

Xingyi (Daniel) Chen

xchen274@jh.edu • danielchen05.github.io • github.com/danielchen05 • linkedin.com/in/danielchen-jhu

RESEARCH INTERESTS

Statistical genomics and biomedical data science; development and evaluation of statistical and machine-learning methodology for high-dimensional genomic and biomedical data.

EDUCATION

Johns Hopkins University

Baltimore, MD

B.S. in Applied Mathematics & Statistics, minors in Computational Medicine and Mathematics

Expected May 2027

- GPA: 3.99/4.00 | Dean's List, 6 semesters
- Relevant Coursework:

Mathematics & Statistics: Real Analysis I/II (with Measure Theory), Probability, Mathematical Statistics, Applied Statistics & Data Analysis, Monte Carlo Methods, Optimization, Computational Mathematics (*in progress*), Multivariable Calculus, Linear Algebra, Differential Equations, Causal Inference (*audit*)

Computational Biology & Genomics: Computational Genomics: Data Analysis, Computational Biology & Bioinformatics, Algorithms for Genomic Sequence Data (*in progress*), Genomic Data Visualization, Computational Medicine, Systems Pharmacology, Translating Healthcare AI (*in progress*)

RESEARCH EXPERIENCE

Research Assistant – Hicks Lab

January 2025 – Present

Johns Hopkins University, Department of Biostatistics | PI: Stephanie Hicks, PhD

- Developed a B-spline-based feature selection method for identifying isoforms with nonlinear age-dependent trajectories in bulk RNA-seq, including algorithm design, model fitting, feature ranking, and cross-dataset validation.
- Applied the framework to human hippocampal aging to identify age-associated isoform patterns; presented preliminary findings at Biology of Genomes, Cold Spring Harbor Laboratory, 2026, with a methods-focused manuscript in preparation.
- Developed and released SpotSweeper-py, a Python/scverse implementation of neighborhood-based quality control for spatial omics; validated concordance with the R/Bioconductor implementation, benchmarked computational performance and parameter sensitivity, and led the first-author F1000Research publication.
- Built reusable Visium HD spatial transcriptomics analysis infrastructure, including 10+ reproducible pipelines and the pip-installable *visiumhd-utils* package for preprocessing, QC, destriping, visualization, and figure generation, supporting an NIH R01 grant proposal.

Quantitative Sciences Undergraduate Research Experience (QSURE) Intern – Shah/McPherson Lab

New York, NY

Memorial Sloan Kettering Cancer Center | PI: Sohrab Shah, PhD; Mentors: Andrew McPherson, Matt Myers

Summer 2026

- Developed a deletion-specific one-sided scan statistic for low-coverage scDNA-seq, extending a multi-sample scan framework to test copy number 0 against the best-fitting nonzero state and aggregate evidence across cells.
- Designed simulation studies benchmarking CBS, PELT, FPOP, dynamic programming segmentation, and scan-statistic approaches across sequencing depth, bin size, event length, aneuploidy, and subclonal prevalence; identified failure modes of general-purpose methods and characterized the proposed method's resolution limits.
- Calibrated the statistic using empirical null distributions and conducted robustness analyses, quantifying recovery of rare subclonal deletions under controlled settings and sensitivity to bin alignment, carrier prevalence, and noise estimation.
- Implemented an end-to-end Python workflow for preprocessing, interval scanning, posterior carrier scoring, parallel computation, and candidate visualization; applied the method to MSK-SPECTRUM ovarian cancer scDNA-seq data to investigate candidate deletions and technical artifacts.

Independent Research Project – Donor-Aware scRNA-seq Evaluation

March 2026 – Present

Johns Hopkins University, Department of Biomedical Engineering | PI: Alexis Battle, PhD

- Led a donor-aware, multi-cohort benchmark of scRNA-seq cell-type annotation to quantify when and by how much random cell-level evaluation overestimates cross-donor generalization across atlas-scale and independent validation datasets.

- Compared random cell-level, donor-held-out, and cross-site transfer evaluation across HVG, PCA, Harmony, and scVI representations using macro F1, class-specific error analyses, and generalized linear mixed models to separate evaluation-scheme effects from donor- and batch-level heterogeneity.
- Characterized how task separability and cohort composition modulate the magnitude of performance inflation, including settings in which ceiling or floor performance can mask differences between random and donor-aware evaluation.
- Built a modular, reproducible Python benchmarking framework for multi-cohort analysis and cross-dataset validation; preparing a first-author preprint with reproducible code and outputs.

Research Assistant – Beer Lab

November 2023 – January 2025

Johns Hopkins University, Department of Biomedical Engineering | PI: Michael Beer, PhD

- Automated extraction and preprocessing of gene expression datasets from public databases and performed exploratory analyses and visualizations in R using tidyverse and ggplot2.

Bioinformatics Intern – Wei Lab

Shanghai, China

Lin Gang Laboratory | PI: Wu Wei, PhD

Summer 2024

- Reviewed and presented methodological summaries of five spatial-omics studies, highlighting statistics/ML approaches and potential follow-up analyses; supported two ongoing bioinformatics projects with data cleaning and exploratory analyses.

PUBLICATIONS & MANUSCRIPTS

[Spotsweeper-py](#) – spatially-aware quality control metrics for spatial omics data in the Python ecosystem.

Chen X, Totty M, Hicks SC. *F1000Research*. 2026; 15:33. DOI: [10.12688/f1000research.175162.1](https://doi.org/10.12688/f1000research.175162.1)

Chen X, Erdogdu B, Perteu M, Hicks SC. *Spline-based feature selection for identifying time-dependent isoform trajectories in transcriptomic data*. Manuscript in preparation.

Chen X, Li J, Battle A. *How much do random cell-level splits overestimate scRNA-seq cell-type annotation performance? A donor-aware multi-cohort benchmark across representations and transfer settings*. Manuscript in preparation.

PRESENTATIONS

Chen X, Myers M, McPherson A. *Identifying homozygous deletions in single-cell DNA sequencing data*. Oral presentation at *QSURE Symposium*, Department of Epidemiology & Biostatistics, Memorial Sloan Kettering Cancer Center, August 2026.

Chen X, Erdogdu B, Perteu M, Hicks SC. *Machine learning reveals tissue-agnostic and region-specific isoform aging markers in the human hippocampus*. Poster presentation at *Biology of Genomes*, Cold Spring Harbor Laboratory, May 2026.

TEACHING & LEADERSHIP EXPERIENCE

Undergraduate Lead Teaching Assistant

August 2024 – Present

Johns Hopkins University, Department of Mathematics | Supported Course: Differential Equations & Applications

- Promoted from Course Assistant to Lead TA; led weekly discussion section for 25 students and delivered 13 independently prepared instructional sessions per semester.
- Mentored undergraduate TAs, coordinated course logistics, co-developed quizzes/review sessions, and earned an average student teaching-evaluation score of 98%.

Co-Instructor & Peer Leader

August 2026 – Present

Johns Hopkins University | Supported Course: Intro to Hopkins: Arrive & Thrive

- Co-instruct a first-year seminar introducing incoming students to academic, professional, and student-support resources at Johns Hopkins.
- Mentor first-year students through their transition to university and facilitate discussions on academic success, community, and campus engagement.

SELECTED COURSE-BASED RESEARCH PROJECTS

BRCA-Mutant Breast Cancer scRNA-seq Analysis – Semester Project, Foundations of Computational Biology

Johns Hopkins University | Mentor: Chris Bradburne, PhD

January 2025 – May 2025

- Led differential-expression analysis for a BRCA-mutant breast cancer scRNA-seq project, comparing expression across sample clades using Welch's t-test and Mann-Whitney U-test and generating publication-ready visualizations in Scanpy.
- Prioritized six candidate genes with differential expression patterns associated with BRCA-mutant tumor evolution, assessing robustness using fold change and consensus voting across statistical tests.

Visium Spatial Transcriptomics Dataset Comparison Analysis – Final Project, Genomic Data Visualization

Johns Hopkins University | Mentor: Jean Fan, PhD

January 2025 – March 2025

- Analyzed Visium spatial transcriptomics data from two kidney tissue sections following acute injury, using deconvolution, PCA, t-SNE, and k-means clustering to compare spatial patterns at 4 hours vs 2 days post injury.
- Interpreted five marker genes with distinct spatial expression patterns in context of kidney injury and recovery.

COVID-19 Regression Analysis – Semester Project, Freshman Experience in Applied Mathematics & Statistics

Johns Hopkins University | Mentor: Prajakta Bedekar, PhD

September 2023 – December 2023

- Modeled post-vaccination COVID-19 antibody levels using multivariable regression and compared age-stratified effects.

SKILLS

Programming & Computing: Python, R, Bash; Unix, HPC/SLURM, Git/GitHub, LaTeX

Statistics & Machine Learning: B-spline modeling and feature selection, scan statistics and empirical-null calibration, generalized linear mixed models, regression/classification, simulation studies, Monte Carlo methods, cross-validation/resampling, clustering and dimensionality reduction

Genomics & Scientific Computing: bulk RNA-seq/isoform analysis, scRNA-seq, scDNA-seq, spatial transcriptomics; Scanpy, AnnData, scikit-learn, pandas, NumPy, SciPy, Bioconductor, tidyverse, ggplot2

Reproducibility & Software: Jupyter/Quarto, Snakemake, parallel computing, Python package development, PyPI, Zenodo

Languages: Mandarin (native), English (fluent), French (elementary)