Dongyuan Wu

Graduate 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 July 3, 2021

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.

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

Grading and office hour responsibilities

- 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)

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

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.

Publications

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.

Wu, D., Ellis, D., and Datta, S. (2020). COVID-19: Reduced Lung Function and Increased Psycho-emotional Stress. *Bioinformation*, 16(4), 293-296.

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

"COVID-19: Reduced Lung Function and Increased Psycho-emotional Stress." UF PHHP Research Day. February 2021. Virtual.

Softwares

- RSBID: Resampling Strategies for Binary Imbalanced Datasets
- COVID19FL: Florida's COVID-19 Data Visualization

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, Photoshop, Illustrator, MS Office

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)