

DONGYUAN WU

DOCTORAL STUDENT

Department of Biostatistics
University of Florida
2004 Mowry Rd, CTRB 5th Floor
Gainesville, FL 32610
☎ +1 (352) 870 4802
✉ dongyuanwu@ufl.edu
📄 dongyuanwu.github.io

Last updated on November 16, 2022

EDUCATION

University of Florida <i>Doctor of Philosophy in Biostatistics</i> GPA: 3.98/4.00	Gainesville, FL Aug. 2020 - present
University of Florida <i>Master of Science in Biostatistics</i> GPA: 4.00/4.00	Gainesville, FL Aug. 2018 - May 2020
Minzu University of China <i>Bachelor of Science in Applied Statistics</i> GPA: 3.77/4.00	Beijing, China Sept. 2013 - July 2017

WORK EXPERIENCE

Graduate Research Assistant <i>Department of Biostatistics, University of Florida</i> The Role and Mechanisms of Lipid and Lipoprotein Dysregulation in Sepsis <ul style="list-style-type: none">○ Analyzed demographic and clinical data with descriptive statistics, examined the existence of significant differences among different groups of sepsis patients.○ Aligned RNA sequencing data into expression matrix, did further downstream analyses such as differential expression analysis, gene co-expression network analysis, functional annotation, and so on.○ Dealt with missing values and unbalanced issues of lipidomics data, and then conducted suitable comparisons based on different clinical outcomes, phenotypes, etc.○ Did power analysis and calculated sample sizes for grant purpose.	Gainesville, FL Aug. 2020 - present
Graduate Teaching Assistant <i>Department of Biostatistics, University of Florida</i> <ul style="list-style-type: none">○ Held office hours to respond to students' questions, graded assignments and exams, and helped with courses development.	Gainesville, FL Aug. 2020 - present
OPS Student Research Assistant <i>Department of Community Health and Family Medicine, University of Florida</i> <ul style="list-style-type: none">○ Collaborated with investigators to identify problems and provided biostatistical consultation including analysis of data, interpretation of results, and preparation of reports.	Gainesville, FL May 2019 - Apr. 2020
Research Assistant <i>Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences</i> <ul style="list-style-type: none">○ Provided a variety of statistical analysis for projects using appropriate statistical and computing methodologies, and assisted in the interpretation and presentation of results.○ Imported data from SQL, generated reproducible templates for adverse drug reaction weekly report, monthly report, and annual report by using R Markdown, and assisted to develop a platform that can automatically display these reports for different drugs.	Beijing, China July 2017 - June 2018

RESEARCH
EXPERIENCE

Cell-Cell Communication on Spatial RNA-seq Data

Supervisor: Dr. Susmita Datta

Gainesville, FL
Jan. 2022 - present

- Propose a generalized linear regression model to infer cell-cell communication from spatial transcriptomics data.
- Apply Hamiltonian Monte Carlo algorithm with Horseshoe priors to estimate and infer the communication strength of a specific ligand-receptor pair between two cell types.

Immune Cell Responses to SARS-CoV-2 Infection

Supervisor: Dr. Susmita Datta

Gainesville, FL
Apr. 2021 - Feb. 2022

- Preprocessed datasets, such as filtering, normalization, and resampling.
- Applied MAST for multiple single-cell RNA sequencing data related to immune cells, to figure out the differential expression genes among healthy donors, mild COVID-19 patients, and severe COVID-19 patients.
- Made gene co-expression networks using PIDC based on healthy, mild COVID-19, and severe COVID-19, separately. Then found the hub genes in modules and detected the differentially connected genes.
- Built random forests to calculate the Gini importance scores for genes.
- Did functional annotation for those detected important genes to improve the insight of gene expression and network patterns associated with immune cells for SARS-CoV-2 infection.

Analysis of Alzheimer's Disease scRNAseq Data

Supervisor: Dr. Susmita Datta

Gainesville, FL
Sept. 2019 - Apr. 2020

- Used 2-dimensional t-SNE plots to visualize the single-cell RNA sequencing data.
- Preprocessed dataset, such as splitting, filtering, normalization, and clustering.
- Applied hurdle models specifically designed for sequencing-based single-cell gene expression data, including CRE and MAST, to detect differentially expressed genes between AD-pathology and no-pathology.
- Did network analysis and functional annotation for those differentially expressed genes.

Risk Factors of ADR for a Traditional Chinese Medicine

Supervisor: Dr. Wei Yang

Beijing, China
Oct. 2016 - May 2017

- Processed 30,888 data in advance, such as data cleaning, data standardization.
- Analyzed all data with descriptive statistics, examined the existence of significant differences between normal group and adverse drug reaction (ADR) group by applying various hypothesis testing methods, and reported the results by using R Markdown.
- Associated resampling methods, including RUS, ROS, and SMOTE, with classification algorithms, such as decision trees, AdaBoost, random forests, and LASSO, to improve the accuracy of classifiers for the minority class (i.e., ADR group).

Evaluation of TCM Clinical Practice Guidelines

Supervisor: Dr. Wei Yang

Beijing, China
Aug. 2015 - Sept. 2016

- Processed more than 20,000 items of Traditional Chinese Medicine (TCM) data in advance, including data cleaning, data standardization.
- Analyzed data in 76 TCM syndrome and 11 subjects with descriptive statistics, and reported the results by using R Markdown based on the idea of reproducible research.

REFEREED
PUBLICATIONS

DelRocco, N., Wang, Y., **Wu, D.**, Yang, Y., Shan, G. (2022). New Confidence Intervals for Relative Risk of Two Correlated Proportions. *Statistics in Biosciences*. doi: 10.1007/s12561-022-09345-7

Wu, D., Zhang, R., and Datta, S. (2022). Unraveling T Cell Responses for Long Term Protection of SARS-CoV-2 Infection. *Frontiers in Genetics*, 13:871164. doi: 10.3389/fgene.2022.871164

Malaty, J., **Wu, D.**, and Datta, S. (2021). Flexible Nasal Endoscopic Procedures in Family Medicine: Indications and Effectiveness. *Family Medicine*, 53(10), 886-889. doi: 10.22454/FamMed.2021.332061

Ellis, D., **Wu, D.**, and Datta, S. (2021). SAREV: A review on statistical analytics of single-cell RNA sequencing data. *Wiley Interdisciplinary Reviews: Computational Statistics*, e1558. doi: 10.1002/wics.1558

ORAL PRESENTATIONS “UTRCOV2: Unraveling T cell responses for long term protection of SARS-COV-2 infection.” Annual International Conference on Critical Assessment of Massive Data Analysis. July 2021. Virtual.

POSTER PRESENTATIONS “UTRCOV2: Unraveling T cell responses for long term protection of SARS-COV-2 infection.” UF College of Medicine Research Day. April 2022. Gainesville, FL.
 “COVID-19: Reduced Lung Function and Increased Psycho-emotional Stress.” UF College of Public Health & Health Professions Research Day. February 2021. Virtual.

SOFTWARES

- [PairedRR](#): A Shiny app to calculate the confidence intervals for relative risk of two correlated proportions.
- [RSBID](#): An R package to provide resampling strategies for binary imbalanced datasets.

AWARDS AND HONORS

Outstanding Achievement Award	<i>International Center, University of Florida</i>	2019, 2022
Student Travel Award	<i>Department of Biostatistics, University of Florida</i>	2021
Grinter Fellowship	<i>University of Florida</i>	2020
Outstanding Master Graduate	<i>Department of Biostatistics, University of Florida</i>	2020
First-class Scholarship (TOP 3%)	<i>Minzu University of China</i>	2017
Honorable Mention	<i>Interdisciplinary Contest in Modeling</i>	2016
First Prize in Beijing Region	<i>National Mathematical Modeling Contest</i>	2016
Wu Xianhong Scholarship	<i>College of Science, Minzu University of China</i>	2015, 2016

SKILLS R, SAS, Python, L^AT_EX, C/C++, WinBUGS, MySQL

CERTIFICATES

- [SAS Certified Professional: Advanced Programming Using SAS 9.4](#) (2020)
- [SAS Certified Specialist: Base Programming Using SAS 9.4](#) (2019)
- [Applied Data Science with Python Specialization](#) by UMich on Coursera (2020)
- [Genomic Data Science Specialization](#) by JHU on Coursera (2020)
- [Data Science Specialization](#) by JHU on Coursera (2019)