Xingyi (Daniel) Chen

765-775-7928 • xchen274@jh.edu • Baltimore, MD • www.linkedin.com/in/danielchen-jhu

RESEARCH INTEREST

Statistical genomics and biomedical data science; applied statistics and machine learning for medicine and healthcare.

EDUCATION

Johns Hopkins University Baltimore, MD
Bachelor of Science Expected May 2027

- Major: Applied Mathematics & Statistics. Minors: Computational Medicine, Mathematics.
- Cumulative GPA: 3.99. Dean's List Recipient (6 semesters).
- Relevant Coursework:
  Biology: Genetics, Computational Biology (A+), Computational Medicine (A+), Computational Genomics: Data Analysis, Genomic Data Visualization, Systems Pharmacology, Healthcare AI Evaluation (current), Genomics Sequence Data Analysis (current)
  Mathematics: Multivariable Calculus, Differential Equations, Linear Algebra, Real Analysis I/II (with Measure Theory), Calculus-Based Probability & Statistics (A+), Applied Statistics & Data Analysis (A+), Data Science (A+), Optimization, Monte Carlo Methods (A+), Computational Mathematics (current), Causal Inference (audit)

RESEARCH EXPERIENCE

Research Assistant – Hicks Lab January 2025 – Present
Johns Hopkins University, Department of Biostatistics | PI: Dr. Stephanie Hicks

- Designed isoform-level machine learning models for human brain aging using bulk RNA-seq, identifying tissue-agnostic and hippocampus-specific transcript usage markers through feature selection and cross-dataset validation; presented findings at Biology of Genomes, CSHL, 2026.
- Developed 10+ reusable Jupyter pipelines and a pip-installable *visiumhd-utils* package for Visium HD preprocessing, QC, destriping, visualization, and publication-ready figure generation for a NIH R01 grant proposal.
- Ported SpotSweeper from R/Bioconductor to Python, preserving spatially aware quality-control functionality and contributing to a published F1000Research software workflow.
- Benchmarked GeneCover against scRNA-seq gene panel selection methods for Xenium panel design, assessing whether compact gene sets preserve biological coverage, marker diversity, cell-type separability, and downstream classification performance.

QSURE Intern – Shah Lab New York, NY
Memorial Sloan Kettering Cancer Center | PI: Dr. Sohrab Shah; Mentors: Dr. Andrew McPherson, Matt Myers Summer 2026

- Developed computational methods to detect homozygous deletions from scDNA data.

Research Assistant – Beer Lab November 2023 – January 2025
Johns Hopkins University, Department of Biomedical Engineering | PI: Dr. Michael Beer

- Performed exploratory analysis & visualization for gene expression datasets (R: tidyverse, ggplot2)
- Programmed dataset extraction from databases and assisted with downstream data cleaning and processing.

Bioinformatics Intern – Wei Lab Shanghai, China
Lin Gang Laboratory | PI: Dr. Wu Wei Summer 2024

- Synthesized 5 spatial-omics papers into internal talks; highlighted ML/statistics applications & proposed follow-ups.
- Supported 2 in-progress bioinformatics projects with data cleaning and preliminary analysis.

MANUSCRIPTS & PRESENTATIONS

Spotsweeper-py – spatially-aware quality control metrics for spatial omics data in the Python ecosystem.
Chen X, Totty M, and Hicks SC. *F1000Research*. 2026; 15:33. DOI: 10.12688/f1000research.175162.1

Chen X, Erdogdu B, Pertea 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.

SELECTED COURSE RESEARCH PROJECTS

Random Cell-Level Splits Introduce Systematic Bias in scRNA-seq Cell Type Annotation
– Final Project, Computational Genomics: Data Analysis
Johns Hopkins University | Mentor: Dr. Alexis Battle March 2026 – April 2026

- Benchmarked scRNA-seq cell type annotation under random cell-level and donor-held-out evaluation schemes using healthy PBMC data, demonstrating systematic performance inflation from donor-level leakage.
- Trained and evaluated multinomial logistic regression classifiers across HVG, PCA, Harmony, and scVI representations, using macro F1, accuracy, per-class F1, confusion matrices, and cross-site transfer analyses.
- Performed donor ablation and cell-type-specific error analyses, showing that classification difficulty is driven by biological structure and that T-cell subtypes remain challenging across representations.
- Built a fully reproducible research package with ordered Jupyter notebooks, modular Python code, manuscript, presentation, GitHub repository, and archived Zenodo snapshot.

BRCA-Mutant Breast Cancer scRNA-seq Analysis – Semester Project, Foundations of Computational Biology
Johns Hopkins University | Mentor: Dr. Chris Bradburne January 2025 – May 2025

- Led differential expression part of scRNA-seq analysis: compared gene expression across defined sample clades on custom-built hierarchical trees with dual statistical tests (t-test and U-test) and produced publication ready volcano plots in Python (Scanpy).
- Discovered 6 candidate driver genes whose up or down regulation tracks BRCA-mutant tumor evolution: verified robustness and quantified effects using fold change metric and consensus voting.
- Authored the discussion section of the manuscript, highlighting biological significance while outlining future work.

Visium Spatial Transcriptomics Dataset Comparison Analysis – Final Project, Genomic Data Visualization
Johns Hopkins University | Mentor: Dr. Jean Fan January 2025 – March 2025

- Analyzed Visium spatial transcriptomics data from two kidney tissue sections post-injury using R.
- Applied deconvolution, dimensionality reduction (t-SNE, PCA), and k-means clustering to identify cell-type specific responses at 4 hours vs 2 days after acute kidney injury.
- Identified and interpreted 5 significant marker genes, highlighting spatial patterns and their significance with respect to kidney recovery mechanisms.

COVID-19 Regression Analysis – Semester Project, Freshman Experience in Applied Mathematics & Statistics
Johns Hopkins University | Mentor: Dr. Prajakta Bedekar September 2023 – December 2023

- Conducted regression analysis for COVID-19 antibody levels after vaccination and compared age subgroup results: wrote a 20-page summary paper highlighting methodologies, key results, and future work.

LEADERSHIP EXPERIENCE

Undergraduate Lead Teaching Assistant August 2024 – Present
Johns Hopkins University, Department of Mathematics | Supported Course: Differential Equations & Applications, Calculus III

- Promoted from Course Assistant to Lead TA; led weekly discussion section for 25 students and delivered 13 lectures every semester using self-prepared notes.
- Mentored undergraduate TAs, supported course logistics, co-developed quizzes/review sessions, and earned a 98% average student rating.

SKILLS

Programming: Python (proficient), R (fluent), SQL, C/C++, Java
Libraries: Scanpy, AnnData, pandas, numpy, matplotlib; ggplot2, tidyverse; scikit-learn
Data/ML: EDA, regression/classification, clustering, reproducible workflows
Technical: Unix, Git/GitHub, LaTeX.
Languages: Mandarin (native), English (bilingual), French (elementary)