DAVID HUANG

Raleigh, NC
919-914-2614 | hczdavid@gmail.com
https://www.linkedin.com/in/caizhi-david-huang-9bb04a139/

CORE SKILL SET

- Excellent in R programming skills (statistical modeling, data manipulation and visualization, R package/app development)
- Extensive statistical methods/modeling and interpretation experience in SAS and R
- Extensive experience with metagenomic data (bioinformatics and statistics)
- Extensive experience with advanced statistical methods, machine learning and deep learning
- Proficient in Python, Julia, Shell scripting and Linux environment

- Proficient in scientific writing and presentation
- Multiple productive summer internship experiences with outstanding performance.
- Strong teamwork ability with multiple research collaborations in both academic and industry
- Multiple first-author research publications/manuscripts

EDUCATION

North Carolina State University, Raleigh, North Carolina: 2018 to Present
Ph.D. Candidate – Bioinformatics (focus on statistical genetics) (GPA 3.9)
Master of Statistics (concentrate on Biostatistics with statistic Ph.D. level course training) (GPA 3.96)

- **Relevant Coursework**: Statistical Inference, Neural Networks, Statistical Learning, Statistical Method, Bioinformatics, Linear Model, Genetic/Categorical/Multivariate/Longitudinal Data Analysis, Computational Methods for Molecular Biology

Central South University, Changsha, Hunan, China: 2010 to 2014
Bachelor of Science – Bioengineering (GPA 3.58)

WORK EXPERIENCE

Procter & Gamble | Mason, OH, USA | 2022 Summer
R&D Intern - Ph.D. Statistician
- Implemented connectivity map (Cmap) algorithms and designed the analysis strategy to do method comparison. Developed multi-omics data analysis R Shiny tool including data integration, differential abundance & biological functional analysis, and predicting model using machine learning. (Manager evaluation: Excessed expectations)

Reynolds American Inc. | Winston-Salem, NC, USA | 2021 Summer
Scientific & Regulatory Affairs Intern
- Analyzed the longitudinal PATH dataset using a mixed effect model and multistate Markov model to understand the harm reduction effect of e-cigarette and smokeless products. Analysis was performed in R and SAS. The manuscript is under preparation. (Manager evaluation: Outstanding)

Merck & Co. | Philadelphia, PA, USA (Remote) | 2020 Summer
Biostatistics Intern
- Developed visualization tools (R package and Shiny app) for immunogenicity results to draw forest plot and RCDC plot. Proposed a pseudo-value approach to compare different generations of multivalent vaccines studied based on a constrained multiple linear regression model. The manuscript is under preparation. (Manager evaluation: Outstanding)

National Institute of Environmental Health Science, Biostatistics & Computational Biology Branch | 2019 Summer
Biostatistics Intern
- Developed a CPU-parallel version of a complete deconvolution method for bulk RNA-Seq data and evaluated the method using synthetic data from the single-cell sequencing. Co-first author paper was published. The method was implemented in the *CDSeq* R package at CRAN.

North Carolina Central University | Durham, North Carolina, USA | 2015 to 2018
Research Technician and Lab manager
- Managed the biological lab including funding, order, equipment, and training. Supported the lab statistical analysis and data visualization using R. Conducted animal experiments and cell culture to study how ethanol promotes oro-esophageal carcinogenesis through Notch-Pax9 signaling.
- Led the analysis project about using peak-random forest model to predict oral cancer risk in patients with oral leukoplakia.

North Carolina State University, Bioinformatics Research Center| 2018 to Present
Research Assistant
Phylogeny-guided Microbiome OTU-specific Association Test (leading)
- Developed a novel association test method for microbiome data using kernel machine model to detect the association between phenotypes and OTUs. The manuscript is under review. The method was implemented in the *POSTm* R package at CRAN.

Taxa-specific Longitudinal Microbiome Association Test (leading)
- Developing a novel longitudinal microbiome data association test method using kernel machine and functional regression model to access the temporal impact of microbiome on clinical outcome.

A Meta-analysis of The Vaginal Microbiome and Preterm Birth (leading)
- Processed 12 amplicon sequence datasets at a Linux environment and conducted the cross-study analysis using machine learning, Bayesian approach and mixed effect model. The manuscript is under preparation.

Terrain Identification from IMU Streams using a Recurrent Neural Network (co-leading)
- Processed the time series data with overlapping/slicing; employed SMOTE for data augmentation; implemented the recurrent neural network with LSTM cell using Keras of TensorFlow in Python; tuned the structure and hyperparameters based on the model performance evaluated using training/testing loss and F1 score.

National Institute of Environmental Health Science, Biostatistics & Computational Biology Branch | 2020 to Present
Special Volunteer
Identifying Clinical Parameters that are Associated with The Severity of COVID-19 Infection (co-leading)
- Annotated the electronic healthy records (EHR) data and classified patients based on COVID-19 severity; analyzed the associations between COVID severity and clinical parameters using statistical testing, ordinal logistic regression and machine learning.

North Carolina Central University | Durham, North Carolina, USA | 2019 to 2021
Collaborator
Development of Targeted Therapy of NRF2high Esophageal Squamous Cell Carcinoma
- Analyzed RNA-Seq data by using an unsupervised approach, performing gene-set and differential abundance analysis.

ADDITIONAL CREDENTIALS

SKILLS	- Software: R, Python, SAS, Julia, Shell script, Latex, Microsoft Office, Endnote - Experiment Skills: Molecule experiment, Cell culture experiment, Animal experiment
R PACKAGE & APP	- CDSeq R package: a complete deconvolution method. (CRAN) - POSTm R package: phylogeny-guided microbiome associate test (CRAN) - VSARgraph R package/app: visualization tools for immunogenicity results (Merck internal use)
HONORS & AWARDS	- GGI 2021 Retreat Poster Presentation 2nd Prize - PepsiCo 2021 Data Challenge 2nd Price - NCSU 2021 PackHacks (Info Insight) 1st Price - Bioinformatics Graduate Traineeship (2018-2020) - Outstanding Graduates Award (2014) National Scholarship for Encouragement (2013) - Zhangyuan Tungsten Scholarship (2013) Macrolink Group Class A Scholarship (2012) - The Second Prize Scholarship (2011-2013) Outstanding Student Awards (2011-2013)
VOLUNTEERING	- Agape Way, Chapel Hill, NC, 2014-2018: managed social media platforms, proofread articles for publication on WeChat, assisted in translating student letters to donors.
PUBLICATION	- **Huang C.**, Benjamin C., et al. Phylogeny-guided microbiome OTU-specific association test (POST). *Microbiome 2022 10.1: 1-15.* - **Huang C.**, Craig G., et al. Meta-analysis reveals the vaginal microbiome is a better predictor of earlier than later Preterm Birth. (Submitted) - Kang, K., **Huang, C.**, et al. CDSeqR: fast complete deconvolution for gene expression data from bulk tissues. *BMC Bioinformatics 2021, 22(1), 1-12.* (**Co-first author**) - Odera J., Xiong Z., **Huang C.**, et al. NRF2/ACSS2 axis mediates the metabolic effect of alcohol drinking on esophageal squamous cell carcinoma. *Biochemical Journal 2020, 477(16), 3075-3089.* - Fu J., Xiong Z., **Huang C.**, et al. Hyperactivity of the transcription factor Nrf2 causes metabolic reprogramming in mouse esophagus. *Journal of Biological Chemistry 2018, 294(1), 327-340.*