# Noor Sohail

[nsohail19.github.io](nsohail19.github.io)

## Education

**University of Michigan**                                                     Ann Arbor, MI
*B.Sc Informatics*                                                                   May 2020

## Work Experience

**Bioinformatician II**                                                April 2023 - Present
*Harvard University*                                                           Boston, MA

- Conducted exploratory analyses on genomic datasets for Harvard and pharmaceutical collaborators, identifying trends across age, sex, disease states, and experimental conditions.
- Devised, documented, and implemented a robust workflow to handle zero-inflated data, improving the accuracy of correlations for these sparse datasets.
- Generated custom, client-facing reports with R Markdown, translating results into actionable insights through clear visualizations.
- Applied a range of statistical and ML techniques to high-dimensional data, utilizing tools such as PCA, Louvain clustering, UMAP, k-nearest neighbors, and imputation.
- Developed and led workshops for first-time programmers, covering topics from introductory R and shell to advanced analyses of genomic datasets.

**Computational Biologist**                                       April 2021 - April 2023
*Memorial Sloan Kettering Cancer Center*                          New York, NY

- Published peer-reviewed research articles on the impact of targeted CRISPR deletions as well as performance evaluations of new bioinformatics algorithms.
- Improved data processing pipelines by implementing data recovery strategies while also accelerating run time by parallelizing key steps.
- Built Jupyter Notebook templates to automate routine data clean-up and quality control, reducing turnaround time for new studies.
- Maintained and improved pipelines for processing hundreds of genomic datasets using SQL, WDL, Docker, and AWS.

## Technical Skills

- Python (pandas, numpy, tensorflow), R (tidyverse), C++, SQL, bash scripting.
- LINUX/UNIX, git, LaTeX, AWS S3, HPC, slurm, LSF.
- Jupyter notebooks, R markdown.

## Publications

1. Kaplan HS, BL Logeman, K Zhang, TA Yawitz, C Santiago, **Sohail N**, et al. "Sensory input, sex and function shape hypothalamic cell type development." *Nature*, 2025.
2. Persad S, Choo ZN, Dien C, **Sohail N**, et al. "SEACells infer transcriptional and epigenomic cellular states from single-cell genomics data." *Nature Biotechnology*, 2023.
3. Barriga FM, Tsanov KM, Yu-Jui H, **Sohail N**, et al. "MACHETE identifies interferon-encompassing chromosome 9p21.3 deletions as mediators of immune evasion and metastasis." *Nature Cancer*, 2022.
4. Vilseck JZ, **Sohail N**, et al. "Overcoming Challenging Substituent Perturbations with Multisite λ-Dynamics: A Case Study Targeting β-Secretase 1." *Journal of Physical Chemistry Letters*, 2019.
5. Triska P, Kaneva K, Merkurjev D, **Sohail N**, et al. "Landscape of Germline and Somatic Mitochondrial DNA Mutations in Pediatric Malignancies." *Cancer Research*, 2019.

## Presentations

- **Sohail N**, Billingsley JM, Geist, F, Hoong H, Ho SH. "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 *CZI Assembling Tissue References Workshop* (2022).
- **Sohail N**, Triche TJ. "Deconvolution with Mitochondrial Variation". Oral presentation at *Van Andel Institute Summer Internship Research Symposium* (2019).
- **Sohail N**, Triche TJ. "MTseeker: Mitochondrial Variant Analysis". Oral presentation at *Van Andel Institute Summer Internship Research Symposium* (2018).
- **Sohail N**, Triche TJ. "MTseeker: Mitochondrial Variant Analysis Tools for Bioconductor". Poster presentation at *Bioconductor* (2018).
- **Sohail N**, Vilseck JZ, Hayes RL, Brooks CL III. "Validating Lambda Dynamics for Computing Free Energy Differences". Poster presentation at *University of Michigan Undergraduate Research Opportunity Program Symposium* (2018).

## Research Experience

| | |
|---|---|
| **Summer Research intern** | Jan 2020 - Mar 2020 |
| *Weill Cornell Medicine* | Ann Arbor, MI |

Enhanced lab-generated results by scraping publicly available data to increase analytical depth.

| | |
|---|---|
| **Summer Research Intern** | May 2018 - Aug 2018 |
| *Van Andel Research Institute* | May 2019 - Aug 2019 |
| | Grand Rapids, MI |

Contributed new features to MTseeker, an R package, to improve functionality.

| | |
|---|---|
| **Undergraduate Researcher** | Oct 2017 - Dec 2019 |
| *University of Michigan* | Ann Arbor, MI |

Simulated molecules interacting with one another using statistical mechanics for drug discovery.

## Other Activities

**Lab Prep Assistant**
*University of Michigan*
Prepared reagents and materials for an advanced biology lab course.

Jan 2020 - Mar 2020
Ann Arbor, MI

**Mathematic Tutor**
*Office of Academic Multicultural Initiatives*
Mentored students one-on-one in Calculus 1 and 2.

Jan 2018 - Apr 2018
Ann Arbor, MI

**Writer**
*Wolverine CuiZine*
Composed food related articles related to the theme of the magazine.

Sep 2017 - Mar 2020
Ann Arbor, MI

**Committee Director**
*Model United Nations at the University of Michigan*
Created educational material on historic and global issues to facilitate debate among student committees.

Sep 2016 - Mar 2020
Ann Arbor, MI