Justin Hong

500 W. 120th St.
351H Engineering Terrace
New York NY 10027

978-339-3108
justin.hong@columbia.edu
justinhong.me

Education

2022 – **Computer Science Ph.D. Candidate (Computational Biology Track)**
Columbia University, *4.09/4.00*
Advisor: Elham Azizi

2019 – 2020 **M.S. Electrical Engineering and Computer Science**
University of California, Berkeley, *4.00/4.00*
Advisors: Yun Song, Kannan Ramchandran
Thesis: A Likelihood-based Deconvolution of Bulk Gene Expression Data Using Single-cell References [T1].

2015 – 2019 **B.A. Computer Science & Molecular and Cellular Biology (emph. Immunology)**
University of California, Berkeley, *3.95/4.00*
Graduated with Highest Distinction and EECS Department Honors.

Research Experience

2022 - Graduate Student Researcher. Columbia University
Advisor: Elham Azizi. Hierarchical generative modeling for sample-level heterogeneity in single-cell RNA sequencing data [J1, W1]. Scalable causal discovery method applicable to inferring gene regulatory networks from Perturb-seq data [J2, J1]. Interpretable models for inferring cell-cell interactions from spatial transcriptomics data [P1].

Summer 2024 Machine Learning & NGS Intern. Sanofi
Advisor: Vikram Agarwal. Modeled the effect of mRNA sequences on *in vitro* transcription (IVT) efficiency. Took the project from processing raw Next Generation Sequencing (NGS) data to ultimately providing sequence designs for the next iteration of in-house lab experiments.

2021 – 2022 Research Engineer. Yosef Lab, UC Berkeley
Advisor: Nir Yosef. Developed open-source software for deep probabilistic modeling of single-cell omics data, `scvi-tools` [J6]. Contributed to novel research work concerning RNA velocity inference and meta-analysis of large-scale scRNA-seq datasets [W3, J5].

2019 – 2020 Graduate Student Researcher. Song Lab, UC Berkeley
Advisor: Yun Song. Developed a method for the deconvolution of bulk RNA-seq samples using a single-cell RNA-seq reference [J7].

2018 – 2019 Undergraduate Student Researcher. BLISS Lab, UC Berkeley
Advisor: Kannan Ramchandran. Developed a robust method for the federated learning regime in the presence of adversaries [W4].

2016 – 2017 Research Assistant. Brem Lab, UC Berkeley
Advisor: Rachel Brem. Trained in experimental wet lab protocols concerning gene transformations and knockouts in yeast.

Honors & Awards

2023	1st Place GSK.ai CausalBench Challenge [W2]
2020	Outstanding Graduate Student Instructor Award, UC Berkeley
2019	Graduation with Highest Distinction, UC Berkeley (equiv. *summa cum laude*)
2019	EECS Department Honors, UC Berkeley
2018	Jim and Donna Gray Scholarship, UC Berkeley
2017	Upsilon Pi Epsilon, UC Berkeley

Publications

Star symbol (*) denotes equal contributions as a co-first author.

Journal Articles

[J1] Pierre Boyeau*, **Justin Hong***, Adam Gayoso, Martin Kim, José L McFaline-Figueroa, Michael I Jordan, Elham Azizi, Can Ergen, and Nir Yosef. “Deep generative modeling of sample-level heterogeneity in single-cell genomics”. In: *Nature Methods (accepted in principle)* (2025).

[J2] Achille Nazaret, Joy Linyue Fan, Vincent-Philippe Lavallée, ... **Justin Hong** ..., David Blei, Dana Pe’er, and Elham Azizi. “Joint representation and visualization of derailed cell states with Decipher”. In: *Genome Biology* 26.1 (2025), p. 219. [URL].

[J3] Fadi Sheban*, Truong San Phan*, Ken Xie*, Florian Ingelfinger*, ... **Justin Hong** ..., Nir Yosef, Assaf Weiner, and Ido Amit. “ZEB2 is a master switch controlling the tumor-associated macrophage program”. In: *Cancer Cell* (2025). [URL].

[J4] Jiacheng Gu, ... **Justin Hong** ..., and Jellert T Gaublomme. “Mapping multimodal phenotypes to perturbations in cells and tissue with CRISPRmap”. In: *Nature Biotechnology* (2024). [URL].

[J5] Adam Gayoso, Philipp Weiler, Mohammad Lotfollahi, Dominik Klein, **Justin Hong**, Aaron Streets, Fabian J Theis, and Nir Yosef. “Deep generative modeling of transcriptional dynamics for RNA velocity analysis in single cells”. In: *Nature Methods* (2023), pp. 1–10. [URL].

[J6] Adam Gayoso*, Romain Lopez*, Galen Xing*, Pierre Boyeau, Valeh Valiollah Pour Amiri, **Justin Hong**, Katherine Wu, Michael Jayasuriya, Edouard Mehlman, Maxime Langevin, Yining Liu, Jules Samaran, Gabriel Misrachi, Achille Nazaret, Oscar Clivio, Chenling Xu, Tal Ashuach, Mariano Gabitto, Mohammad Lotfollahi, Valentine Svensson, Eduardo da Veiga Beltrame, Vitalii Kleshchevnikov, Carlos Talavera-López, Lior Pachter, Fabian J. Theis, Aaron Streets, Michael I. Jordan, Jeffrey Regier, and Nir Yosef. “A Python library for probabilistic analysis of single-cell omics data”. In: *Nature Biotechnology* 40.2 (2022), pp. 163–166. [URL].

[J7] Dan D Erdmann-Pham*, Jonathan Fischer*, **Justin Hong**, and Yun S Song. “Likelihood-based deconvolution of bulk gene expression data using single-cell references”. In: *Genome Research* 31.10 (2021), pp. 1794–1806. [URL].

ARTICLES IN CONFERENCE PROCEEDINGS

[J1] Mathieu Chevalley, ... **Justin Hong** ..., and Patrick Schwab. “The CausalBench challenge: A machine learning contest for gene network inference from single-cell perturbation data”. In: *Fourth Conference on Causal Learning and Reasoning (CLeaR 2025)* (2025). [URL].

[J2] Achille Nazaret*, **Justin Hong***, Elham Azizi, and David Blei. “Stable Differentiable Causal Discovery”. In: *Proceedings of the 41st International Conference on Machine Learning (ICML)* (2024). [URL].

REFEREED WORKSHOP PAPERS

[W1] Mathias Perez*, **Justin Hong***, Aaron Zweig, and Elham Azizi. “Domain-Invariant Feature Learning for Patient-Level Phenotype Prediction from Single-Cell Data”. In: *AI for Science workshop, NeurIPS 2025* (2025). [URL].

[W2] Achille Nazaret* and **Justin Hong***. “BetterBoost-Inference of Gene Regulatory Networks with Perturbation Data”. In: *Machine Learning for Drug Discovery (MLDD) Workshop at ICLR, Oral Presentation* (2023). [URL]. 1st place submission of GSK.ai CausalBench Challenge.

[W3] Pierre Boyeau*, **Justin Hong***, Adam Gayoso, Michael Jordan, Elham Azizi, and Nir Yosef. “Deep generative modeling for quantifying sample-level heterogeneity in single-cell omics”. In: *Machine Learning in Computational Biology (MLCB), Oral presentation* (2022). [URL].

[W4] Avishek Ghosh*, **Justin Hong***, Dong Yin, and Kannan Ramchandran. “Robust Federated Learning in a Heterogeneous Environment”. In: *ICML Workshop on Privacy and Security in ML* (2019). [URL].

PREPRINTS

[P1] **Justin Hong***, Khushi Desai*, Tu Duyen Nguyen, Achille Nazaret, Nathan Levy, Can Ergen, George Plitas, and Elham Azizi. “AMICI: Attention Mechanism Interpretation of Cell-cell Interactions”. In: *bioRxiv* (2025), pp. 2025–09. [URL].

[P2] Pierre Boyeau*, **Justin Hong***, Adam Gayoso, Martin Kim, José L. McFaline-Figueroa, Michael I. Jordan, Elham Azizi, Can Ergen, and Nir Yosef. “Deep generative modeling of sample-level heterogeneity in single-cell genomics”. In: *bioRxiv* (2024). [URL].

[P3] Ross M Giglio, Nicholas Hou, Adeya Wyatt, **Justin Hong**, ..., Nir Yosef, Elham Azizi, and José L. McFaline-Figueroa. “A heterogeneous pharmaco-transcriptomic landscape induced by targeting a single oncogenic kinase”. In: *bioRxiv* (2024). [URL].

[P4] Stephen R Quake and Tabula Sapiens Consortium. “Tabula Sapiens reveals transcription factor expression, senescence effects, and sex-specific features in cell types from 28 human organs and tissues”. In: *bioRxiv* (2024), pp. 2024–12. [URL].

THESIS WORK

[T1] **Justin Hong**, Dan D Erdmann-Pham, Jonathan Fischer, and Yun S Song. “A Likelihood-based Deconvolution of Bulk Gene Expression Data Using Single-cell References”. Master’s Thesis. University of California, Berkeley, 2021. [URL].

Presentations

INVITED TALKS

May 2025	Yosef and Amit Lab Meeting, Weizmann Institute
Feb 2025	Research Talk, MBZUAI
Feb 2025	Guest Lecture, CB803: Single Cell Biology and Bioinformatics, MBZUAI
Feb 2025	Idaghdour Lab Meeting, NYU Abu Dhabi

CONTRIBUTED TALKS

Jul 2024	Irving Institute for Cancer Dynamics Intensive Workshop
Dec 2023	Irving Institute for Cancer Dynamics Monthly Research Meeting
Nov 2022	Machine Learning in Computational Biology Workshop (MLCB), Oral Presentation
May 2022	CZI Assembling Tissue References Workshop
Dec 2021	UC Berkeley Computational Biology Skills Seminar

POSTERS

Sep 2024	Single Cell Genomics
Sep 2024	Machine Learning in Computational Biology Workshop (MLCB)
Jul 2024	International Conference on Machine Learning (ICML)
Oct 2023	Single Cell Genomics
Jul 2023	ICML Workshop on Computational Biology

Software

2025	AMICI – A method for inferring cell-cell interactions from spatial transcriptomics data.
2023	SDCD – A method for inferring causal graphs from labeled interventional data.
2022-2024	MrVI – A library for deep sample-level meta-analysis of single-cell omics data.
2021 – 2022	scvi-tools – A library for deep probabilistic analysis of single-cell omics data.
2019 – 2020	RNA-Sieve – A package for deconvolution of bulk RNA-seq data with single-cell RNA-seq references.

Industry Experience

Summer 2025 Algo Developer Intern. HUDSON RIVER TRADING

2024 – 2025 Research Consultant. VISTERRA

2020 – 2021 Software Engineer. NURO
Developed infrastructure software for the evaluation and introspection of the autonomy

software stack. Mentored new hires and a summer intern.

Summer 2019	Software Engineer Intern. Nuro Developed software for remote large-scale bot fleet management. Co-inventor of a patent relevant to this project.
Summer 2018	Software Engineer Intern. Palantir Developed code editing and execution software within the Foundry platform.
Summer 2017	Software Engineer Intern. Affinity Developed software for customer relationship management in the venture capital space.

Teaching Experience

2023	Teaching Assistant. Columbia University *Course Title*: Probabilistic Models and Machine Learning (STCS 6701) *Responsibilities*: Held office hours and recitations, developed and graded assignments.
2019–2020	Head Graduate Student Instructor. UC Berkeley *Course Title*: Probability and Random Processes (EE 126) *Responsibilities*: Developed course content, assignments, and exams. Organized staff of over ten student instructors. Lectured as a substitute. Coordinated the course transition to fully online during the COVID-19 pandemic.
2018–2019	Undergraduate Student Instructor. UC Berkeley *Course Title*: Probability and Random Processes (EE 126) *Responsibilities*: Taught discussions sections, led office hours, created course content.
2016–2017	Undergraduate Student Instructor / Course Tutor. UC Berkeley *Course Title*: Structure and Interpretation of Computer Programs (CS 61A) *Responsibilities*: Taught discussion sections, led office hours, graded exam content.

Journal and Conference Reviewing

2025	AI for Science Workshop, NeurIPS 2025
2025	Science Advances
2024	Machine Learning in Computational Biology Workshop (MLCB)
2023	Machine Learning in Computational Biology Workshop (MLCB)
2023	ICML Workshop on Computational Biology

Mentorship

Su 2025	Mathias Perez, IICD Alliance Program Summer Intern, École Polytechnique
Sp/Su 2024	Khushi Desai, Master's Student, Columbia University

Summer 2024	Léa Bohbot, IICD Alliance Program Summer Intern, École Polytechnique
Summer 2023	Tu Duyen Nguyen, IICD Alliance Program Summer Intern, École Polytechnique