Qihao Qi

555 San Antonio Rd, Mountain View, CA 94040
Phone: 202-436-0829 E-Mail: Qihao.Qi@ucsf.edu

Education

University of Chicago, Booth school of business, Chicago, USA 3/2017 – 3/2020
Master of Business Administration

George Washington University, Washington, D.C., USA 8/2010 – 12/2011
Part time Non-degree graduate student at Department of Statistics
- Coursework: Mathematical Statistics I & II, Probability (Ph.D. level)

Georgetown University, Washington, D.C., USA 1/2008 – 12/2008
Master of Science in Biochemistry and Molecular Biology

Southern Medical University, Guangzhou, China 9/2003 – 6/2007
Bachelor of Engineering in Biomedical Engineering – Medical Imaging Track

Publication

First-author, "Non-negative matrix factorization of gene expression profiles: a plug-in for BRB-ArrayTools" *Bioinformatics*

Professional Experience

The University of California, San Francisco (UCSF) **San Francisco, CA**

Bioinformatics Programmer III at Institute for Neurodegenerative Diseases 12/2016 - Present

- Analyzed and evaluated IonTorrent and Illumina RNA-Seq datasets. Recommended purchase decision.
- Led RNA-Seq statistical data analysis projects with collaborators from other institutions. Conducted differentially expression analysis, GO pathway analysis, permutation test and built gene co-expression network.
- Created a new algorithm that can produce reliable results by discovering hidden confounding variables from RNA-Seq data. Built negative binomial model to model RNA-Seq expression, and created step-wise algorithm to automatically select strong predictors. Parallelized the algorithm to significant reduce computing time.
- Analyzed several CRISPR knockout iPSC RNA-Seq datasets to discover significant genes.Used PCA analysis to detect outliers.

The EMMES Corporation - *A consulting company provides IT and statistical analysis services* **Rockville, MD**

Senior Bioinformatics Statistical Analyst 2/2015 – 12/2016

- Analyzed DNA-Seq data sequenced from cancer samples of a 3D in vitro mouse model. Improved accuracy of GATK variants calling results by creating a new mapping strategy for BWA alignment.
- Built reliable end-to-end NGS pipelines with BWA and SAMTools. Packaged the pipeline within Docker image. Deployed the pipeline on Google Cloud. Optimized the workflow to reduce costs and processing time by 2.5X.
- Identified possible oncogenes from leukemia RNA-Seq datasets with edgeR R package. Set up design matrix. Examined hidden relationship among samples. Performed Gene Set Enrichment Analysis on GO, KEGG genelists.

- Spearheaded development of BRB-SeqTools, a Linux software built with C++ and Qt for analyzing DNA and RNA sequencing data. Received 5 out of 5 (outstanding) in annual performance reviews in Technology Innovation, Timeliness of Performance and Business Relations.
  - Theoretical evaluated existing methods and tools for analyzing Illumina data.
  - Significantly shortened development time with agile development approach. Completed development from concept to final product in half the required time.
  - Validated analysis results and wrote user manual.
- Mentored junior engineers. Trained new hires. Drafted business proposal. Rewarded $8M government contract.
- Presented market trends of genomic data analysis solutions to company key leaders. Uncovered an emerging market and outlined missing in-house technology capabilities. Recommended a practical solution for transferring and storing mass Next-generation sequencing data.

Bioinformatics Statistical Analyst 10/2014 – 1/2015

- Learned JavaScript and Node.js on the fly and created a prototype for RNA-Seq data analysis tool using new skills with Tophat2, SAMTools and HTSeq. Evaluated existing variant calling programs, such as GATK, freebayes and BCFTools on RNA-Seq data.
- Improved alignment performance of RNA-Seq workflow by implementing STAR 2-PASS approach. Benchmarked STAR aligner on multiple NGS datasets.

Bioinformatics Programmer 2/2009 – 9/2014

- Led a team of engineers to re-architect the computational infrastructure of BRB-ArrayTools, a Bioinformatics statistical analysis software used in 68+ countries. Eliminated 15% of codes and improved computing speed by over 30% by re-coding of the back-end computing engine.
- Created an interactive 3D visualization application using OpenGL, R and Tcl/Tk. This tool uses Principal components analysis algorithm to project the high-dimensional gene expression data into a 3D space.
- Built statistical models for predicting risk of having a type of cancer with various methods, including Logistic Regression, Lasso, Random Forest and Boosting.
- Built a C++ Rserve back-end servicer that enables desktop application fast transfer big data set to R.
- Developed a new website for BRB-ArrayTools using JavaScript and Bootstrap CSS library. Launched a message board system with phpBB.
- Troubleshot installation issues of our software via email and phone. Helped users to choose the best analysis methods for their study objects and understand results.
- Supervised activities of new customer support engineers. Achieved the highest rating in customer satisfaction.
- Instructed annual software users workshops at NIH to advertise our software products.

National Cancer Institute (NCI), National Institutes of Health **Bethesda, MD**

Guest Researcher (Summer Intern) 5/2008 – 8/2008

- Worked with world-leading scientists at Division of Cancer Treatment and Diagnosis, NCI to develop a machine learning algorithm, Non-negative Matrix Factorization (NMF) in R. This unsupervised clustering algorithm can identify hidden patterns. Improved the usability of NMF on negative data by adding the Semi-NMF algorithm.
- Published Qi Q. et al. (2009) Nonnegative matrix factorization of gene expression profile: A plug-in for BRB-ArrayTools. Bioinformatics. 25(4). 545-547