HONGYU SHI

New York, NY

☎ 347-820-2226 ✉ hyshi@seas.upenn.edu

https://www.linkedin.com/in/hongyu-shi-a98258a3/

OBJECTIVE

Looking for Software Development Internship (2022)

SOFTWARE ENGINEER SKILLS

- **Machine Learning**
- **Front End Development**
- **Data Visualization**
- **Data Analysis**
- Database
- Java Back End Development

TECHNICAL SKILLS

Programming: **Java, Python, R**, JavaScript, C++, MySQL

EDUCATION

Gerstner Sloan Kettering Graduate School of Biomedical Sciences | Jul 2018 – Present

PhD Candidate in Biomedical Sciences (Computational Oncology) (GPA: 4.0/4.0)

University of Pennsylvania | Jan 2020 - Present

Master student (part-time) in Computer and Information Technology (GPA: 4.0/4.0)

Hong Kong University of Science and Technology | Sep 2014 – Jun 2018

B.Sc. in Biochemistry and Cell Biology, minor in Information Technology (GPA: 3.9/4.3)

- Relevant Coursework: Object-Oriented Programming and Data Structures, Database Management Systems, Data Mining

Georgia Institute of Technology | Aug 2016 – Dec 2016

Exchange Student (GPA: 4.0/4.0)

- Relevant Coursework: Algorithms

PROJECTS AND RESEARCH

Shah Lab @ MSKCC, 2021

CloneAlign 2.0: integration and visualization of single cell datasets

- Implemented CloneAlign 2.0 with Pyro and PyTorch. CloneAlign is a tool based on machine learning methods (Bayesian hierarchical modeling) to integrates single cell RNA and DNA datasets.
- Improved the prediction accuracy of CloneAlign by 20% and reduced the runtime by 42%.

 ☎ 347-820-2226 ✉ hyshi@seas.upenn.edu

- Implemented a visualization application to present prediction results with React.js and D3.js.
- GitHub: https://github.com/AlexHelloWorld/clonealign_pyro

Wang Lab @ HKUST, 2018

FLORA: Fast long-noncoding RNA analysis pipeline

- Developed FLORA, an efficient computational pipeline to find novel lncRNAs and predict their functions from large gene expression datasets. lncRNA is a special group of RNA molecules in cells that can regulate gene expression.
- Identified lncRNA LINC01614 as a potential tumor driver in gastric cancer.
- Characterized gastric tumor subtypes that have implications on cancer patients' survival
- GitHub: https://github.com/AlexHelloWorld/FLORA
- Publication: https://www.nature.com/articles/s41388-021-01743-3

Schultz Lab @ MSKCC, 2019-2020

Investigate clinical associations of genomic data in cancer patients

- Used cox regression to identify genomic features associated with overall survival in colon cancer patients.
- Characterized co-occurrent genomic alterations in sarcoma patients using MSK-IMPACT datasets.
- Quantified tumor subclonal heterogeneity and conducted correlation analysis with clinical features.

PUBLICATIONS

1. Chen, Y.*, Cheng, W. Y.*, **Shi, H**.*, Huang, S., Chen, H., Liu, D., Xu, W., et al. (2021). Classifying gastric cancer using FLORA reveals clinically relevant molecular subtypes and highlights LINC01614 as a biomarker for patient prognosis. *Oncogene*, *40*(16), 2898–2909.
2. Kemeny, N. E., Chou, J. F., Capanu, M., Chatila, W. K., **Shi, H**., Sanchez-Vega, F., Kingham, T. P., et al. (2021). A Randomized Phase II Trial of Adjuvant Hepatic Arterial Infusion and Systemic Therapy With or Without Panitumumab After Hepatic Resection of KRAS Wild-type Colorectal Cancer. *Annals of Surgery*, *274*(2), 248–254.
3. Chow, A., Schad, S., Green, M. D., Hellmann, M. D., Allaj, V., Ceglia, N., Zago, G., et al. (2021). Tim-4+ cavity-resident macrophages impair anti-tumor CD8+ T cell immunity. *Cancer Cell*, *39*(7), 973-988.e9.

ADDITIONAL CREDENTIALS

HONORS & AWARDS

- The HKUST Academic Achievement Medal
- The D.H Chen Foundation Life Science Scholarship
- HKSAR Government Scholarship Fund – Reaching Out Award