

Noor Sohail

 [nsohail19.github.io](https://github.com/nsohail19) |  github.com/nsohail19 |  Boston, MA

EDUCATION

University of Michigan

B.Sc. in Life Science Informatics

May 2020

Ann Arbor, MI

Relevant coursework: Intro Bioinformatics, Biochemistry, Organic Chemistry, Data Structures & Algorithms, Foundations of Computer Science, Web Systems, Discrete Math, Linear Algebra, Probability, Real Analysis

RESEARCH EXPERIENCE

Harvard T.H. Chan School of Public Health

Bioinformatician II, Harvard Chan Bioinformatics Core

Apr. 2023 – Present

Boston, MA

- Lead end-to-end single-cell and spatial transcriptomic analyses to uncover disease mechanisms in immunology, dermatology, and neuroscience studies, contributing to multiple publications.
- Designed an imputation-based workflow that improves gene co-expression estimates in sparse, zero-inflated single-cell data to prioritize drug targets for a pharmaceutical partner.
- Contrast conditions and identify disease-associated cell states with pseudobulk differential expression, gene set enrichment, neighborhood-based differential abundance, and cell-cell communication analysis.
- Create and teach workshops on programming basics and advanced data analysis, equipping hundreds of Harvard researchers with practical computational skills.

Memorial Sloan Kettering Cancer Center

Computational Biologist, Single-cell Analysis Innovation Lab

Apr. 2021 – Apr. 2023

New York, NY

- Served as primary data analyst across many translational projects, including single-cell RNA-seq analysis of CRISPR-engineered mouse tumors revealing how chromosome 9p deletions drive immune evasion (*Nature Cancer*, 2022).
- Assembled an 8M-cell COVID-19 atlas by integrating published datasets and summarizing them into metacells (SEACells), reducing sparsity and making the data tractable for analysis.
- Developed a parallelized pipeline that recovers weak interferon signals from raw reads by rescuing multimapped reads shared across the IFN gene family.
- Engineered a pipeline to call mitochondrial DNA variants from whole-cell ATAC-seq data to trace lineages in cancer.
- Implemented scalable WDL/AWS/Docker processing and QC workflows supporting hundreds of samples across dozens of projects.

Van Andel Research Institute

Summer Research Intern, Triche Lab

May 2019 – Aug. 2019; May 2018 – Aug. 2018

Grand Rapids, MI

- Maintained and extended MTseeker, an R/Bioconductor package for calling mitochondrial DNA variants and annotating their predicted protein-level effects using a variant database.
- Characterized germline and somatic mitochondrial DNA mutations across pediatric cancers by comparing paired tumor and normal sequencing data (*Cancer Research*, 2019).

University of Michigan

Undergraduate Researcher, Brooks Lab

Oct. 2017 – Dec. 2019

Ann Arbor, MI

- Computed protein–ligand binding free energies for β -Secretase 1 drug candidates using molecular dynamics (CHARMM) and statistical mechanics (*J. Phys. Chem. Lett.*, 2019).

PUBLICATIONS

- [1] Rai A, Iatrou A, Valenzuela Arzeta I, ..., Ho Sui S, Daskalakis NP, Mavrikaki M. "Aging-related Transcriptomic Changes with Spatial Resolution in the Human Prefrontal Cortex." *bioRxiv* (preprint), 2026. ([link](#))
- [2] Crisler WJ, **Sohail N**, Steuart SJ, Vazquez-Machado M, Mahajan A, ..., Clark RA, LaChance A. "In morphea, cytotoxic resident memory T cells induce chronic, immunogenic endothelial cell injury via necroptosis." *JCI Insight*, 2026. ([link](#))
- [3] Collins DR, Olatotse MJ, Urbach JM, ..., **Sohail N**, ..., Walker BD, Tsibris A. "CD8+ T cell recall cytotoxicity during antiretroviral therapy is associated with limited HIV-1 reservoir size and activity." *medRxiv* (preprint), 2025. ([link](#))
- [4] Kaplan HS, Logeman BL, Zhang K, Yawitz TA, Santiago C, **Sohail N**, ..., Dulac C. "Sensory input, sex and function shape hypothalamic cell type development." *Nature*, 2025. ([link](#))
- [5] Persad S, Choo ZN, Dien C, **Sohail N**, ..., Pe'er D. "SEACells infer transcriptional and epigenomic cellular states from single-cell genomics data." *Nature Biotechnology*, 2023. ([link](#))

[6]

Barriga FM, Tsanov KM, Ho YJ, **Sohail N**, . . . , Lowe SW. “MACHETE identifies interferon-encompassing chromosome 9p21.3 deletions as mediators of immune evasion and metastasis.” *Nature Cancer*, 2022. [\(link\)](#)

[7]

Vilseck JZ, **Sohail N**, . . . , Brooks CL III. “Overcoming Challenging Substituent Perturbations with Multisite λ -Dynamics: A Case Study Targeting β -Secretase 1.” *Journal of Physical Chemistry Letters*, 2019. [\(link\)](#)

[8]

Triska P, Kaneva K, Merkurjev D, **Sohail N**, . . . , Gai X. “Landscape of Germline and Somatic Mitochondrial DNA Mutations in Pediatric Malignancies.” *Cancer Research*, 2019. [\(link\)](#)

SELECTED PRESENTATIONS

- **Sohail N**, Billingsley JM, Geist F, Hoong H, Ho Sui S. “Co-Expression Workflow for Single-Cell RNA Sequencing Data.” Poster presentation at ISMB, 2024.
- **Sohail N**, Moorman A, Pe’er D. “SEACells: Using Single Cell Aggregation to Facilitate Data Integration.” Oral presentation at Chan Zuckerberg Initiative Assembling Tissue References Workshop, 2022.
- **Sohail N**, Triche TJ. “MTseeker: Mitochondrial Variant Analysis Tools for Bioconductor.” Poster presentation at Bioconductor, 2018.

TEACHING & MENTORSHIP

Harvard Chan Bioinformatics Core | *Workshop Instructor*

Apr. 2023 – Present | Boston, MA

- Revised the scRNA-seq workshop, wrote three new workshops, and migrated the training website to Quarto to streamline material development:
 - **Introduction to Spatial Transcriptomics** [\(link\)](#)
Analysis of spatial transcriptomics data (Visium HD).
 - **Introduction to single-cell RNA Sequencing (scRNA-seq)** [\(link\)](#)
End-to-end single-cell analysis taught in both R and Python, from experimental design to cell type annotation.
 - **Pseudobulk and Related Approaches for scRNA-seq** [\(link\)](#)
Sample-level statistical approaches for comparing conditions in single-cell data.
 - **Introduction to Python** [\(link\)](#)
Introductory Python for data analysis with pandas, Matplotlib, and Seaborn.

Van Andel Research Institute | *Undergraduate Mentor*

Jun. 2019 – Aug. 2019 | Grand Rapids, MI

- Supervised two undergraduate students in computational research.

Office of Academic Multicultural Initiatives | *Mathematics Tutor*

Jan. 2018 – Apr. 2018 | Ann Arbor, MI

- Tutored first-generation college students one-on-one in Calculus I and II.

SKILLS

Genomics	Xenium, Visium / Visium HD, STEREO-seq, single-cell RNA-seq, scATAC-seq, bulk RNA-seq
Programming	Python, R, bash, SQL, C++
Visualization & Reporting	ggplot2, Matplotlib, Seaborn, Quarto, Jupyter
Infrastructure	Linux, HPC (Slurm, LSF), AWS, Docker, WDL, git