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 September 16, 2022

Education

University of Florida — Gainesville, FL
Doctor of Philosophy in Biostatistics — *Aug. 2020 - present*

University of Florida — Gainesville, FL
Master of Science in Biostatistics — *Aug. 2018 - May 2020*

Minzu University of China — Beijing, China
Bachelor of Science in Applied Statistics — *Sept. 2013 - July 2017*

Work Experience

Graduate Research Assistant — Gainesville, FL
Department of Biostatistics, University of Florida — *Aug. 2020 - present*

The Role and Mechanisms of Lipid and Lipoprotein Dysregulation in Sepsis

- 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.

Graduate Teaching Assistant — Gainesville, FL
Department of Biostatistics, University of Florida — *Aug. 2020 - present*

Grading and office hour responsibilities

- PHC 6052: Introduction to Biostatistical Methods (Fall 2021)
- PHC 6051: Biostatistical Methods II (Spring 2021)
- PHC 6020: Clinical Trial Methods (Spring 2021)
- PHC 6790: Biostatistical Methods Using SAS (Fall 2020)
- PHC 6050: Statistical Methods for the Health Sciences I (Fall 2020)
- PHC 6089: Public Health Computing (Fall 2020)

Helping with courses development

- PHC 6020: Clinical Trials Methods (Summer 2022)
- PHC 6088: Statistical Analysis of Genetic Data (Spring 2022)
- PHC 6097: Statistical Learning with Applications in Health Sciences (Summer 2021)

OPS Student Research Assistant — Gainesville, FL
Department of Community Health and Family Medicine, University of Florida — *May 2019 - Apr. 2020*

- Collaborated with investigators to identify problems and provided biostatistical consultation including analysis of data, interpretation of results, and preparation of reports.

Research Assistant Beijing, China
Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences *July 2017 - June 2018*

- 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.

Research Experience

Cell-Cell Communication on Spatial RNA-seq Data Gainesville, FL
Supervisor: Dr. Susmita Datta *Jan. 2022 - present*

- Inferring cell-cell communication from spatial transcriptomics data.

Immune Cell Responses to SARS-CoV-2 Infection Gainesville, FL
Supervisor: Dr. Susmita Datta *Apr. 2021 - Feb. 2022*

- Preprocessed datasets, such as filering, 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 Gainesville, FL
Supervisor: Dr. Susmita Datta *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 Beijing, China
Supervisor: Dr. Wei Yang *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 Beijing, China
Supervisor: Dr. Wei Yang *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: An 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

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

Skills

R, SAS, Python, LaTeX, 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)