YIREN SHAO
West Roxbury, MA 02132
http://www.Yirenshao.com
(984) 261-5682 yiren.shao@outlook.com https://www.linkedin.com/in/yirenjshao/

EDUCATION

DUKE UNIVERSITY, School of Medicine, Durham, NC
Master of Biostatistics, May 2023, GPA: 3.85/4.0
Relevant coursework includes: Data Structure and Algorithm, Statistical Inference, High Dimensional Statistics, Computational Sequence Biology, Statistical Programming, Statistical Computing, Applied Biostatistical Method I&II, Generalized Linear Model

TONGJI UNIVERSITY, School of Mathematical Science, Shanghai, China
Bachelor of Science, **Applied Mathematics**, July 2021, GPA: 87/100. **Scholarship for Excellent Academic Performance**
Relevant coursework included: Differential Equations, Numerical Analysis, Mathematical Analysis, Complex Analysis, Computing Method, Functional Analysis, Advanced Algebra, Topology, Deep Learning, Database Management

PUBLICATIONS

- Cieri N, Hookeri N, Stromhaug K, Li L, Keating J, Díaz-Fernández P, Gómez-García de Soria V, Stevens J, Kfuri-Rubens R, **Shao Y**, Kooshesh KA, Powell K, Ji H, Hernandez GM, Abelin J, Klaeger S, Forman C, Clauser KR, Sarkizova S, Braun DA, Penter L, Kim HT, Lane WJ, Oliveira G, Kean LS, Li S, Livak KJ, Carr SA, Keskin DB, Muñoz-Calleja C, Ho VT, Ritz J, Soiffer RJ, Neuberg D, Stewart C, Getz G, Wu CJ. **Systematic identification of minor histocompatibility antigens predicts outcomes of allogeneic hematopoietic cell transplantation.** Nat Biotechnol. 2024 Aug 21. doi: 10.1038/s41587-024-02348-3. PMID: 39169264.
- **Shao Y**, Gao Q, Wang L, Nixon A, Chan C, Li Q, Xie J; **B-Lightning: Using bait genes for marker gene hunting in single-cell data with complex heterogeneity (https://github.com/yirenshao/B-Lightning).** Manuscript under revision in Briefings in Bioinformatics.

ABSTRACTS

- Cieri N, **Shao Y**, Powell K, Ho VT, Stevens J, Lane WJ, Kim HT, Neuberg D, Stewart C, Getz G, Ritz J, Soiffer RJ, Wu CJ. **Acute GvHD of the Gut is Associated with Minor Histocompatibility Antigens Cross-reactive against Gut-tropic Viral Epitopes**. Abstract accepted for oral presentation at the 2024 American Society of Hematology (ASH) Meeting.
- Cieri N, **Shao Y**, Powell K, Smith C, Ho VT, Kerrigan M, Stevens J, Lane WJ, Kim HT, Neuberg D, Stewart C, Getz G, Ritz J, Soiffer RJ, Wu CJ. **Minor Histocompatibility Antigens Cross-reactive against Gut-tropic Viral Epitopes facilitate acute GvHD of the Gut.** Submitted to Tandem Meeting of ASTCT and CIBMTR 2025.

RECENT ACADEMIC EXPERIENCE

DANA-FARBER CANCER INSTITUTE, Boston, MA 2023-2024
Computational Biologist

Project: Systematic identification of minor histocompatibility antigens predicts outcomes of allogeneic hematopoietic cell transplantation

- Applied the pipeline for minor histocompatibility antigen (mHAg) prediction on whole-exomes from a cohort of 220 HLA-matched allogeneic hematopoietic cell transplant (allo-HCT) recipients (and their relative donors)
- Extended and applied the mHAg pipeline to and additional cohort of 300 unrelated donor transplants
- Processed, integrated and analyzed single-cell RNA sequencing data obtained from colon biopsies of allo-HCT patients with graft-versus-host disease (GvHD) to evaluate the gene expression profile and its overlap with that of colonic samples obtained from healthy donors
- Contributed to **Systematic identification of minor histocompatibility antigens predicts outcomes of allogeneic hematopoietic cell transplantation**. **Nat Biotechnol**. 2024 Aug 21. doi: 10.1038/s41587-024-02348-3.
- **This work serves as a solid foundation for delineation of post allo-HSCT mHAg repertoire in a personalized fashion**

Project: Quantification of viral cross-reactivity across patient-specific minor histocompatibility antigen landscape

- Built up a comprehensive viral epitope library by using HLAthena1.0, NetMHCpan4.1 and IEDB based on viral protein sequences from UniProt
- Designed and implemented a computational algorithm incorporating two ways of measurement (6AA homogeneity and levinstein distance) to quantify homology between viral epitopes and patient-specific GvHD mHAgs inferred from WES
- **The library can be used for identification of potential cross-reactive mHAgs**
- **In silico analysis revealed that cross-reactivity against CMV contributed to the pathophysiology of GI acute GvHD**

Project: Identification of Major HLA Mismatch Derived Epitopes for prognostication in partially-matched transplants

- Built up a pipeline to identify major HLA mismatch derived epitopes based on HLAthena1.0 and NetMHCpan4.1
- Streamlined and docked the pipeline on Terra

Project: Deciphering the Molecular Basis of Cutaneous Chronic GvHD Following Blaschko Lines

- Analyzed bulk RNA-seq and WES data of affected and unaffected dermis and epidermis (AD, AE, UD, UE) biopsies from a post-alloHCT patient who developed skin chronic Graft-versus-Host-Disease presenting along Blaschko lines
- Quantified mitochondrial DNA heteroplasmy in AD, AE and UE bulk RNA-seq data by mitochondrial reads
- Analyzed 4 published bulk RNA-seq data and 1 published single-cell RNA-seq data of affected and unaffected skin tissues to finalize a gene panel for cGvHD-related minor histocompatibility antigen (mHAg) prediction based on whole-exome sequencing (WES) of germline DNA from donor and recipient (D-R)
- Assessed the presence of concomitant virome in AD, AE and UE bulk RNA-seq data via VirPy and Pathseq using VZV/HSV1 infected cell lines as positive controls
- Identified somatic mutations by implementing CGA pipeline on AD, UD, AE, UE and germline WES and predicted corresponding neoantigens using HLAthena and NetMHCpan4.1
- Identified transposable elements using SQuiRE on RNA-Seq data from AD, AE, UE, as well as publicly available unaffected skin bulk RNA-seq and predicted corresponding neoantigens via HLAthena and NetMHCpan4.1

Project: Delineating the TCR specificity of Proliferating Lymphocytes before PTCy Administration

- Genotyped and analyzed single-cell RNA-seq data of MRD, URD, MMURD and HAPLO's PBMC 3 days post-alloHCT to identify donor-derived proliferating T cells
- Analyzed single-cell TCR-seq data to define clonal expansion and clonotypes and mapped TCR clonotypes to single cell RNA-seq clusters to select top expanded clonotypes
- Implemented VDJdb to explore T-cell specificity based on the defined clonotypes

DUKE UNIVERSITY, School of Medicine, Durham, NC 2022-2024

Research Assistant

Project: B-Lightning: Using Bait Genes for Marker Gene Hunting in Single-cell Data with Complex Heterogeneity

- Designed and realized an iterative algorithm called B-Lightning as a gene marker fishing method within R and C++
- Modified SPLATTER to generate simulated single-cell RNA-seq data of complicated heterogeneity
- Refined and implemented the algorithm on the simulated and experimental single-cell RNA-seq data, including lung ciliated cells data, T cells data, breast cancer data and Alzheimer's Disease data
- **The algorithm outperforms some traditional differential expressed genes detecting methods in terms of sensitivity/specificity and has been used to identify a synchronized pattern in effective and memory T cell transformation under the context of SARS-COV-2 infections**
- **Manuscript in advanced stage of revision in Briefings in Bioinformatics (https://github.com/yirenshao/B-Lightning)**

PRESENTATIONS

- "**Interpretable Dimension Reduction for Pseudo-time Trajectories**," The 18th Annual CFAR Fall Scientific Retreat, Duke Center for AIDS Research, Duke University, Durham, NC, September 22, 2022
- "**An Iterative Gene Fishing Method to Detect Senescence Gene Markers**," CHSI Meta-lab Seminar, Duke Center for Human Systems Immunology, Duke University, Durham, NC, September 27, 2022

VOLUNTEER EXPERIENCE

EARLY BIRD REHABILITATION CENTER, Shanghai, China 2017-2018

Teaching Assistant

- Assisted teachers in tutoring kids with Autism, Hyperkinetic Disorder or Intellectual Disability