**Yujue Wang, M.S.**

Newton, MA; (347)-789-0342; [spsc83@gmail.com](mailto:spsc83@gmail.com)

**PROFILE**

* 1+ years of data lake development and multiple cloud data transfer development experience.
* 3+ years of working experience in bioinformatics, including website development, single-cell DNA whole genome sequencing (WGS) pipeline development optimization and implementation, and single nucleotide variations (SNVs) and short insertions and deletions (INDELs) from the next generation sequencing data annotating and classifying.
* 6+ years of working experience in finance (programming and project managing).

**SKILLS**

reading riding

Python, R, HTML, CSS, JavaScript, SQL, Bash, C, Delphi, Java, Perl, Flask, Django, jQuery, MySQL, SQLite, Mandarin Chinese (native speaker)

**PUBLICATIONS and WEBSITE DEVELOPMENTS**

Sun S\*, **Wang Y\***, Maslov AY, Dong X, Vijg J. SomaMutDB: a database of somatic mutations in normal human tissues. ***Nucleic Acids Res****.* 2022 Jan 7;50(D1): D1100-D1108. doi: 10.1093/nar/gkab914.

Zhao Y, **Wang Y**, Shi Y, Morrow B, *et al*. Chromatin regulators in the TBX1 network confer risk for conotruncal heart defects in 22q11.2DS. ***npj Genomic Medicine****.* 2023 Accepted.

Dong X\*, Zhang L\*, **Wang Y**\*, Lee M., Maslov AY, Wang T, Gorbunova V, Vijg J. Identifying genome structural variations in single cells. In submission.

https://vijglab.einsteinmed.org/SomaMutDB/

<https://github.com/biosinodx/SCcaller>

<https://github.com/biosinodx/SCcaller3_PEA>

<https://www.vijglab.org/>

<https://srirajlab.com/>

https://yujuewangresume.net/

**WORK EXPERIENCE**

**GeneDx, Data Science Team,** MD, USA. Apr. 2022-present **Senior Data Engineer**

As a senior data engineer, I manage the data lake of GeneDx’s exome sequencing data on Azure cloud storage. And I help genetic counselors send clinical data to customers with different cloud environments.

* Mocha is a Python-based application that serves as a robust solution for managing genomic and exomic data lakes on Azure cloud storage. It enables the archival and retrieval of data within the data lake. My role encompassed the comprehensive development of backend Python scripts, creation of unit tests, automation of document updates, integration of both unit tests and automated document updates into GitHub actions, as well as the design and execution of function tests. Presently, I am in the process of developing the front-end interface and database components of Mocha, utilizing Django, Bootstrap, jQuery, and SQLite as the foundational technologies.
* The cloud data send-out system automatically delivers clinical data to GeneDx's customers' cloud storage based on the requirements specified by GC (Genetic Counselors). It supports various cloud storage options, including AWS S3, Azure cloud storage, Google Cloud Platform, sFTP, and Citrix Sharefile. Throughout this project, I collaborated closely with GCs to gather their requirements, and subsequently, I developed the complete script and conducted unit tests.

**Albert Einstein College of Medicine, Genetics Department,** *NY, USA.*Apr. 2019-Mar. 2022 **Research Technician.**

As a research technician, I help the postdocs, professors, and labs with their projects.

* SomaMutDB is a website database that compiles all up-to-date available somatic mutation data in healthy tissues, including in-house and publicly available data. It provides multiple functions, such as data visualization, browsing, signature analysis, and data download. I designed the structure of the website system and programmed the front and back end. I implemented six mutation signature analyzing tools on the website. And I set up the MySQL database for it. The paper on this website database was published in Nucleic Acids Research. I am the co-first author.
* PEA (phasing, enhanced reference genome and assembly) method can identify genome structure variations (SVs) from single-cell WGS data. I designed and programmed the pipeline under the supervision of a postdoc. The paper on the PEA method was submitted to *Nature Methods*. I am the co-first author.
* SCcaller can identify SNVs and INDELs from single-cell sequencing data. I simplified the whole pipeline of SCcaller from 200 commands and almost 800 parameters per cell to just one command and four parameters per cell, reducing the IO by 92.8%. And I sped up the pipeline by ten times with the parallel running technic. I also fixed the bugs for the previous version in INDELs calling.
* I annotated and classified SNVs and INDELs identified from the WGS data in >1500 22q11.2 deletion syndrome patients into damaging LoF, damaging frameshift, damaging Missense, benign Missense, splice-disrupting, synonymous variants using multiple state-of-art algorithms, software, and public databases including VEP, Bystro, spliceAI, etc. under the supervision of a postdoc. I built an SQLite database for it. And I implemented multiple statistical tests based on the database, including Fisher’s exact test, binomial test, and variant-Set Test for Association using Annotation infoRmation (STAAR). The paper of the project is on the way, and I am the second author.
* I expanded the DNA mapping pipeline’s availability from only for the human genome to 7 different species for one lab. And I implemented it for 48 samples of 6 different species for the lab.
* I built and maintained the lab websites for two professors on the wowchemy and WIX platforms.
* I developed an HPC job submission website for students and postdocs at Albert Einstein College of Medicine to facilitate their computational analysis.

**Albert Einstein College of Medicine**, **Genetics Department,** *NY, USA.*Jul. 2018-Apr. 2019, **volunteer and waiting for my H1b visa approval.**

As a volunteer, I helped postdocs to implement their analysis pipelines on HPC. And I learned the necessary knowledge (single-cell WGS sequencing, DNA mapping, variant calling, etc.) for optimizing the SNV caller for single-cell WGS data.

**China UnionPay Merchant Services Company, Ltd.**, Ufood Division, *Shanghai, China*. Dec. 2016-Nov. 2017, **Project Manager**.

Ufood provides solutions for restaurant management. As a project manager, I supervised 10 employees and managed the accounts of over 1000 different vendors. We designed applications for PC, tablets, and POS terminals, a website for restaurant management, in addition to creating customized WeChat accounts for vendors and customers to facilitate reservations, ordering, and payments.

**China UnionPay Merchant Services Company, Ltd.**, MIS-POS (merchant integrated system – point of sale) Department, *Shanghai, China*. May 2012-Nov. 2016, **Programmer**.

I worked with senior personnel of international companies (Vanguard, McDonald's, Yum, NBA Play zone, Bestseller, ZARA, Decathlon, etc.) to develop and implement their integrated payment system (IPS) for the Chinese market. IPS module is deployed on the computers of cashiers or vending machines to drive hardware and communicate with bank servers securely.

**EDUCATION**

**East China University of Science and Technology**, *Shanghai, China*, **M.S.**, Control Science and Engineering, Sep. 2007 - May 2012

**Tongji University**, *Shanghai, China,* **B.S.**, Electrical Engineering and Automation, Sep. 2003 – Jun. 2007

**AWARDS AND HONORS**

The employee of the Year in 2014 and 2015, China UnionPay Merchant Services Company, Ltd.