concepts and contributing to the students' practical understanding of statistical analysis.</li><li>&gt; Led interactive classroom discussions and facilitated group activities to encourage critical thinking and foster a collaborative learning environment.</li><li>&gt; Created and implemented innovative teaching materials, including assignments and digital learning platforms (Canvas), leading to improved student learning outcomes and course effectiveness.</li><li>&gt; Delivered detailed grading and constructive feedback, contributing to measurable improvements in student performance and academic success.</li></ul>	

## PUBLICATIONS & PREPRINTS

Key: † indicates co-first authors; \* indicates corresponding authors.

1. **Dong, X.**, Leary, J. R., Yang, C., Brusko, M. A., Brusko, T. M., & Bacher, R.\* (2023). *Data-driven selection of analysis decisions in single-cell RNA-seq trajectory inference* (p.2023.12.18.572214). bioRxiv. <https://doi.org/>

2. Shapiro, M. R.<sup>†</sup>, **Dong, X.**<sup>†</sup>, Perry, D. J.<sup>†</sup>, McNichols, J. M., Thirawatananond, P., Posgai, A. L., Peters, L. D., Motwani, K., Musca, R. S., Muir, A., Concannon, P., Jacobsen, L. M., Mathews, C. E., Wasserfall, C. H., Haller, M. J., Schatz, D. A., Atkinson, M. A., Brusko, M. A., Bacher, R.<sup>\*</sup>, & Brusko, T. M.<sup>\*</sup> (2023). Human immune phenotyping reveals accelerated aging in type 1 diabetes. *JCI Insight*. <https://doi.org/10.1172/jci.insight.170767>
3. Jacobsen, L. M.<sup>†</sup>, Diggins, K.<sup>†</sup>, Blanchfield, L.<sup>†</sup>, McNichols, J. A., Perry, D. J., Brant, J., **Dong, X.**, Bacher, R., Gersuk, V. H., Schatz, D. A., Atkinson, M. A., Mathews, C. E., Haller, M. J., Long, S. A.<sup>\*</sup>, Linsley, P. S.<sup>\*</sup>, & Brusko, T. M.<sup>\*</sup> (2023). Responders to low-dose ATG induce CD4 T cell exhaustion in type 1 diabetes. *JCI Insight*. <https://doi.org/10.1172/jci.insight.161812>
4. **Dong, X.**, & Bacher, R. (2023). Analysis of Single-Cell RNA-seq Data. In B. Fridley & X. Wang (Eds.), *Statistical Genomics* (Vol. 2629, pp. 95–114). Springer US. [https://doi.org/10.1007/978-1-0716-2986-4\\_6](https://doi.org/10.1007/978-1-0716-2986-4_6)
5. **Dong, X.**, & Bacher, R. (2022). Data-driven assessment of dimension reduction quality for single-cell omics data. *Patterns*, 3(3), 100465. <https://doi.org/10.1016/j.patter.2022.100465>
6. Darden, D. B., **Dong, X.**, Brusko, M. A., Kelly, L., Fenner, B., Rincon, J. C., Dirain, M. L., Ungaro, R., Nacionales, D. C., Gauthier, M., Kladde, M., Brusko, T. M., Bihorac, A., Moore, F. A., Loftus, T., Bacher, R., Moldawer, L. L., Mohr, A. M., & Efron, P. A.<sup>\*</sup> (2021). A Novel Single Cell RNA-seq Analysis of Non-Myeloid Circulating Cells in Late Sepsis. *Frontiers in Immunology*, 12, 696536. <https://doi.org/10.3389/fimmu.2021.696536>
7. Ross, J. J., Wasserfall, C. H., Bacher, R., Perry, D. J., McGrail, K., Posgai, A. L., **Dong, X.**, Muir, A., Li, X., Campbell-Thompson, M., Brusko, T. M., Schatz, D. A., Haller, M. J., & Atkinson, M. A. (2021). Exocrine Pancreatic Enzymes Are a Serological Biomarker for Type 1 Diabetes Staging and Pancreas Size. *Diabetes*, 70(4), 944–954. <https://doi.org/10.2337/db20-0995>

## SOFTWARE

### GitHub software package

Escort: Methodology and an R package implementing Escort. This method evaluates a single-cell RNA-seq dataset's suitability for trajectory inference and for quantifying trajectory properties influenced by analysis decisions. [Available at: GitHub].

## PRESENTATIONS

### National talk

Invited talk: “Data-driven Selection of Trajectory in Single-cell RNA-seq Data.” International Biometric Society Eastern North American Region (ENAR) Annual Meeting. Mar 2024. Baltimore, MD, USA.

### Local talks

1. “Data-driven Trajectory Construction in Single-Cell RNA-seq Data.” Mark C. K. Yang Award Ceremony. Oct 2023. University of Florida, Gainesville, FL, USA.
2. “Linear Regression with Highly Correlated Variables.” Jan 2023. Brusko Lab, University of Florida Diabetes Institute, Gainesville, FL, USA.
3. “Exploring the CausalCell Approach to Single-Cell Data.” Sept 2022. Brusko Lab, University of Florida Diabetes Institute, Gainesville, FL, USA.
4. “Human immune phenotyping (HIP): Age Prediction Model.” Apr 2022. Brusko Lab, University of Florida Diabetes Institute, Gainesville, FL, USA.
5. “Trajectory Goodness of Fit in Single-Cell RNA-seq Data.” Dec 2021. Brusko Lab, University of Florida Diabetes Institute, Gainesville, FL, USA.
6. “Understanding RNA Velocity.” Aug 2021. Brusko Lab, University of Florida Diabetes Institute, Gainesville, FL, USA.
7. “Gene Selection for Trajectory Inference in Single-Cell RNA-seq Data.” Jan 2021. Brusko Lab, University of

Florida Diabetes Institute, Gainesville, FL, USA.

### National posters

1. "Data-driven Selection of Trajectory in Single-cell RNA-seq Data." American Statistical Association (ASA) Section on Statistics in Genomics and Genetics (SSGG). May 2024. Pittsburgh, PA, USA.
2. "Data-driven Evaluation of Trajectories in Single-Cell RNA-seq Data." International Biometric Society Eastern North American Region (ENAR) Annual Meeting. Mar 2023. Nashville, TN, USA.

### Local posters

1. "Immune Phenotypic Profiling across the Human Lifespan Demonstrates Accelerated Immune Aging in Subjects with Type 1 Diabetes." 2022 College of Medicine Research Day. Apr 2022. University of Florida, Gainesville, FL, USA.
2. "Analyzing Inclusion Criteria of 7000 Cochrane Systematic Reviews." Undergraduate Research Symposium. Apr 2018. University of Illinois at Urbana-Champaign, Champaign, IL, USA.
3. "Distortions in Scientific Literature - A Replication Analysis of Greenberg's Citation Network of 302 Alzheimer's Science Research Papers." Undergraduate Research Symposium. Apr 2018. University of Illinois at Urbana-Champaign, Champaign, IL, USA.

### TEACHING

---

*Teaching Assistant*, University of Florida

- > PHC4094 - Introduction to Biostatistics for Health Science and Public Health. Fall 2021.
- > PHC6052 - Introduction to Biostatistical Methods. Fall 2019 & Fall 2020.
- > PHC6059 - Introduction to Applied Survival Analysis. Fall 2022.
- > PHC6088 - Statistical Analysis of Genetic Data. Spring 2021, Spring 2022 & Spring 2023.
- > PHC6089 - Public Health Computing. Spring 2020.

### ACADEMIC SERVICE

---

#### Journal reviewer

- > BMC Bioinformatics, Scientific Reports, Bioinformatics

#### Service in statistics

- > Session moderator. Annual Meeting of the Florida Chapter of the American Statistical Association (ASA). Mar 2023.
- > Session moderator. Annual Applied Statistics Symposium of the International Chinese Statistical Association (ICSA). Jun 2022.

### HONORS & AWARDS

---

Mark C. K. Yang Student Presentation Award	Oct 2023
Student Competition Award, American Statistical Association Florida Chapter	Mar 2023
Certificate in Data Science	Nov 2018
Successful Participant Award, Mathematical Contest in Modeling	Jan 2016

### PROFESSIONAL SOCIETY MEMBERSHIPS

---

International Biometric Society (ENAR), American Statistical Association (ASA)

### SKILLS

---

- > **Programming Languages:** Proficient in R/Rstudio for data analysis and visualization, complemented by skills in SAS and Python for diverse statistical applications.
- > **Statistical Analysis Techniques:** Proficient in a variety of statistical methods for biomedical research, including multivariable analysis and machine learning algorithms. Experienced with bioinformatics tools for RNA-seq analysis, notably Seurat and DESeq2.
- > **Data Coordination:** Skilled in managing and coordinating multi-site data, ensuring data integrity and consistency across collaborative research projects.
- > **Consulting & Mentorship:** Effective in providing statistical consulting and mentorship, aiding researchers and students in data analysis and interpretation across diverse fields.

- > **Communication:** Fluent in English and Chinese languages, with strong written and verbal communication skills, facilitating collaboration and scientific dissemination.