

SHAOHENG LIANG

SL103@rice.edu / shaohengliang@gmail.com | web: shaohengliang.com | cell: 832-542-1987

EDUCATION

Rice University, Houston, Texas

2017 – 2022

Ph.D. in Computer Science | Dissertation: Interpretable and Efficient Machine Learning in Cancer Biology | GPA: 4.0/4.0

supervised by Dr. Ken Chen, **Department of Bioinformatics and Computational Biology, MD Anderson Cancer Center**

co-advised by Dr. Luay Nakhleh, **Department of Computer Science, Rice University**

(Selected courses: Bioinformatics - Sequence Analysis, Bioinformatics - Network Analysis, Advanced Bayesian Inference, Advanced Topics in Optimization, Survival Analysis, Design and Analysis of Algorithms, Neural Networks)

Certificate in Teaching and Learning

(Courses: Principles of Effective Teaching, Research on Teaching and Learning, Practicum in Teaching, Teaching Portfolio)

Tsinghua University, Beijing, China

2013 – 2017

Bachelor of Engineering in Electronic Information Science and Technology | Research advisor: Dr. Jiansheng Chen | GPA: 89/100

(Selected courses: Media and Recognition, Pattern Recognition, Auditory-Visual Information systems, Digital image processing, Signals and Systems, Digital Signal Processing, Statistical Signal Processing, Database (SQL), Quantum and statistical physics)

RESEARCH

Graduate Researcher

2017 – 2022

Computational Cancer Genomics Lab, MD Anderson Cancer Center, Houston, TX | Advisor: Dr. Ken Chen

(with group members (^) or collaborators in MDACC¹, UHouston², WUSTL³, and BCM⁴):

- Developed a white-box positive-example-only ML model based naïve Bayes formulation, with a webapp, for predicting MHC-I antigen presentation, achieving state-of-the-art performance, comparable to neural networks | Dr. Gregory Lizée¹
- Developed a Siamese-network-based peptide embedding method to map patient peptides to HLA alleles | Dr. Gregory Lizée¹
- Classified and analyzed malignant cells in HPV+ HNSCC patient samples using HPV genes, epithelial markers, inferred copy number variation, phylogeny inference, and a novel intra-tumor heterogeneity score | Dr. Maura Gillison¹
- Led the sequence and statistical analysis for single-cell CRISPR screening | Dr. Weiyi Peng²
- Co-developed an approach to integrate and analyze scRNA and scATAC profile data | Dr. Jinzhuang Dou[^]
- Innovated a unified manifold-based feature selection method for clusters, subclusters, and trajectories to help design panels including CyTOF and MerFISH | Dr. Muharrem Muftuoglu¹, Dr. Weiyi Peng², Dr. Li Ding³
- Invented an autoencoder with sine activation function to decipher cell cycle and other periodic processes
- Modeled cell type abundances using beta-binomial distribution and created a web tool to help find the best sample size | Dr. Jason Willis¹, Dr. Eduardo Vilar¹
- Studied stratified non-parametric statistical tests and effect sizes to identify DEGs under the batch effect | Dr. Rui Chen⁴
- Analyzed trajectories of CD4/CD8 T cells from individuals vaccinated for SARS-CoV-2 | Dr. Muharrem Muftuoglu¹, Dr. Li Li¹
- Used BioBert to build directional GRN from corpus | Ramiz Iqbal[^]
- Analyzed miRNA in gene edited NK cells | Dr. May Daher¹, Dr. Katy Rezvani¹
- Developed a Mahalanobis distance-based approach to mitigate batch-effect in longitudinal data
- Used GAN to predict and interpret developmental processes of retinal cells
- Analyzed gene expression and cell-cell communication in KRAS+/- PDAC mice | Jincheng Han¹, Dr. Ronald DePinho¹
- Developed analytics for DNA-barcode-based lineage tracing | Dr. Anil Sood¹
- Quantifying transfection efficiency from scRNA-seq of engineer CAR-NK products | Dr. Katy Rezvani¹
- Performed analysis of CRISPR-engineered CAR-NK/T products for acquiring FDA approval | Dr. Katy Rezvani¹
- Analyzed copy number profile of AML scATAC-seq data to identify malignant cells | Dr. Hussein Abbas¹
- Discovered gene sets from scRNA profile of multiple myeloma patients using matrix factorization | Dr. David Symer¹

- Studied HLA binding affinity of spike protein of SARS-CoV-2 variants | Dr. Xianli Jiang[^], Dr. Muharrem Muftuoglu¹, Dr. Li Li¹
- Performed differential gene expression analysis on COVID-19 patients with different severity | Dr. Weiyi Peng²
- Helped analyze copy number profile of pancreatic cancer cells | Anirban Maitra¹
- Collaborated with five labs in analyzing breast cell data and retina cell data as an integral part of the *Human Cell Atlas* project, and other data including AML, CAR-NK, CAR-T, CRISPR, etc. | Various groups and collaborators

Undergraduate Researcher

2015 – 2016

Image Recognition and High-Speed Image Processing Lab, Tsinghua University, Beijing, China | Advisor: Dr. Jiansheng Chen

- Developed a method to detect symmetric axes by dynamic programming to achieve two orders of magnitude improvement in speed
- Reimplemented a MATLAB-based gradient descent-based facial marking method in C++/OpenCV
- Contributed to a saliency detection method when searching for a fixed aspect ratio rectangle to apply the method to more scenarios

PUBLICATIONS

(Google scholar: <https://scholar.google.com/citations?user=r3Tz1xMAAAAJ&hl=en&oi=ao>) (“*” indicates co-first authorship)

Journal publications:

1. Jinzhuang Dou*, **Shaoheng Liang***, Vakul Mohanty, Qi Miao, Yuefan Huang, Qingnan Liang, Xuesen Cheng, Sangbae Kim, Jongsu Choi, Yumei Li, Li Li, May Daher, Rafet Basar, Katayoun Rezvani, Rui Chen, Ken Chen. “Bi-order multimodal integration of single-cell data.” *Genome Biology* 23, 1 (2022): 1-25
2. Jiakai Hou*, **Shaoheng Liang***, Chunyu Xu, Yanjun Wei, Yunfei Wang, Yukun Tan, Nidhi Sahni, Daniel McGrail, Mike Davies, Song (Stephen) Yi, Yiwen Chen, Cassian Yee, Ken Chen, Weiyi Peng. “Single-cell CRISPR immune screens reveal immunological roles of tumor intrinsic factors.” *NAR Cancer* 4, no. 4 (2022): zcac038.
3. Li Li, et al. “Loss of metabolic fitness drives tumor resistance after CAR-NK cell therapy and can be overcome by cytokine engineering.” *Science Advances* (in press)
4. **Shaoheng Liang**, Jason Willis, Jinzhuang Dou, Vakul Mohanty, Yuefan Huang, Eduardo Vilar, and Ken Chen. "Sensei: how many samples to tell a change in cell type abundance?." *BMC bioinformatics* 23, no. 1 (2022): 1-22.
5. Li Li*, Muharrem Muftuoglu*, **Shaoheng Liang***, Mahesh Basyal, L. V. Jiangxing, Mehmet Emin Akdogan, Ken Chen, Michael Andreeff, Christopher R. Flowers, and Simrit Parmar. "In-depth analysis of SARS-CoV-2-specific T cells reveals diverse differentiation hierarchies in vaccinated individuals." *JCI insight* 7, no. 7 (2022).
6. Maria Monberg, Heather Geiger, Jaewon J. Lee, Roshan Sharma, Alexander Semaan, Vincent Bernard, Justin Wong et al. "Occult polyclonality of preclinical pancreatic cancer models drives in vitro evolution." *Nature communications* 13, no. 1 (2022): 1-16.
7. **Shaoheng Liang**, Vakul Mohanty, Jinzhuang Dou, Qi Miao, Yuefan Huang, Muharrem Müftüoğlu, Li Ding, Weiyi Peng, and Ken Chen. "Single-cell manifold-preserving feature selection for detecting rare cell populations." *Nature Computational Science* 1, no. 5 (2021): 374-384.
8. **Shaoheng Liang**, Qingnan Liang, Rui Chen, and Ken Chen. "Stratified Test Accurately Identifies Differentially Expressed Genes Under Batch Effects in Single-Cell Data." *IEEE/ACM Transactions on Computational Biology and Bioinformatics* (2021).
9. Fang Wang, Qihan Wang, Vakul Mohanty, **Shaoheng Liang**, Jinzhuang Dou, Jincheng Han, Darlan Conterno Minussi et al. "MEDALT: single-cell copy number lineage tracing enabling gene discovery." *Genome biology* 22, no. 1 (2021): 1-22.
10. Muharrem Muftuoglu, Li Li, **Shaoheng Liang**, Duncan Mak, Angelique J. Lin, Junxiang Fang, Jared K. Burks, Ken Chen, and Michael Andreeff. "Extended live-cell barcoding approach for multiplexed mass cytometry." *Scientific reports* 11, no. 1 (2021): 1-13.
11. Hussein A. Abbas, Vakul Mohanty, Ruiping Wang, Yuefan Huang, **Shaoheng Liang**, Feng Wang, Jianhua Zhang et al. "Decoupling Lineage-Associated Genes in Acute Myeloid Leukemia Reveals Inflammatory and Metabolic Signatures Associated with Outcomes." *Frontiers in Oncology*: 2897.
12. **Shaoheng Liang**, Fang Wang, Jincheng Han, and Ken Chen. "Latent periodic process inference from single-cell RNA-seq data." *Nature communications* 11, no. 1 (2020): 1-8.
13. Qingnan Liang, Rachayata Dharmat, Leah Owen, Akbar Shakoor, Yumei Li, Sangbae Kim, Albert Vitale et al. "Single-nuclei RNA-

seq on human retinal tissue provides improved transcriptome profiling." *Nature communications* 10, no. 1 (2019): 1-12.

14. Fang Wang, **Shaoheng Liang**, Tapsi Kumar, Nicholas Navin, and Ken Chen. "SCMarker: ab initio marker selection for single cell transcriptome profiling." *PLoS computational biology* 15, no. 10 (2019): e1007445.

Conference Proceedings:

15. **Shaoheng Liang**, Qingnan Liang, Rui Chen, and Ken Chen. "Stratified Test Alleviates Batch Effects in Single-Cell Data." *Algorithms for Computational Biology* 12099 (2020): 167.
16. **Shaoheng Liang**, Jiansheng Chen, Zhengqin Li, and Gaocheng Bai. "Linear time symmetric axis search based on palindrome detection." In *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 1799-1803. IEEE, 2016.
17. Jiansheng Chen, Gaocheng Bai, **Shaoheng Liang**, and Zhengqin Li. "Automatic image cropping: A computational complexity study." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 507-515. 2016.

Conference abstracts:

18. Muharrem Muftuoglu, Li Li, Mahesh Basyal, Shaoheng Liang, Carissa J. Ball, Ken Chen, Prasanna Kumar, Arnaud Lesegretain, Naval Daver, and Michael Andreeff. "Single-Cell Proteomics Identifies Leukemia Landscape Associated with Clinical Outcomes in R/R AML Treated with MDM2i (Milademetan) and FLT3i (Quizartinib): Putative Role of CD68 and Diversity Index." *Blood* 138 (2021): 3443.

Preprints and manuscripts under review:

19. **Shaoheng Liang**, Jinzhuang Dou, Ramiz Iqbal, and Ken Chen. "Batch-Corrected Distance Mitigates Temporal and Spatial Variability for Clustering and Visualization of Single-Cell Gene Expression Data." *bioRxiv*.
20. Jiakai Hou, Yanjun Wei, Jing Zou, Roshni Jaffery, **Shaoheng Liang**, Caishang Zheng, Ken Chen et al. "Integrated multi-omics analyses identify key anti-viral host factors and pathways controlling SARS-CoV-2 infection." *Research Square* (2022): rs-3.

TEACHING

Instructor: "Multiomics integration and feature selection", Gulf Coast Consortia (GCC) Single Cell Omics Workshop	2022
Research Mentor: For summer interns	2021
Teaching demonstrations: Practicum in Teaching (UNIV 502), 3 mini sessions	2021
Teaching Assistant: Statistical Machine Learning (COMP 540), 3 semesters	2018 – 2020
Teaching Assistant: Statistical Models and Algorithms for Data Science (COMP 340), 2 semesters	2018 – 2019

WORK EXPERIENCE

NuProbe USA, Inc. (50+ employee venture-backed biotechnology startup)	Houston, TX
Research Scientist II, Bioinformatics	2023 –
Intern Scientist, Bioinformatics	2022
➤ Highly efficient combinatorial optimization of multiplex sequencing panel in Python/C++ hybrid programming	
➤ Deployment of panel designing services on Google Cloud Platform	
➤ Algorithms for recognizing conflicting primers	

Schneider-Electric China Research & Development Center	Shanghai, China
Intern Software Engineer, Webpage design and SVN version control configuration	2014

CONSULTING ACTIVITIES

Torus Biosystems (Medford, MA), Multiplex qPCR panel design	2023
--	------

INVITED TALKS

"Manifold preserving feature selection", MD Anderson Data Science Forum Series	2022
"Manifold preserving feature selection", 2nd Annual GCC Single Cell Omics Cluster Symposium	2021
"Manifold preserving feature selection", Cancer Target Discovery and Development (CTD ²) Network Oregon Health & Science	

University (OSHU) site visit	2021
“Manifold preserving feature selection”, Texas Single-Cell (TXSC) Seminar Series	2021
“Batch effect, batch stratified test, and batch orthogonal analysis”, Texas Single-Cell (TXSC) Seminar Series	2019
“Latent variable model helps delineate cell states”, Texas Single-Cell (TXSC) Seminar Series	2018

CONFERENCE TALKS

“Stratified Test Alleviates Batch Effects in Single-Cell Data”, International Conference on Algorithms for Computational Biology (AICoB)	2021
“Manifold preserving feature selection”, CZI Seed Networks Computational Biology Meeting	2021
“Latent periodic process inference from single-cell RNA-seq data”, Intelligent Systems for Molecular Biology (ISMB)	2019
“Latent variable model helps delineate cell states”, Rice University Data Science Conference	2018

POSTER PRESENTATIONS

International Conference on Research in Computational Molecular Biology (RECOMB) Satellite: Computational Cancer Biology (CCB), San Diego, CA	2022
CZI Seed Networks Computational Biology Meeting	2021
Algorithms and Models for Single-Cell Genomics, Irvine, CA	2019
American Society of Human Genetics (ASHG) Annual meeting, Houston, TX	2019
Rice University Data Science Conference, Houston, TX	2019
International Conference on Research in Computational Molecular Biology (RECOMB) Satellite: Computational Cancer Biology (CCB), Washington, DC	2019

SERVICES

Guest editor: <i>Emerging Topics in Life Sciences on AI in the life sciences</i>	2021
Peer reviewer: BMC Bioinformatics, PLOS Computational Biology, Scientific Reports, RECOMB 2020, RECOMB-CCB2019, AICoB 2019, ISMCO 2020	2019-2022

HONORS AND AWARDS

Hongqian Electronics Scholarship for Scientific Innovation	2016
University Scholarship for Academic Excellence	2016
National Olympiad in Informatics in Provinces, first prize	2013
National Olympiad in Informatics in Provinces, first prize	2012

MEDIA COVERAGE

GenomeWeb: New Single-Cell Algorithm Finds Gene Panels Used to ID Rare Cells	2021
--	------