Zhiyun Gong

+1-412-726-0427 | zhiyun.gong.gr@dartmouth.edu | dreamycloud.xyz

Zhiyun Gong | Zhiyun | Zhiyun_G

Education

- **Dartmouth College** *Aug 2021 - present*
  PhD. in Quantitative Biomedical Sciences Hanover, NH, USA
- **Carnegie Mellon University** *May 2021*
  MS in Computational Biology Pittsburgh,PA, USA
- **Xi'an Jiaotong liverpool University / University of Liverpool, UK** *Jul 2019*
  BSc in Bioinformatics Suzhou, China

Experience

- **Biomedical Data Science Department, Dartmouth College** *Jun 2022 - present*
  PhD candidate/Research assistant, Supervised by Michael Whitfield Lebanon, NH, USA
  - Developed data processing and analysis pipelines for microarray data, bulk and single cell RNA-seq data using Bpipe, Snakemake, and Bash
  - Identified intrinsic molecular subtypes of Systemic Sclerosis (SSc) using semi-supervised learning on transcriptomics data
  - Developed a supervised learning method for predicting SSc intrinsic subtypes
  - Developed a R shiny-based web application for SSc cell atlas, where multiple scRNA-seq datasets were integrated and processed using a unified pipeline in Snakemake
- **Biomedical Data Science Department, Dartmouth College** *Mar 2022 - Jul 2022*
  Research Assistant Lebanon, NH
  - Recovered facial sketched using 2D facial landmark coordinates extracted from interview videos
  - Developed a CNN-LSTM model using Keras to predict the presence of each PHQ-8 item
  - Achieved good classification performance on 5 items
- **Computational Biology Department, CMU** *May 2020 - May 2021*
  Research Assistant, Supervised by: Christopher Langmead Pittsburgh,PA, USA
  - Integrated synchronous and asynchronous parallel Bayesian Optimization algorithms with different acquisition functions into a user-friendly Shiny interface that allows researchers to get optimization suggestions for their experimental parameters without writing code
  - Tested different modes of the optimization algorithm on suitable simulations and possibly real-world experiments
- **Department of Biological Sciences / Computational Biology Department, CMU,** *Feb 2020 - May 2020*
  Research Assistant, Supervised by: Jonathan, Minden and Christopher Langmead Pittsburgh,PA, USA
  - Perfored data engineering on original isoelectric focusing (IEF) time series data generated by the instrument
  - Developed different machine learning models including Convolutional Neural Network and Recurrent Neural Network to the data in order to classify the experiment results as good or bad, which will help the researcher decide whether to proceed with the 2nd dimension of separation of the proteins in 2D difference in-gel electrophoresis
  - Developed an R Shiny app where the users can get predictions of the quality by the trained models for new IEF data
- **Department of Biological Sciences, Xi'an Jiaotong Liverpool University** *Sep 2018 – May 2019*
  Final Year Project Student, Supervised by: Siew Who Choo Suzhou, China
  - Performed quality control of raw RNA-Seq data and mapped reads to reference genomes using Hisat2.
  - Assembled cerebellum transcriptomes of pangolin, cat, human, and mouse using Cufflinks.
  - Predicted protein sequences using TransDecoder and identified homologous proteins via OrthoVenn.
  - Wrote R scripts to conduct cross-species differential expression analysis.

- Conducted Gene Ontology enrichment analysis to interpret functional roles of differentially expressed genes.

- **MRC Cancer Unit, University of Cambridge** *Jul 2018 – Sep 2018*
  Visiting Student Researcher, Supervised by: Shamith Samarajiwa Cambridge, UK
  - Built an R-based pipeline for microarray data processing, differential expression analysis, and quality control.
  - Performed motif finding and filtering to identify interferon-stimulated genes (ISGs).
  - Developed an interactive web interface (JavaScript, D3.js, Bootstrap, jQuery) linked to a Neo4J graph database of ISGs and pathways for real-time gene list querying and visualization.

PROJECTS

- **SCREENER: Visualizing Geographic Patterns of Systemic Sclerosis and Environmental Exposure** *2023*
  Tools: R Shiny, Leaflet, Geospatial Data Integration
  - Built an interactive web application allowing users to overlay Systemic Sclerosis (SSc) incidence rates with superfund site locations across U.S. zip codes.
  - Integrated Medicare-derived incidence data, environmental site data, and geographic shapefiles using ‘tigris‘ and custom preprocessing scripts.
  - Enabled user-driven map navigation, state-level zoom, incidence range filtering, and disease selection (SSc vs other connective tissue diseases).
  - Included Moran’s I clustering to identify statistically significant hotspots of disease occurance
- **Glioblastoma Subtype Discovery with Multi-Omics Data** *2020*
  Tools: R, Similarity Network Fusion (SNF), Consensus Clustering, Differential Expression, Enrichment Analysis
  - Integrated gene expression, methylation, and CNV data using SNF to cluster glioblastoma patients into subtypes.
  - Identified differentially expressed genes and enriched pathways associated with each subtype.
- **Survival Prediction of Glioblastoma Patients Using Radiomic Features** *Dec 2020*
  Tools: Python, Radiomics, Cox Regression, scikit-learn
  - Extracted radiomic features from multi-modal MRI images.
  - Built Cox proportional hazards models to predict survival outcomes based on radiomic biomarkers.
- **Gene Regulatory Network Inference from Single-Cell Data** *Dec 2020*
  Tools: Monocle, bnlearn, GO Enrichment, R
  - Constructed pseudo-time trajectories with Monocle and grouped cells into developmental stages.
  - Fit Bayesian networks to infer gene interactions within and across stages; performed GO analysis on stage-specific networks.
- **Database of eCLIP Assays for Sequence Similarity Search** *Mar 2020*
  Tools: Bash, Bedtools, BLAST, PureCLIP, ENCODE, STAR
  - Downloaded and indexed CLIP-Seq peak regions to build a searchable blast database of RNA-binding protein (RBP) motifs.
  - Built a bash-based pipeline for analyzing raw eCLIP data and querying known RBP profiles.
- **Gene Co-expression Visualization Web App** *Dec 2019*
  Tools: Go, HTML/CSS, D3.js
  - Developed a web interface to visualize gene co-expression clusters from uploaded expression matrices.
  - Implemented correlation computation, graph generation, and DFS traversal in Go for backend functionalities.

PRESENTATIONS

Talks

- **"Identification of Systemic Sclerosis Intrinsic Subtypes in the ASSET Clinical Trial Using PBC Gene Expression"**
  Zhiyun Gong, Timothy Sullivan, Tammara Wood, David Fox, Dinesh Khanna, Michael Whitfield
  ACR Convergence, Washington, D.C., Nov 2024

- **"RUNX1 Is Expressed in a Subpopulation of Dermal Fibroblasts and Is Increased with Systemic Sclerosis Disease Severity"**
  Rezvan Parvizi, **Zhiyun Gong**, Helen Jarnagin, Tamar Abel, Dillon Popovich, Madeline Morrisson, Tammara Wood, Sasha Shenk, Monique Hinchcliff, Jonathan Garlick, Patricia Pioli, Michael Whitfield
  ACR Convergence, Washington, D.C., Nov 2024

- **"Single Nuclei Multiome and Spatial Transcriptomic Analysis of Early, Untreated SSc Skin Identifies Signaling Interactions Between Macrophages and Fibroblasts"**
  Helen Jarnagin, Dillon Popovich, Rezvan Parvizi, Rosemary Gedert, Lam C. Tsoi, Rachael Wasikowski, **Zhiyun Gong**, Madeline Morrisson, Laurent Perreard, Fred Kolling IV, Dinesh Khanna, Johann Gudjonsson, Michael Whitfield
  ACR Convergence, Washington, D.C., Nov 2024

Posters

- **"Identification and Prediction of Systemic Sclerosis Intrinsic Subtypes Using Semi-Supervised and Supervised Learning on Gene Expression Data of Multiple Cohorts"**
  Zhiyun Gong, Rezvan Parvizi, Helen Jarnagin, Haobin Chen, Madeline Morrisson, Tammara Wood, Monique Hinchcliff, Michael Whitfield
  ACR Convergence, Washington, D.C., Nov 2024

- **"Deconvolution Reveals Differences in Cell Type Composition Among Systemic Sclerosis Intrinsic Subtypes Defined by Semi-supervised Learning in Multiple Cohorts"**
  Zhiyun Gong, Rezvan Parvizi, Helen C. Jarnagin, Haobin Chen, Madeline J. Morrisson, Tammara A. Wood, Michael L. Whitfield
  8th Systemic Sclerosis World Congress, Prague, Czech Republic, Mar 2024

- **"Identifying a Signature of Adipocyte to Fibroblast Transition in Systemic Sclerosis"**
  Helen Jarnagin, Rezvan Parvizi, Madeline Morrisson, **Zhiyun Gong**, Patricia A. Pioli, Michael L. Whitfield
  8th Systemic Sclerosis World Congress, Prague, Czech Republic, Mar 2024

- **"RUNX1 Expression and Binding Site Accessibility Is Associated with Increased Disease Severity in Systemic Sclerosis"**
  Rezvan Parvizi, **Zhiyun Gong**, Helen C. Jarnagin, Diana M. Toledo, Tamar R. Abel, Dillon Popovich, Madeline J. Morrisson, Sasha Shenk, Monique E. Hinchcliff, Jonathan A. Garlick, Patricia A. Pioli, Michael L. Whitfield
  ACR Convergence, San Diego, CA, Nov 2023

SKILLS

- **Programming Languages:** R, Python, Golang, Java, Bash
- **Web Technologies:** HTML5, CSS, JavaScript, D3.js,R shiny, Bootstrap
- **Database Systems:** SQL, Spark, MongoDB, Neo4j
- **Data Science & Machine Learning:** scikit-learn, keras, Tensorflow, Seaborn, pandas, tidyverse, ggplot2
- **Cloud Technologies:** AWS, Google cloud
- **DevOps & Version Control:** Git, GitHub, Docker, Kubernetes
- **Bioinformatics:** Snakemake, Microarray, RNA-seq, scRNA-seq data analysis
- **Computational Biology:**
- **Research Skills:** Data Integration, Reproducible Research, Scientific Writing, Visualization and Communication

HONORS AND AWARDS

- **Student and Resident Research Award** *Nov 2024*
  Rheumatology Research Foundation
  - Competitive national award supporting research presentations at ACR Convergence
- **CQB Travel Award** *Mar 2024*
  Center for Quantitative Biology, Dartmouth College
  - Travel support for presenting at the Systemic Sclerosis World Congress
- **Guarini Travel Award** *Mar 2024*
  Guarini School of Graduate and Advanced Studies, Dartmouth College
  - Awarded for excellence in research presentation and conference participation
- **ACR Convergence Student and Resident Scholarship** *Nov 2023*
  Rheumatology Research Foundation
  - Scholarship to support attendance and presentation at ACR Convergence 2023
- **ACR Convergence Student and Resident Scholarship** *Nov 2022*
  Rheumatology Research Foundation
  - Scholarship to support attendance and presentation at ACR Convergence 2022
- **Research Excellence Award** *May 2021*
  Carnegie Mellon University
  - Recognized for outstanding academic and research performance during MS in Computational Biology
- **University Academic Excellence Award** *Aug 2018*
  Xi'an Jiaotong Liverpool University
  - Awarded for academic distinction in undergraduate Bioinformatics program

ADDITIONAL INFORMATION

Languages: English (Full profession proficiency), Mandarin Chinese (Native), Russian (Elementary proficiency)