Guojie Zhong

1130 St. Nicholas Avenue, New York, NY, USA| Phone: (+01)7187305993 | Email: gz2294@cumc.columbia.edu

PUBLICATIONS

- Ren, X.#, **Zhong, G.**#, Zhang, Q., Zhang, L., Sun, Y., and Zhang, Z. (2020). Reconstruction of cell spatial organization from single-cell RNA sequencing data based on ligand-receptor mediated self-assembly. *Cell Res* 30, 763-778. (#, Contributed Equally)
- Zhang, Q., He, Y., Luo, N., Patel, S.J., Han, Y., Gao, R., Modak, M., Carotta, S., Haslinger, C., Kind, D., Peet G.W., **Zhong, G.**, Lu, S., Zhu, W., Mao, Y., Xiao, M., et al.. (2019). Landscape and Dynamics of Single Immune Cells in Hepatocellular Carcinoma. *Cell* 179, 829-845 e820.
- **Zhong, G.**, Ahimaz, P., Edwards, N.A., Hagen, J.J., Faure, C., Lu, Q., Kingma, P., Middlesworth, W., Khlevner, J., El Fiky, M., et al. (2022). Identification and validation of candidate risk genes in endocytic vesicular trafficking associated with esophageal atresia and tracheoesophageal fistulas. *HGG Adv* 3, 100107.
- **Zhong, G.**, and Shen, Y. (2022). Statistical models of the genetic etiology of congenital heart disease. Curr Opin Genet Dev 76, 101967.
- **Zhong, G.**, Choi, Y.A., and Shen, Y. (2022). Integration of gene expression data in Bayesian association analysis of rare variants. *bioRxiv*, 2022.2005.2013.491893.
- Edwards, N., **Zhong, G.**, Ahimaz, P., Kenny, A., Kingma, P., Wells, J., Shen, Y., Chung, W., and Zorn, A. (2022). Discovering the Developmental Basis of Trachea-Esophageal Birth Defects: Evidence for Endosome-opathies. The FASEB Journal 36.
- Wang, Y., Tiruthani, K., Li, S., Hu, M., **Zhong, G.**, Tang, Y., Roy, S., Zhang, L., Tan, J., Liao, C., et al. (2021). mRNA Delivery of a Bispecific Single-Domain Antibody to Polarize Tumor-Associated Macrophages and Synergize Immunotherapy against Liver Malignancies. *Adv Mater* 33, e2007603.

EDUCATION

Yuanpei College, **Peking University**, Beijing, China. *2015.09 – 2019.07*
Bachelor of Science in Integrated Science Program
Core Courses: Integrated Mathematics and Physics I, Multi-Variable Calculus and Linear Algebra, Biostatistics, Probability Theory and Statistics, Introduction to Stochastic Process, Introduction to Computation, Data Structure and Algorithm, Machine Learning, Cell Biology, Molecular Biology, Genetics, Biochemistry.
Awards: 2014-2015 academic year Outstanding Freshman Scholarship in Peking University

Department of Integrated Biology, **University of California, Berkeley**, CA, USA. *2017.08 – 2017.12*
Exchange Student in Berkeley Bioscience Study Abroad Program (BBSA)
Courses: Data Structures, Foundation of Data Science, Biophysical Neurobiology

Department of Systems Biology, **Columbia University**, New York, NY, USA. *2019.08 – present*
PhD program in Cellular, Molecular and Biomedical Sciences, Systems Biology track
Courses: Machine Learning for Functional Genomics (A+), Neural Networks Deep Learning (A), Foundation of Graphic Models (B+), Deep Sequencing (A), Microbiome & Health (P, covid-19), Computational Genomics (P, covid-19)

RESEARCH EXPERIENCES

Undergraduate Researcher, Biomedical Pioneering Innovation Center (BIOPIC), Peking University.
Advisor: Dr. Zemin Zhang

Single cell RNA-seq of infiltrated T cells in carcinoma of colon and rectum. *2017.03-2017.07*

- Successfully separated several kinds of T cells, including: T helper cells, cytotoxic T cell and regulatory T cells, from eight patients' malignant tissue, normal adjacent tissue and peripheral blood, using Fluorescence Activated Cell Sorting (FACS).
- Confirmed that regulatory T cells are significantly enriched in all patients' malignant tissue based on statistical analysis. Proposed a hypothesis regarding the cause of exhaustion of cytotoxic T cells.
- Built up single cell RNA sequencing library, using several technologies including reverse transcription, PCR preamplification, Real-time Quantitative PCR and cDNA purification.

Undergraduate Researcher, Biomedical Pioneering Innovation Center (BIOPIC), Peking University.
Advisor: Dr. Zemin Zhang, Dr. Xianwen Ren

3D single cell interaction network reconstruction based on ligand-receptor mediated self-assembly. *2018.03-2019.07*

- Developed and utilized a new algorithm, CSOmap, which can give an inference of cellular spatial organization as well as cellular interaction. Different from mapping-based algorithms which require references generated from imaging technologies such as *in situ* sequencing, this algorithm only takes

single cell transcriptome data as input, and uses an unsupervised machine learning model to predict ligand-receptor based cell spatial organization.

- Applied this new algorithm to five published cancer datasets, including liver carcinoma, lung carcinoma, carcinoma of colon and rectum, melanoma, head and neck cancer. CSOmap can successfully recapitulate spatial characteristics of tumor microenvironments for multiple cancers, prioritize molecular determinants of cellular interactions, and generate biological insights consistent with literature via *in silico* interference.

PhD Student, Department of Systems Biology, Columbia University.
Advisor: Dr. Yufeng Shen

Integration of gene expression data in Bayesian association analysis of rare variants.

2019.08-2022.02

- The statistical power to identify risk genes by rare *de novo* variants is generally low due to rarity of genotype data. Previous studies have shown that disease risk genes usually have high expression in relevant cell types, although for many diseases the identity of these cell types are largely unknown. Recent efforts in single cell atlas in human and model organisms produced large amount of gene expression data.
- We developed two new methods, xTADA and VBASS, that integrate expression data to improve power of rare variants association analysis. Optimized for bulk RNA-seq and single-cell transcriptomics data respectively, xTADA and VBASS model the association of disease risk as a function of expression profiles of relevant tissue or cell types in Bayesian frameworks. VBASS uses both analytical likelihood function and neural network approximations in joint probability calculation, and it learns the importance of cell types jointly from expression and genetics data. On simulated data, both methods show proper error rate control and better power than extTADA, the state-of-the-art Bayesian method. We applied the methods to published datasets and identified more candidate risk genes than extTADA with supports from literature or data from independent cohorts.

Predict mode of function of damaging missense variants based on protein sequence, structure, and latent representation of functions.

2022.02-present

- Accurate prediction of the functional impact of missense variants is one of the bottlenecks in discovering genetic causes of diseases and implementing genomic medicine. Conventional methods all focused on predicting whether a variant pathogenicoty. The next frontier of the field is to predict mode of molecular functions.

SKILLS

- Solid background in applications of Machine Learning, Deep Learning and Statistical Learning to biological questions.
- Solid programming experience in Python, R, Matlab, Java.
- Solid experiences in analysis of single cell functional genomics data and whole genome sequencing (WGS) data.