# Chunhui Gu

7900 Cambridge St ● Houston, TX 77054 ●chunhui.gu@uth.tmc.edu ● 404.952.6630

## EDUCATION

**The University of Texas Health Science Center at Houston**, Houston, Texas September 2020 - Present
Doctor of Philosophy in Biostatistics Cumulative GPA: 3.916/4.0

**Rollins School of Public Health, Emory University**, Atlanta, GA September 2018 - May 2020
Master of Science in Public Health, Biostatistics Cumulative GPA: 3.96/4.0
**Certification: SAS Base Programming Certification**

**Fudan University**, Shanghai, China September 2013 - July 2018
Bachelor of Medicine in Clinical Medicine Rank: 80/203

## SKILLS QUALIFICATIONS

**Experience:** 2 years of part-time working experience in bioinformatics and biomedical informatics and 1-year full-time internship in specialist hospitals and general hospitals
**SQL:** familiar with using SQL in SAS, R (SQLite), Teradata, and MySQL
**Programming:** Proficient in R, SAS, and Python; familiar with HTML, Java, and C
**Writing:** Excellent academic writing skills developed through curriculum projects, work, and thesis
**Analytical**: Strong analytical skills developed through Research Assistant positions
**Team-work:** 2 years of Research Assistant experience in multi-disciplinary teams (biologists, doctors, epidemiologists, and computer scientists)

## EXPERIENCE

*Graduate Research Assistant |Advisor: Prof. Ehsan Irajizad, Ph.D.* June 2022 – Present
*Supervisor: Prof. Kim-Anh Do, Ph.D.*

**In-depth proteomic profiling of gastric cancer tumor cells uncovers previously unidentified cancer testis antigens**

- Learned structure and properties of mass spectrometry proteomics data
- Built in-house R programming package for automating proteomics analysis pipeline including special data structure, statistical testing, and visualization modules
- Identified missing proteins using combined information in mass spectrometry protein expression and mRNA expression

**Using graph neural networks with protein-protein-interaction for enhancing protein identification**

- Reviewed status quo in using information from other sources to improve mass spectrometry-based protein detection
- Learned current state-of-art graph neural network models, such as GCNConv, GraphSAGE, and GAT
- Implemented graph neural network deep learning framework for enhancing mass spectrometry proteomics protein detection with protein-protein interaction information

*Graduate Research Assistant |Advisor: Prof. Guoqiang Zhang, Ph.D.* September 2020 - May 2022

**A review of propensity-score matching method in studies that explore factors associated with COVID-severity**

- Reviewed use of propensity-score matching method in selected papers
- Generalized commonly used propensity-score models, matching strategies, and covariates used in this topic

- Summarized shortcomings in using propensity-score matching in this topic

**Studied how to query through MongoDB database and explore the effect of hydroxychloroquine in treating COVID-19 by propensity-score matching analysis**

- Performed basic MongoDB operations, aggregation framework, improving MongoDB performance, MongoDB Python developer, and MongoDB diagnostics and debugging
- Studied theory and application of propensity-score methods in covariate selection, missing imputation, propensity-score model selection, diagnostics, and analysis
- Reviewed published papers exploring effect of hydroxychloroquine in treating COVID-19
- Two curriculum projects were done for practice purposes:
  - Stochastic process I: Comparing the length of stay of COVID-19 patients receiving and not receiving hydroxychloroquine treatment using discrete-time Markov chain
  - Multivariate analysis: Estimating the effect of preventative use of hydroxychloroquine and interaction between age in four severity biomarkers after COVID infection

**Does Cancer History Drive COVID-19 Outcome? A Large-scale Matched Cohort Analysis**

- Debugged SAS code from other biostatisticians
- Translated SAS code into R code to test the compatibility of R and SAS in our routine analyses
- Developed R in-house packages to automate analysis pipeline
- Offered additional expertise from clinical medicine and biostatistics perspective in topic development and modeling

*Research Assistant |Advisor: Prof. Nelson Chen, Ph.D.* May 2021 - May 2022

**Expansion of a Risk Prediction Model for Healthcare facility-onset Clostridioides difficile infection in Patients Receiving Systemic Antibiotics**

- Conducted ANOVA to compare mean of covariates of patients among case and control groups, constructed multivariate logistic regression model with variables significantly associated with hospital-onset CDI, and evaluated at hospital admission.
- Evaluated regression model and point-based risk prediction model by receiver operating characteristic curve (ROC curve), positive predictive value, negative predictive value, sensitivity, specificity, and accuracy at various point cutoffs.

**Oral Thiazide Diuretic Comparison in Acute Decompensated Heart Failure**

- Compared efficacy and safety of two thiazide diuretic in weight change, 24-hour-urine-output (UOP), and length of stay with Chi-square and ANCOVA in SAS.
- Used multivariate logistic regression model to analyze ICU transfer rate between two groups after adjusting for other factors.

## INTERNSHIP EXPERIENCE

**Children's Hospital of Atlanta,** Atlanta, U.S. May 2019 - May 2020

*Intern bioinformatics analyst |Advisor: Prof. Rabindra M Tirouvanziam, Ph.D.*

- Processed raw data from microarray and RNA-seq using R (beadarray) and Linux command line (bowtie)
- Analyzed microarray and RNA-seq data by differential expression analysis and genes enrichment analysis using R (limma and DESeq2). Summarized analysis results with chromosome ideogram, heatmap, bubble chart, and other plots by JavaScript, Python, and R.
- Built well-documented in-house R packages for automated RNA-seq data analysis and plotting.
- Conducted research on genes associated with a good prognosis in severe influenza infection by PCA, ingenuity pathway, modular analysis, Cibersort, etc.

**Fudan University's Huashan Hospital,** Shanghai, China October 2017 - April 2018

*Clinical Rotation and Surgical Assistant | Advisor: Zhiwei Yang, M.D.*

- Requested medical history from the inpatient department. Performed dermal suture 10 times and retraction with supervising doctors over 30 times. Conducted promulgated follow-ups about medical knowledge.
- Used quantitative and qualitative epidemiology methods to conduct and analyze community diagnosis to prevent high incidence of diseases in communities.

**Pathogenic Microbes Department of Fudan University**, Shanghai, China June 2016 - August 2016

*Researcher Assistant | Advisor: Prof. Xunjia Cheng, Ph.D.*

- Completed basic operations independently, such as gel electrophoresis, plasmid extraction, and PCR.
- Studied and finished advanced operations such as fluorescent focal microscopy operation, PCR primer design, and the culture of engineering bacteria by the method of plasmid transfer.

## HONORS

| | |
|---|---|
| Fudan University Outstanding Graduate Student Scholar | 2018 |
| Third prize-winning contest group in Clove Programming Competition | 2016 |
| Fudan University Academic Excellent Scholarship | 2013/2014/2015 |
| Excellent Volunteer of Shanghai Children's Museum | 2014 |

## VOLUNTEER EXPERIENCE

**Children's Museum,** Shanghai, China February 2014 - February 2015

Organizer and Volunteer

- Led all the volunteers and organized the museum schedule, ensuring a smooth working environment for the volunteers and employers.
- Guided and supervised over 300 children around the museum.

## ADDITIONAL SKILLS

**Statistical Software and Programming:** SAS, Python, R programming, SQL, SUDAAN, Java, C, SPSS, Linux command line, MongoDB

**Technique:** Propensity-score methods, Linear and Non-linear regression, Bayesian models, Survival analysis, Complex survey design and analysis, Clinical trials, Simulation, Times series, Differential expression analysis

**Software:** Microsoft Office (Word, Excel, and PowerPoint), Photoshop, Premiere, and Lightroom

**Languages**: English (fluent), Mandarin Chinese(fluent)

## PUBLICATIONS

**Gu, C**., Cai, Y., Hsiao, F.-C., Dennison, J., Vykoukal, J., Farhmann, J., Do, K.-A., Ajani, J. A., Katayama, H., Irajizad, E., & Hanash, S. (Addressing reviews). In-depth proteomic profiling of gastric cancer tumor cells uncovers previously unidentified cancer testis antigens. Cells.

Kim, Y., Zhu, L., Zhu, H., Li, X., Huang, Y., **Gu, C**., Bush, H., Chung, C., & Zhang, G.-Q. (2022). Characterizing cancer and COVID-19 outcomes using electronic health records. PLoS One, 17(5), e0267584.
Margaroli, C., Moncada Giraldo, D., Arafat Gulick, D., Dobosh, B., Giacalone, V., Forrest, O., Sun, F., Gu, C., Kissick, H., & Wu, R. (2021). Broad transcriptional firing represses bactericidal activity in human airway neutrophils. Available at SSRN 3543838.

Margaroli, C., Moncada-Giraldo, D., Gulick, D. A., Dobosh, B., Giacalone, V. D., Forrest, O. A., Sun, F., **Gu, C.,** Gaggar, A., & Kissick, H. (2021). Transcriptional firing represses bactericidal activity in cystic fibrosis airway neutrophils. Cell Reports Medicine, 2(4), 100239.

Tilton, C. S., Sexton, M. E., Johnson, S. W., **Gu, C.,** Chen, Z., Robichaux, C., & Metzger, N. L. (2021). Evaluation of a risk assessment model to predict infection with healthcare facility–onset Clostridioides difficile. American Journal of Health-System Pharmacy, 78(18), 1681-1690.

## Other Education Programs:

**Coursera Online Courses:**
Python for Everyone Specialization|University of Michigan
Introduction to Scripting in Python Specialization|Rice University
Bayesian Statistics: From Concept to Data Analysis| University of California, Santa Cruz
Bayesian Statistics: Techniques and Models|University of California, Santa Cruz
Version Control with Git| Atlassian
Genomic Data Science Specialization|Johns Hopkins University
MongoDB University courses: Aggregation framework, performance, Python developer, Diagnostics, and debugging