CURRICULUM VITAE

Name: Smita Krishnaswamy, PhD
Email: smita.krishnaswamy@yale.edu
Website: https://www.krishnaswamylab.org/
Github: https://github.com/KrishnaswamyLab

Title: Associate Professor
Affiliations: Departments of Computer Science (Yale SEAS) and Genetics (YSM)
Program for Applied Mathematics
Program for Computational Biology and Bioinformatics
Program for Interdisciplinary Neuroscience
Wu-Tsai Institute
Yale Institute for the Foundations of Data Science
Yale Cancer Center
Yale Center for Biomedical Data Science
Yale University
Affiliate member, MILA Quebec AI Institute

Education:

2008	PhD, Computer Science and Engineering, University of Michigan, Ann Arbor, MI (Advisors: Igor L. Markov, John P. Hayes, Thesis: "Design Analysis and Test of Logic Circuits under Uncertainty," Winner EDAA Outstanding Dissertation Award)
2002	BSE, Computer Engineering, University of Michigan, Ann Arbor, MI
2002	BA, Mathematics, Kalamazoo College, Kalamazoo, MI

Career/Academic Appointments:

2020-present	Associate Professor, Yale University, New Haven, CT
2024-2027	Affiliate member, MILA Quebec Artificial Intelligence Institute (lead by Yoshua Bengio)
Fall 2024	Centre Researches Mathematiques (CRM)-Simons Professor, University of Montreal, Canada
2015-2020	Assistant Professor, Yale University, New Haven, CT
2010-2015	Postdoctoral Fellow, Columbia University. NY
2009-2010	Research Staff Member, IBM TJ Watson Research Center, Yorktown Heights, NY
2008-2008	Postdoctoral Researcher, IBM TJ Watson Research Center, Yorktown Heights, NY
2002-2008	Graduate Student Research Assistant, University of Michigan, Ann Arbor, MI

HONORS AND RECOGNITION:

International/National/Regional:

2023	Roberts innovation fund
2022	FASEB Early Career Excellence in Science Award
2022	Yale Innovation Summit, 1st Place in Biotech Pitch Contest
2022	Nucleate Bio, Golden Ticket Award
2022	Blavatnik Fellowship for Innovation
2021	Sloan Faculty Fellowship (CS)
2021	NSF CAREER Award (CISE Directorate)

2020	IEEE Machine Learning for Signal Processing, Best Student Paper Award
2019	IEEE Viz Best Workshop Submission, Honorable Mention
2019	University of Michigan, Computer Science, Alumni Merit Award
2018	Damon Runyon-Rachleff Innovation Award, Finalist
2010	IBM Research Division Accomplishment of 2010, DeltaSyn Project
2009	Outstanding Dissertation Award, Electronic Design Automation Association
2005	Design Automation and Test in Europe, Best Paper Award

University:

2022-present	Specter Fellowship in Neuroscience, Yale School of Medicine
2018-2021	Kingsley Fellowship, Yale School of Medicine
2017	Pew Fellowship Nominee for Yale

RESEARCH & SCHOLARSHIP

The primary focus of my research is on Artificial Intelligence for extracting knowledge from scientific data. Towards this end, I develop methods that build mathematical and geometric priors, as well as signal processing techniques within machine learning frameworks for exploratory analysis of big biomedical data. The problems I work on are motivated by the high-throughput, high-dimensional data that has become ubiquitous in the biomedical and health sciences as a result of breakthroughs in measurement technologies like single cell RNA-sequencing, as well as vast improvements in health record data collection and storage. While these large datasets containing millions of cellular or patient observations hold great potential for understanding the generative state space of the data, as well as drivers of differentiation, disease and progression, they also pose new challenges in terms of noise, missing data, measurement artifacts, and the so-called "curse of dimensionality." My research has been addressing this issue, by bridging the gap between computational techniques and biomedical understanding by focusing on unsupervised exploration of the structure of the state space revealed by these large datasets using a concept known as manifold learning. Manifold learning involves discovering the low-dimensional, smoothly varying intrinsic structure of heterogeneous populations of cells or patients, since they are often sampled from a coherent developmental or clinical context. In my lab, we develop two types of techniques to learn the data manifold, extract patterns from it, and represent it in various ways for biological and clinical interpretation. These include: graph topological & spectral methods, and geometric deep learning.

My research involves both fundamental contributions to machine learning such as methods for manifold learning, graph signal processing, fast algorithms for diffusion-based earth mover's distance, and graph density estimation as well as applications. I have developed several innovative and highly used methods in this space to denoise data (MAGIC), visualize data (PHATE, Multscale PHATE), to discover multi granular clusters (diffusion condensation, SAUCIE) and archetypal factors (AAnet), to understand dynamics and trajectories in data (trajectoryNET), etc. These capabilities have been applied to a wide variety of biomedical systems including data pertaining to human embryonic stem cell development, the immune system (type 1 diabetes, and infectious diseases), immunotherapy (clinical trials of anti-PD1 together with radiation), cancer (epithelial to mesenchymal transition in breast cancer, obesity-driven pancreatic cancer), structural biology (RNA folding, and protein optimization), neuronal systems (fMRI data, retinal bipolar cells in connection with macular degeneration, primary cortical neurons during development in connection with autism, connectivity data in c elegans) as well as more recently to electronic health record data (EICU, gastrointestinal bleeding, COVID-19) in order to discover differentiation trajectories, driver genes, emergent subpopulations, and predictive biomarkers of disease.

PUBLICATIONS: (includes peer-reviewed CS conference publications, *= senior, corresponding):

1. Liu, Chen, Matthew Amodio, L. Shen, Feng Gao, Arman Avesta, Sanjay Aneja, Jay C. Wang, Lucian V. Del Priore, and Smita Krishnaswamy. "Cuts: A framework for multigranular unsupervised medical image segmentation." *preprint arXiv:2209.11359* to appear in Proceedings of the 27th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) 2024.

2. Tong, Alexander, Frederik Wenkel, Dhananjay Bhaskar, Kincaid Macdonald, Jackson Grady, Michael Perlmutter, **Smita Krishnaswamy*,** and Guy Wolf*. "Learnable filters for geometric scattering modules." *IEEE Transactions on Signal Processing* (2024).

3. Venkat A, Carlino MJ, Lawton BR, Prasad ML, Amodio M, Gibson CE, Zeiss CJ, Youlten SE, **Krishnaswamy S***, Krause DS*. Single-cell analysis reveals transcriptional dynamics in healthy primary parathyroid tissue. Genome Research 2024 Jul 8. doi: 10.1101/gr.278215.123. PMID: 38977309.

4. H. Steach, S. Viswanath, Y. He, X. Zhang, N. Ivanova, M. Hirn, M. Perlmutter*, **S. Krishnaswamy*,** Inferring Metabolic States from Single Cell Transcriptomic Data via Geometric Deep Learning, *In International Conference on Research in Computational Molecular Biology, pp. 235-252. Cham: Springer Nature Switzerland, 2024.*

5. S. Leone, X. Sun, M. Perlmutter, **S.Krishnaswamy*** Bayesian Spectral Graph Denoising with Smoothness Prior, *2024 58th Annual Conference on Information Sciences and Systems (CISS)*, pp. 1-6. IEEE, 2024.

6. D. Liao, C. Liu, B. Christensen, G. Huguet, A. Tong, G. Wolf, M. Nickel, I. Adelstein, **S. Krishnaswamy.** Assessing neural Network Representations During Training using Noise-Resilient Diffusion Spectral Entropy, Assessing neural network representations during training using noise-resilient diffusion spectral entropy." In *2024 58th Annual Conference on Information Sciences and Systems (CISS)*, pp. 1-6. IEEE, 2024.

7. C. Xu, L. Goldman, V. Guo, B. Hollander-Brodie, M. Trank-Greene, I. Adelstein, E. De Brower, R. Ying, **S Krishnaswamy*,** M. Perlmutter*, Blisnet: Classifying Signals on Graphs, t*o appear in proceedings of AISTATS 2024.*

8. A. Venkat, J. Chew, F. C. Rodriguez, C. J. Tape, M. Perlmutter, & **S. Krishnaswamy*,** Directed Scattering for Knowledge Graph-Based Cellular Signaling Analysis. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 9761-9765). IEEE.

9. M. R. Zapatero, A. Tong, J.W. Opzoomer, R. O'Sullivan, F.C. Rodriguez, J. Sufi, ... & **S. Krishnaswamy*, C.J Tape,*** (2023). Trellis tree-based analysis reveals stromal regulation of patient-derived organoid drug responses. *Cell*, *186*(25), 5606-5619.

10. G. Huguet, A. Tong, E. De Brouwer, Y. Zhang, G. Wolf, I. Adelstein, & **S. Krishnaswamy*,** (2023). A Heat Diffusion Perspective on Geodesic Preserving Dimensionality Reduction. *in Proceedings of NeurIPS*

11. D. Bhaskar, S. Magruder, E. De Brouwer, A. Venkat, F. Wenkel, G. Wolf, & **S. Krishnaswamy*,** (2023). Inferring dynamic regulatory interaction graphs from time series data with perturbations. in Proceedings of Learning on Graphs (LOG).

12. G. Huguet, A. Tong, M.R. Zapatero, C.J. Tape, G. Wolf, & **S. Krishnaswamy*,** (2023, September). Geodesic Sinkhorn For Fast and Accurate Optimal Transport on Manifolds. In *2023 IEEE 33rd Machine Learning for Signal Processing (MLSP)* (pp. 1-6). IEEE.

13. K. MacDonald, K., D. Bhaskar, G. Thampakkul, N. Nguyen, J. Zhang, M. Perlmutter, M., ... & **S. Krishnaswamy*,** (2023, September). A Flow Artist for High-Dimensional Cellular Data. In *2023 IEEE 33rd Machine Learning for Signal Processing (MLSP)* (pp. 1-6). IEEE.

14. O. Fasina, G. Huguet, A. Tong et al… **S. Krishnaswamy*** "Neural FIM Neural FIM for learning Fisher information metrics from point cloud data," *in Proceedings of ICML 2023.*

15. J.L. Moore, D. Bhaskar, F. Gao, C. Matte-Martone, C., S. Du, E. Lathrop, S. Ganesan, S., ... & **S. Krishnaswamy***, **V. Greco, V.*** Cell cycle controls long-range calcium signaling in regenerating epidermis, *Journal of Cell Biology*, *222*(7), e202302095.

16. M. Kuchroo., M. DiStasio, E. Calapkulu, M. Ige, L. Zhang, S. H. Sheth, ... **S. Krishnaswamy *** & B. P. Hafler*,"Single Cell Analysis reveals inflammatory interactions driving macular degeneration,". *Nature Communications* 14.1 (2023): 2589.

17. G. Huguet, A. Tong, B. Rieck, J. Huang, M. Kuchroo, M. Hirn, G. Wolf, and **S. Krishnaswamy*.** "Time-inhomogeneous diffusion geometry and topology," *SIAM Journal on the Mathematics of Data Science (SIMODS) 5.2 (2023): 346-372.*

18. E. L. Busch, J. Huang, J., A. Benz, G. Lajoie, G. Wolf, **S. Krishnaswamy*, N. Turk-Browne*** et al., "Multi-view manifold learning of human brain-state trajectories." Nature Computational Science 3, 240–253 (2023). https://doi.org/10.1038/s43588-023-00419-0

19. Huguet, G., Magruder, D. S., Fasina, O., Tong, A., Kuchroo, M., Wolf, G., & **S. Krishnaswamy*,** (2022). "Manifold Interpolating Optimal-Transport Flows for Trajectory Inference." *in Proceedings of NeurIPS 2022.*

20. D. Bhaskar, K. MacDonald, O. Fasina, D. Thomas, B. Rieck, I. Adelstein,, & **S. Krishnaswamy*,** (2022). Diffusion Curvature for Estimating Local Curvature in High Dimensional Data. arXiv preprint arXiv:2206.03977, *in Proceedings of NeurIPS 2022.*

21. J. Huang, E.L. Busch, T. Wallenstein, M. Gerasimiuk, A. Benz, G. Lajoie, G., ... & **S. Krishnaswamy,*** (2021). Learning shared neural manifolds from multi-subject FMRI data in Proceedings of *IEEE Machine Learning for Signal Processing.*

22. D. Bhaskar, J.D. Grady, M.A. Perlmutter, **S. Krishnaswamy,*** (2021). Molecular Graph Generation via Geometric Scattering. *arXiv preprint arXiv:2110.06241*. *in Proceedings of* IEEE Machine Learning for Signal Processing.

23. Perdigoto, Ana Luisa, et al. "Immune cells and their inflammatory mediators modify β cells and cause checkpoint inhibitor–induced diabetes." JCI insight 7.17 (2022).

24. M. Amodio, S. E. Youlten, A. Venkat, B. P. San Juan , C. Chaffer, **S. Krishnaswamy***, Single-Cell Multi-Modal GAN (scMMGAN) reveals spatial patterns in single-cell data from triple negative breast cancer, *Patterns 2022.*

25. E. Castro, A.Godavarthi, J. Rubinfien, K. Givechian, D. Bhaskar, D., & **S. Krishnaswamy***, . (2022). Guided Generative Protein Design using Regularized Transformers. *Nature Machine Intelligence* (2022): 1-12.

26. D. B. Burkhardt, B. P. San Juan, J. G. Lock, **S. Krishnaswamy*, C. L. Chaffer*;** Mapping Phenotypic Plasticity upon the Cancer Cell State Landscape Using Manifold Learning. *Cancer Discov* 1 August 2022; 12 (8): 1847–1859. https://doi.org/10.1158/2159-8290.CD-21-0282

27. S. Horoi, J. Huang, B. Rieck, G. Lajoie, G. Wolf, **S. Krishnaswamy***. Exploring the Geometry and Topology of Neural Network Loss Landscapes. In International Symposium on Intelligent Data Analysis (pp. 171-184). Springer, Cham. 2022.

28. M. Kuchroo, J. Huang, P. Wong, J.C. Grenier, D. Shung, A. Tong, C. Lucas., J. Klein, Gamache I., Poujol R., Burkhardt D.B., Gigante S., Godavarthi A., Rieck B., Israelow B., Simonov M., Mao T., Eun Oh J., Silva J., Pesaranghader A., Takahashi T. Odio C.D., Casanovas-Massana A., Fournier J., Yale IMPACT Team, Farhadian S., Dela Cruz C.S., Ko A.I., Wilson F.P., Hussin J., Wolf G., A.S. Iwasaki*., S. **Krishnaswamy*.** "Multiscale PHATE Exploration of SARS-CoV-2 Data Reveals Multimodal Signatures of Disease.' Nature Biotechnology. March 2022, https://www.nature.com/articles/s41587-021-01186-x

29. A. Tong, G. Huguet, D. Shung, A. Natik, M. Kuchroo, G. Lajoie, G. Wolf, & **S. Krishnaswamy***, "Embedding Signals on Knowledge Graphs with Unbalanced Diffusion Earth Mover's Distance." IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2022, pp. 5647-5651, doi: 10.1109/ICASSP43922.2022.9746556.

30. M. Leucken, D. B. Burkhardt, R. Cannoodt, C. Lance … A. Pisco*, **S. Krishnaswamy***, F. Theis*, J. Bloom*,"A sandbox for prediction and integration of DNA, RNA, and protein data in single cells" *in* Proceedings of NeurIPS 2021.

31. A. Tong,G. Wolf, & S. Krishnaswamy, Fixing Bias in Reconstruction-based Anomaly Detection with Lipschitz Discriminators. Journal of Signal Processing Systems (2021). doi:10.1007/s11265-021-01715-6

32. D. Shung, M. Gerasimiuk, G. Wolf, S. **Krishnaswamy,***. (2021, December). MURAL: An Unsupervised Random Forest-Based Embedding for Electronic Health Record Data. In *2021 IEEE International Conference on Big Data (Big Data)* (pp. 4694-4704).

33. Tong, A., Wenkel, F., Macdonald, K., Krishnaswamy, S., & Wolf, G. (2021, October). Data-Driven Learning of Geometric Scattering Modules for GNNs. In *2021 IEEE 31st International Workshop on Machine Learning for Signal Processing (MLSP)* IEEE.

34. M. Kuchroo, A. Godavarthi, A. Tong, G. Wolf, & **S. Krishnaswamy*.** (2021, October). Multimodal Data Visualization and Denoising with Integrated Diffusion. In *2021 IEEE 31st International Workshop on Machine Learning for Signal Processing (MLSP)* (pp. 1-6). IEEE. **(Nominated for Best Paper Award)**

35. M. Amodio, D. Shung, D. B. Burkhardt, P. Wong, M. Simonov, Y. Yamamoto, D. van Dijk, F.P. Wilson, A. Iwasaki, and **S. Krishnaswamy*.** "Generating hard-to-obtain information from easy-to-obtain information: applications in drug discovery and clinical inference." Patterns (2021): 100288.

36. M. Amodio, and **S. Krishnaswamy*,** "Canvas GAN: Bootstrapped Image-Conditional Models." 2021 International Joint Conference on Neural Networks (IJCNN). IEEE, 2021.

37. M. Amodio & **S. Krishnaswamy*.** "Noise Space Optimization for GANs." *2021 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2021

38. A. Tong, G. Huguet, A. Natik, K. MacDonald, M. Kuchroo, R. R. Coifman, G. Wolf, **S. Krishnaswamy*,** Diffusion Earth Mover's Distance and Distribution Embeddings," ICML 2021.

39. D. Shung, J. Huang, E. Castro, J K. Tay, M. Simonov, L. Laine, R. Batra, **S. Krishnaswamy***, Neural network predicts need for red blood cell transfusion for patients with acute gastrointestinal bleeding admitted to the intensive care unit. Sci Rep 11, 8827 (2021). https://doi.org/10.1038/s41598-021-88226-3

40. M. Amodio, **S. Krishnaswamy***, "Multiple-manifold Generation with an Ensemble GAN and Learned Noise Prior." Proceedings Intelligent Data Analysis (IDA), 2021, pp. 24-36

41. Mark W. Moyle, Kristopher M. Barnes, Manik Kuchroo, Alex Gonopolskiy, Leighton H. Duncan, Titas Sengupta, Lin Shao, Min Guo, Anthony Santella, Ryan Christensen, Abhishek Kumar, Yicong Wu, Kevin R. Moon, Guy Wolf, **Smita Krishnaswamy**+, Zhirong Bao+, Hari Shroff+, William Mohler+, Daniel A. Colón-Ramos+; Structural and developmental principles of neuropil assembly in C. elegans, doi: https://doi.org/10.1101/2020.03.15.992222, Nature 591, no. 7848 (2021): 99-104. (+co-senior)

42. D.B. Burkhardt, J.S. Stanley, A. L. Perdigoto, S. A., Gigante, K.C. Herold, G. Wolf, A.J. Giraldez, D. van Dijk*, and **S. Krishnaswamy***, 2019. Quantifying the effect of experimental perturbations in single-cell RNA-sequencing data using graph signal processing. Nature biotechnology, 37(12), 1482-1492 (2021).

43. B. Rieck, T. Yates, C. Bock, K. Borgwardt, G. Wolf, N. Turk-Browne, **S. Krishnaswamy***. Uncovering the Topology of Time-Varying fMRI Data using Cubical Persistence. Advances in Neural Information Processing Systems. 2020;33.

44. Castro, E., Benz, A., Tong, A., Wolf, G., & Krishnaswamy, S. (2020, December). Uncovering the Folding Landscape of RNA Secondary Structure Using Deep Graph Embeddings. In *2020 IEEE International Conference on Big Data (Big Data)* (pp. 4519-4528). IEEE.

45. A. Tong, G. Wolf, **S. Krishnaswamy***, "Fixing Bias in Reconstruction-based Anomaly Detection with Lipschitz Discriminators," 2020 IEEE 30th International Workshop on Machine Learning for Signal Processing (MLSP). IEEE, 2020. (Winner Best Student Paper Award)

46. M. Amodio, D. van Dijk, G. Wolf, **S. Krishnaswamy***, "Learning General Transformations of Data for Out-of-Sample Extensions," 020 IEEE 30th Machine Learning for Signal Processing (MLSP). IEEE, 2020.

47. A. Tong, D. van Dijk, J. S. Stanley III, M. Amodio, K. Yim, R. Muhle, J. Noonan, G. Wolf, **S. Krishnaswamy***, "Interpretable Neuron Structuring with Graph Spectral Regularization," in Advances in Intelligent Data Analysis (IDA) XVIII 2020.

48. A. Tong, J. Huang, G. Wolf, D. van Dijk, **S. Krishnaswamy***, "TrajectoryNet: A Dynamic Optimal Transport Network for Modeling Cellular Dynamics," in Proceedings of the 37th International Conference on Machine Learning (ICML) 2020.

49. J.S. Stanley III, S. Gigante, G. Wolf, and **S. Krishnaswamy***. "Harmonic Alignment." In Proceedings of the 2020 SIAM International Conference on Data Mining, pp. 316-324. Society for Industrial and Applied Mathematics (SIAM), 2020.

50. W. S. Chen, N. Zivanovic, D. van Dijk, G. Wolf, B. Bodenmiller, **Smita Krishnaswamy***, "Uncovering axes of variation among single cell cancer Specimens," Nature Methods 2020, https://doi.org/10.1038/s41592-019-0689-z

51. D. van Dijk, D. Burkhardt, M. Amodio, A. Tong, G. Wolf, and **S. Krishnaswamy***, "Finding Archetypal Spaces with Neural Networks," arxiv: 1901.09078, in Proceedings of IEEE Big Data 2019.

52. N. Brugnone, A. Gonopolskiy, M. W. Moyle, M. Kuchroo, D. van Dijk, K. R. Moon, D. Colon-Ramos, G. Wolf, M. J. Hirn, **S. Krishnaswamy**,* Coarse Graining of Data via Inhomogeneous Diffusion Condensation,, arxiv: 1907.04463, in Proceedings of IEEE Big Data 2019.

53. K.R. Moon, D. van Dijk, Z. Wang, W. Chen, M.J. Hirn, R. R. Coifman, N.B. Ivanova, G. Wolf, and **S. Krishnaswamy**,* "Visualizing structure and transitions in high-dimensional biological data," Nature Biotechnology 2019, vol. 37, n. 12, pp. 1482-1492, doi: 10.1038/s41587-019-0395-5.

54. M. Amodio, D. van Dijk, K. Srinivasan, W. S. Chen, K. R. Moon, A. Campbell, Y. Zhao, X. Wang, M. Venkataswamy, A. Desai, V. Ravi, P. Kumar, R. Montgomery, G. Wolf, and **S. Krishnaswamy**,* "Exploring single-cell data with multitasking deep neural networks," Nature Methods 2019, vol. 16, pp. 1139–1145, doi:10.1038/s41592-019-0576-7.

55. D. B. Burkhardt, J.S. Stanley, G. Wolf and **S. Krishnaswamy**,*"Vertex-Frequency Clustering," IEEE Data Science Workshop (DSW) 2019, pp. 145-149. doi: 10.1109/DSW.2019.8755591

56. D. van Dijk, S. Gigante, A. Strzalkowski, G. Wolf, and **S Krishnaswamy**,* "Modeling Global Dynamics from Local Snapshots with Deep Generative Neural Networks," arXiv, abs/1802.03497, Proceedings of Sampling Theory and Applications (SAMPTA) 2019.

57. S. Gigante, J.S. Stanley III, N. Vu, D. van Dijk, K. Moon, G. Wolf, **S. Krishnaswamy,*** "Compressed Diffusion," arXiv: 1902.00033, Proceedings of Sampling Theory and Applications, (SAMPTA) 2019.

58. M. Amodio and **S. Krishnaswamy**,* "TraVeLGAN: Image-to-image Translation by Transformation Vector Learning." Proceedings of Computer Vision and Pattern Recognition (CVPR) 2019, pp. 8983-8992.

59. O. Lindenbaum, J. S. Stanley III, G. Wolf, and **S. Krishnaswamy**,* Geometry-Based Data Generation. In proceedings of the 32nd Annual Conference on Neural Information Processing Systems (NeurIPS) 2018, pp. 1407-1418.

60. M.S. Shin, Y. Kang, E.R. Wahl, H.-J. Park, R. Lazova, L. Leng, M. Mamula, **S. Krishnaswamy**, Bucala, R., and Kang, I. (2018), "Macrophage migration inhibitory factor regulates U1snRNP immune complex mediated activation of the NLRP3 inflammasome," Arthritis & Rheumatology 2019, vol. 71(1), pp. 109-120. doi: 10.1002/art.40672.

61. van Dijk, D,. Sharma, R. Nainys, J., Yim, K., Kathail, P., Carr, A., Burdziak, C., Moon, K. R., Chaffer, C. L., Pattabiraman, D., Bierie, B., Mazutis, L. and Wolf, Guy, **S. Krishnaswamy***, D. Pe'er*, "Recovering Gene Interactions from Single-Cell Data Using Data Diffusion," Cell 2018, no. 174(3), pp. 716 – 729 doi.org/10.2139/ssrn.3155779

62. M. Amodio, and **S. Krishnaswamy**,* "MAGAN: Aligning Biological Manifolds." in Proceedings of the 35th International Conference on Machine Learning (ICML) 2018, PMLR vol. 80, pp. 205-223.

63. K.R. Moon, J.S. Stanley, D. Burkhardt, D. van Dijk, G. Wolf, and **S. Krishnaswamy**,* "Manifold learning-based methods for analyzing single-cell rna-sequencing data." Current Opinion in Systems Biology 2018, no. 7, pp. 36 – 46.

64. H. Ogura , P. Preston-Hurlburt, A.L. Perdigoto, M. Amodio, **S. Krishnaswamy**, P. Clark, H. Yu, D. Egli, A. Fouts, A. K. Steck, "Identification and Analysis of Islet Antigen–Specific CD8+ T Cells with T Cell Libraries," The Journal of Immunology (2018), ji1800267, p. 109-120.

65. **S. Krishnaswamy**#, R. Sharma#, N. Zivanovic, D. Pe'er, B. Bodenmiller, "Learning time-varying information flow from single-cell epithelial to mesenchymal transition data," PLOS One, 2018, vol. 13 (10), e0203389.

66. Y. Yao, T. Welp, Q. Liu, N. Niu, X. Wang, C. J. Britto, **S. Krishnaswamy**, G. L. Chupp, and R. R. Montgomery. "Multiparameter single cell profiling of airway inflammatory cells." Cytometry Part B: Clinical Cytometry (2017), vol. 92(1): pp. 12-20. doi: 10.1002/cyto.b.21491.

67. D.E. Lowther, B.A. Goods, L.E. Lucca, B.A. Lerner, K. Raddassi, D. van Dijk, A. Hernandez, X. Duan, X., M. Gunel, V. Coric, **S. Krishnaswamy**, J.C. Love, D. A. Hafler, PD-1 marks dysfunctional regulatory T cells in malignant gliomas. JCI insight (2016), vol.1(5), pii: e85935.

68. **S. Krishnaswamy**, M. Spitzer, M. Mingueneau, S.C. Bendall, E. L. Stone, O. Litvin, D. Pe'er, G. P. Nolan, "Conditional Density-based Analysis of T cell Signaling in Single Cell Data," Science 2014, 343 (6213), DOI:10.1126/science.1250689.

69. M. Mingueneau, **S. Krishnaswamy**, M. Spitzer, S.C. Bendall, E. L. Stone, S. M. Hedrick, D. Pe'er, D. Mathis, G. P. Nolan, C. Benoist, "Single-cell mass cytometry of TCR signaling: amplification of small initial differences result in low ERK activation in NOD mice," Proceedings of the National Academy of Sciences (PNAS) 2014, 111 (46), p.16466-16471, DOI:10.1073/pnas.1419337111.

70. E. Zunder, R. Finck, G. Behbehani E.D. Amir, **S. Krishnaswamy**, V.D. Gonzalez, C. G. Lorang, Z. Bjornson, M. H. Spitzer, B. Bodenmiller, W. J. Fantl, D. Pe'er, G. Nolan, "Palladium-based Mass-Tag Cell Barcoding with a Doublet-Filtering Scheme and Single Cell Deconvolution Algorithm," Nature Protocols 2015, 10, p. 316-333, DOI:10.1038/nprot.2015.020

71. E. Amir, K. L. Davis, M. Tadmor, E. Simonds, J. Levine, D. Shenfeld, **S. Krishnaswamy**, G. P. Nolan, D. Pe'er, "viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia," Nature Biotechnology 2013, 31(6): pp. 545-52.

72. **S. Krishnaswamy**, I.L. Markov, J. P. Hayes Design, Analysis and Test of Logic Circuits Under Uncertainty. Lecture Notes in Electrical Engineering 115, Springer 2013, ISBN 978-90-481-9643-2, pp. 1-120.

73. R. Finck, E. F Simonds, A. Jager, **S. Krishnaswamy**, K. Sachs, W. Fantl, D. Pe'er, GP. Nolan, S. C. Bendall, "Normalization of Mass Cytometry Data with Bead Standards," Cytometry Part A, Jan 2011.

74. H.Ren, R. Puri, L.N. Reddy, **S. Krishnaswamy**, C. Washburn, J. Earl, J. Keinert, "Intuitive ECO Synthesis for High Performance Circuits," Proc. DATE (DATE) 2013, pp. 1002-1007.

75. **S. Krishnaswamy**, H. Ren, N. Modi, R. Puri, "Logic Difference Optimization for Incremental Synthesis," Advanced Techniques in Logic Synthesis, Optimizations and Applications, S. Khatri and K. Gulati, eds; Springer, 2011. pp 203-225

76. T. Welp, **S. Krishnaswamy**, A. Kuehlmann, "Generalized SAT-Sweeping for Post- Mapping Optimization,"Proc. Design Automation Conference (DAC), 2012, pp. 814-819.

77. D. A. Papa, **S. Krishnaswamy**, I. L. Markov, "SPIRE: A Retiming-based Physical Synthesis Transformation System," In Proceedings of International Conference on Computer-Aided Design (ICCAD), 2010.
78. **S. Krishnaswamy**, H. Ren, N. Modi, R. Puri, "DeltaSyn: An Efficient Logic- Difference Optimizer for ECO Synthesis," International Conference on Computer- Aided Design (ICCAD) 2009.

79. **S. Krishnaswamy**, I. L. Markov, and J.P. Hayes, "Improving Testability and Soft- error Resiliance through Retiming," Proc. Design Automation Conference (DAC), 2009, pp. 513-518.

80. **S. Krishnaswamy**, S. Plaza, I. L. Markov, and J.P. Hayes, "Signature-based Design and SER Analysis of Logic Circuits," IEEE Trans. on CAD (TCAD), 2009, vol. 28, no. 2.

81. **S. Krishnaswamy**, G. Viamontes, I. L. Markov, and J.P. Hayes, "Probabilistic Transfer Matrices in the Symbolic Reliability Analysis of Logic Circuits," ACM Transactions on Design Automation of Electronic Systems (TODAES), 2008, no. 8.

82. **S. Krishnaswamy**, I. L. Markov, and J.P. Hayes, "On the Role of Timing Masking in Reliable Logic Circuit Design," in Proceedings of Design Automation Conference (DAC), 2008, pp. 924-929.

83. **S. Krishnaswamy,** S. Plaza, I. L. Markov, and J.P. Hayes, "Enhancing Design Robustness with Reliability-aware Resynthesis and Logic Simulation," in Proceedings of International Conference on Computer-Aided Design (ICCAD) 2007, pp. 149-154.

84. **S. Krishnaswamy**, I. L. Markov, and J.P. Hayes, "Tracking Uncertainty with Probabilistic Logic Circuit Testing," IEEE Design and Test (DT), August 2007.

85. **S. Krishnaswamy,** I. L. Markov and J. P. Hayes, "Testing Logic Circuits for Transient Faults," Proc. IEEE Eur. Test Symposium (ETS) 2005, pp. 102-107.

86. **S. Krishnaswamy**, G. F. Viamontes, I. L. Markov and J. P. Hayes, "Accurate Reliability Evaluation and Enhancement via Probabilistic Transfer Matrices", in Proceedings of Design Automation and Test in Europe (DATE) 2005, pp. 282-287 (Best Paper Award Winner, Circuit Test Category)

Invited Speaking Engagements, Presentations, Symposia & Workshops:

2024 Math Seminar Kalamazoo College, Kalamazoo Michigan
2024 Kavli 20th Anniversary Symposium, New Haven, CT
2024 Tea Talk, MILA, Montreal, CA
2024 Modeling and Computation of Optimal Transport with Applications in Biology, SIAM Annual Conference, Virtual
2024 Helmholz AI Conference, Dusseldorf, Germany
2024 Deep learning in Biology, Fondation des Treilles, Tourtour, France
2024 Single Cell Denmark Conference, Odense, Denmark
2024 Helmholz Bioengineering Conference, Munich, Germany
2024 Computational Biology Seminar, Pasteur Institute, Paris, France
2024 Dagstuhl Seminar: Low dimensional Embeddings of Data, Wadern, Germany
2024 Computational Biology Seminar, Pasteur Institute, Paris, France
2024 Data Science on Graphs (DeGAS) seminar series, virtual
2024 NCI Conference on Emerging Math in Cancer, NIH Campus, Bethesda, MD
2024 CZI conference on AI in Single Cell Biology, Redwood City, CA
2024 Probing disease with single cell genomics conference, Cancun, Mexico
2024 IPAM: Mathematical Approaches to Connectomics, UCLA, Los Angeles, CA
2024 Computational Oncology Seminar, Johns Hopkins, Baltimore, MD
2023 Generative AI in Biology Workshop, NeurIPS, New Orleans, LA
2023 Optimal Transport in ML Workshop, NeurIPS, New Orleans, LA
2023 NIH-FDA IIG Symposium on Immunology, Bethesda, MA
2023 Cold Spring Harbor Lab, Single Cell Analysis, New York
2023 NeuroAI, Montreal, Canada
2023 Single cells, systems biology symposium, Freiburg, Germany
2023 NSF-NIH Joint Workshop in Emerging AI in Biology, Virtual
2023 International Council for Industrial and Applied Mathematics (ICIAM), Tokyo, Japan
2023 Intelligent Systems for Molecular Biology, Lyon, France
2023 Applied Math Seminar, University of California, San Diego, CA
2023 AAI Immunology, FASEB award lecture, Washington DC
2023 Danish Single Cell Biology, Copenhagen, Denmark

2023 Harvard, Dept. of Applied Math Seminar, Cambridge, MA
2023 Vanderbilt, Systems Biology Department, Virtual Seminar
2022 Broad Institute/MIT EECS Joint Seminar, Cambridge, MA
2022 Cold Spring Harbor Labs, Cold Springs Harbor, NY
2022 National Institute for Biological Sciences, Bangalore, India
2022 Indian Institute of Science, Bangalore India
2022 Department of Mathematics, UCLA, CA
2022 Computational Genomics Summer Institute (CGSI), UCLA, CA
2022 Facebook AI Research (FAIR) Seminar, Virtual
2022 Atlanta Workshop on Single Cell Omics
2022 TorBug (Toronto Bioinformatics Users Group) Seminar, University of Toronto, Virtual
2022 Rencontres Santé Numérique (Digital Health Meeting), University of Montreal, Virtual
2022 Computational Biology Seminar, Carnegie Mellon University, Virtual
2022 Helmholtz Institute, Pioneer Campus, HPC Seminar Series
2022 Applied Math Seminar, Johns Hopkins University, Virtual
2021 Institute for Cognitive Science, Osnabrueck University, Germany, Virtual
2021 Mathematical Institute for Data Science, Johns Hopkins, Virtual
2021 Applied Math Seminar, University College London, Virtual
2021 Applied Math Seminar, Rochester Institute for Technology, Virtual
2021 Biostatistics Department Seminar, University of Michigan, Virtual
2021 Genetics Seminar, University of Washington St. Louis, Virtual
2021 Human Systems Biology Meeting, Indian Institute of Science, Virtual
2021 University of North Caroline, Chapel Hill, NC, Virtual
2021 ELIS (European Lab for Learning and Intelligent Systems) Seminar, Virtual
2021 Center for Mathematical Sciences and Applications, Harvard, Virtual
2021 ISMB Representation Learning for Biology (RLB 2021), Virtual
2021 SIAM conference on Applied and Combinatorial Discrete Algorithms (ACDA 2021)
2021 Computational Biology Symposium, Harvard Center for Mathematical Sciences and Applications , Virtual
2021 IDP and Human Genetics Seminar, UCLA, Virtual
2021 Chan-Zuckerberg Initiative, Seed Networks Computational Meeting, Virtual
2021 Systems Immunology, Cold Springs Harbor Laboratory, Virtual
2021 Human Translational Immunology Symposium, Yale, Virtual
2021 AACR Grantee Club Meeting, Virtual
2021 Eliciting Structure in Genomics Workshop, IMSI, University of Chicago
2020 Genome Informatics Conference, Hinxton UK, Virtual
2020 Systems Immunology and Aging Conference, Jax Labs, Virtual
2020 European Laboratory for Learning Intelligent Systems (ELLIS), Virtual
2020 International Conference on Machine Learning (ICML), Virtual
2020 Design Automation Conference (DAC), Virtual
2020 Intelligent Systems for Molecular Biology (ISMB), Virtual
2020 Intelligent Data Analysis (IDA), Germany, Virtual
2020 Cold Springs Harbor Laboratory, Systems Biology Conference, Virtual
2020 Michigan State University, Department of Computational Math and Engineering
2019 New York Genome Center, New York, NY
2019 Center for Physics and Biology, Rockefeller University, New York, NY
2019 Centre hospitalier de l'Université de Montréal (CHUM), Montreal, CA

2019 MILA Quebec Research Institute, Montreal, CA
2019 Michigan Institute of Data Science, University of Michigan, Ann Arbor, MI
2019 CSE Distinguished Alumni Seminar, University of Michigan, Ann Arbor, MI
2019 Allerton Conference on Computing and Control, Allerton, Illinois
2019 Sampling Theory and Applications, 2019, Bordeaux, France
2019 IEEE Data Science Workshop, 2019, Minneapolis, MN
2019 IEEE Graph Signal Processing Workshop, 2019, Minneapolis, MN
2019 Human Immune Project Consortium II, Annual Meeting, Rockville, MD
2019 Theory Lunch, Department of Systems Biology, Harvard Medical School
2019 Distinguished Seminar, Institute for Computational Medicine, Johns Hopkins University, Baltimore, MD
2019 Workshop at Bellairs – Single cell and Massively Parallel Approaches, Barbados.
2018 Institute for Computational Biomedicine (ICB), Weill Cornell Medicine, NY, NY
2018 56th Annual Allerton Conference, Monticello, IL
2018 International Conference on Machine Learning, Stockholm, Sweden
2018 ISMB, Chicago, Illinois
2018 Human Variation and Disease Gordon Research Conference, Biddeford, ME
2018 Computational Biology Department, Sloan Kettering Institute, New York, NY
2018 Chan-Zuckerberg Initiative, Investigators Meeting, Monterey, CA
2018 UCLA-Winter School in Computational Genomics, Los Angeles, CA
2018 DOLCIT Seminar, Caltech University, Pasadena, CA
2017 BIRS Challenges and Synergies in the Analysis of Large-Scale Biomedical Data, Oaxaca, Mexico

2017 Human Cell Atlas Meeting, Weizmann Institute Science, Rehovot, Israel
2017 Department of Immunology, Weizmann Institute of Science, Rehovot, Israel
2017 Department of Applied Mathematics, Tel Aviv University, Tel Aviv, Israel
2017 Department of Electrical Engineering, Technion University, Haifa, Israel
2017 Human Cell Atlas Meeting on Computation, Stockholm, Sweden
2016 Department of Systems Biology, University of California, San Francisco, CA
2016 Connect: Why Networks Matter, Young Scientist Symposium, Institute for Science and Technology, Klosterneuberg, Austria
2016 Computational Biology, Young Faculty Seminar, Johns Hopkins University, Baltimore, MD
2016 Department of Electrical and Computer Engineering, NYU Tandon School of Engineering, New York, NY
2016 Precision Medicine Conference, Montainview, CA

TEACHING AND EDUCATIONAL ACTIVITIES

The primary focus of my research is on developing methods for unsupervised and exploratory analysis of the data. Exploration of the data can be likened to exploring a jungle. There are no maps or labels but there are tools (like swiss-army knives). With the deluge of data that the world is currently immersed in from astronomy, biology, finance, and epidemiology alike, I consider it my passion to make the tools of machine learning and artificial intelligence accessible to all students**.** Indeed, many software tools are readily downloadable and usable. However, the concepts behind the tools remain inaccessible to those who may not have the right background in a

combination of linear algebra, probability and statistics, algorithms, and optimization theory. I have designed and taught the following courses.

Courses:

Course Title: Geometric and Topological Methods for Machine learning
Listing: Math 322, Applied Math 322, Computer Science 664,
Dates: Fall 2022
Enrollments: N/A
My role: Course designer and co-primary instructor (with Ian Adelstein)
Description: This course provides an introduction to geometric and topological methods in data science. Our starting point is the manifold hypothesis: that high dimensional data live on or near a much lower dimensional smooth manifold. We introduce tools to study the geometric and topological properties of this manifold in order to reveal relevant features and organization of the data. Topics include: metric space structures, curvature, geodesics, diffusion maps, eigenmaps, geometric model spaces, gradient descent, data embeddings and projections, and topological data analysis (TDA) in the form of persistence homology and their associated "barcodes." Also covered are applications of these methods in a variety of data types.

Course Title: Unsupervised Learning for Big Data
Listing: Computer Science 453/553, Applied Math 553, Computational Biology 555
Dates: Fall Semesters 2016-2021
Enrollments: 31 students (Fall 2016), 18 students (Fall 2017), 46 students (Fall 2018), 27 students (Fall 2019), 74 students (Fall 2020), 82 students (Fall 2021)
My role: Course designer and primary instructor
Description: This course includes topics not normally taught in basic machine learning courses such as dimensionality reduction (including non-linear), data denoising, information theory, optimal transport, graph representation and graph signal processing, as well as self-supervised neural networks. The course includes three problem sets with programming tasks as well as a comprehensive final project in which the student explores their own interests in data science and problems deriving from their datasets of interest. This course has been taken by a wide variety of students ranging from departments such as Genetics, Neuroscience, Forestry and Environmental Science, Astrophysics, Economics, and of course Computer Science (where it is also taken by undergraduates) and Applied Mathematics.

Course Title: Deep Learning Theory and Applications
Listing: Computer Science 452/552, Applied Math 552, Computational Biology 663
Dates: Spring Semesters 2019-present
My role: Course designer and primary instructor
Description: Deep neural networks have gained immense popularity within the last decade due to their success in many important machine learning tasks such as image recognition, speech recognition, and natural language processing. This course provides a principled and hands-on approach to deep learning with neural networks. By the end of the course, students will have mastered the principles and practices underlying neural networks including optimization and training methods, design and architecture, generalization theory, loss landscapes, and will have applied deep learning methods to real-world problems including image recognition, natural language processing, and biomedical applications. The course is based on homework, a final exam, and a final project

(either group or individual, depending on the total number enrolled). This course too is enrolled by an extremely wide variety of students.

Course Title: Advanced Topics in Machine Learning and Data Mining
Listing: Computer Science 745, Applied Math 745, Computational Biology Bioinformatics 745
Dates: Spring Semesters 2017-2018
Enrollments: 16 students (Spring 2017), 19 students (Spring 2018)
My role: Co-Designer and Co-Instructor (together with Dr. Guy Wolf and Dr. Alex Cloninger), Taught roughly 5 lectures and guided discussions.
Description: This is a seminar course that specializes in a current topic in machine learning and data mining each year and then builds both background as well as surveyed new developments in the area throughout the semester. In Spring 2017 the course focused on ML methods for biomedical data analysis, and in Spring 2018 the course focused on topics in Deep Learning. The first half of the course involved lectures by instructors, and the second half of the course involved in-depth (2-hour) presentations by student teams focusing on specific subtopics.

Workshops and Summer Programs:
Title: Machine Learning for Single Cell Analysis
Dates: Fall 2019, Winter 2020, Fall 2021
My role: Course designer and primary instructor
Description: This is a 6-day "bootcamp" taught over two weeks focusing on how to understand unsupervised data analysis by way of a new emerging technology in biomedical science that hundreds of labs across the world are using: single cell data. Single cell data usually involves thousands to millions of cells which are measured in tens of thousands of dimensions. This is the quintessential problem of unsupervised learning. However, the labs that generate the data are often not equipped to analyze it. Therefore, our bootcamp consists of 6 succinct lectures that develop intuition on the concepts as well as hands-on computer lab sessions to familiarize students with tools needed to analyze the data. Topics include preprocessing, dimensionality reduction, visualization, denoising, batch normalization, and data generation. An active team of Yale-student TAs are involved in teaching this workshop online. Materials of this bootcamp are available at krishnaswamylab.org/workshop.

Title: NSF Summer Mathematics Research at Yale (SUMRY)
Dates: Summers 2021-2024
My role: Mentor and lecturer of the Applied Math Group in Sumry
Description: Designed an 8-week research REU program for undergraduates studying mathematics on topics at the intersection of deep learning, graphs spectral theory, data geometry and topology. Students have conducted research on topics such as measuring curvature in high dimensional data (Bhaskar et al. NeurIPS 2022), enhancing the power of graph neural networks using geometric scattering (Xu et al. AIStats 2024), visualizing data that has dynamics (in the form of velocity arrows) (Macdonald et al. MLSP), and examining the space of neural networks trained for specific tasks (in preparation). The mentoring component of this program involved daily meetings with the student group as well as advice in scientific writing, research strategy as well as testing and benchmarking.

Committees:

1. **Department of Genetics, Education Committee:** I am a member of the education committee in the Genetics department and I have contributed to discussions and decisions involving education in Genetics. The education committee has discussed and optimized a wide variety of educational activities in the Genetics department from redesigning the seminar schedule and the way speakers are chosen, to providing postdocs with a venue for speaking (during the lab interaction hours) to scoring in qualification exams. I have participated in all discussions of the committee and gathered opinions from students in the department barring the meetings during my maternity leave in 2017.
2. **Computer Science advisory panel to the University Science Strategy Committee (USSC):** in the 2017-2018 year. This panel also included key figures in Computer Science such as former chair of CS and pioneer in security research Prof. Joan Feigenbaum, current Computer Science chair Prof. Zhong Shao, prominent robotics professor Prof. Brian Scassellati, and world-renowned Natural Language Processing expert Prof. Dragomir Radev . This committee advised the USSC on how to move computer science forward at Yale generally. We a white paper entitled "How can Yale help define and shape our increasingly computational world?" that contained many innovative proposals including the "CS+X" proposal to allow students in any department to combine their major with computer science. For instance X can be economics, political science, biology or even music or art. The idea here is that CS would be integrated as a "vertical theme" to enable these particular domains in the modern world.
3. **Data Science Advisory Committee to the University:** I started to serve on the data science advisory committee in 2021. This committee plans educational and departmental data science activities across the university, and writes white papers that departments will execute over the next 10 years in cross disciplinary data science. Other professors involved include Provost Scott Stroebel, Dean of Engineering Jeff Brock, Sterling Professors Ronald Coifman and Dan Spielman.
4. **Basic Science YSM panel for the Yale Taskforce for AI Task:** Presented and proposed possibilities for the use of AI in advancing basic science both in terms of the research process, as well as in hypothesis generation. Others on the panel included Prof. Steven Kleinstein, Prof. Yuval Kluger, Prof. Tony Koleske and others.

MENTORING

I strive to create an active and nurturing research environment in my laboratory where students and researchers from all levels (undergraduates, medical students, Ph.D. students, postdocs, clinical fellows and associate researchers) come together to work on a wide variety of problems in biomedical and genomic data science. I specifically encourage openness and out-of-the-box thinking by myself setting the tone and listening to anyone's ideas at any level and then providing them with scientific and computational resources, mentorship or collaborators to carry for their ideas. I also involve lab member in several unique activities designed to enhance their research productivity and creativity:

i) I set up a number of different forums in addition to the canonical lab meetings and weekly one-on-one meetings. These include hackathons, when the entire lab comes together for 2-day stretches to work intensively on a computational problem involving the analysis of biomedical data. Last year our hackathon resulted in a key project in the lab called MELD (Manifold Enhancement of Latent Dimensions) which can be used enhance comparisons on data collected on the same system under different treatment conditions.

ii) I seek opportunities for my lab members to attend a wide variety of scientific events and meetings outside the lab including biological (e.g. Single Cell Genomics Conference, AACR conference) and computer science conferences (NIPS, ICML) , workshops, external hackathons such as the Data Hack

Yale (http://datahack.yale.edu/) which my graduate student Scott Gigante and undergraduate student Adam Ericksen won first place in 2018 and Cytohack (www.cytohack.org) conducted by the Broad Institute in Cambridge, MA.

iii) Active lab slack channels where lab members can discuss any topic of research at any time, whether it is to ask questions, post results, share research news or brainstorm ideas. I notice that our lab slack is active nearly 24-7.

iv) We have begun a "blog" as well as a help slack such that we can give advice to other researchers who are interested in learning about data science or our particular data science tools.

v) Our lab spearheaded the effort to start an open problems in single cell analysis effort, together with Helmholz institute in Germany as well as Chan-Zuckerberg Initiative Biohub. This effort is intended as a community-contributed effort to collect problems (i.e., tasks), methods, datasets as well as metrics for scoring benchmarks in a scalable automated platform for benchmarking methods for single cell analysis. As part of this effort, we had a jamboree in May which was joined by 50 individuals from across the world and already involved contributions in tasks such as clustering, dimensionality reduction, trajectory inference, multimodal data integration, etc. Further, we proposed a coding competition based on the multimodal data inference task to NeurIPS 2021, which was accepted in their competition track and is under way presently.

vi) I have advised my mentees in innovation and entrepreneurship efforts, guiding two groups of my students in applications to the Yale Blavatinik Innovation Fund as well as the Nucleate Bio incubation programs. Currently my student groups are finalists for both in the 2022 competition.

vii) I also advise a large number of undergraduates in their senior thesis, often tutoring them in prevalent theories, brainstorming project ideas, and directing their efforts fruitfully in their chosen topics.

Current Lab Members

Postdoctoral Members:

Name of trainee: Dhananjay Bhaskar
Position and period of mentorship: Postdoctoral Researcher and Yale Boehringher-Ingelheim Fellow June 2021-present
Research project: Topological data analysis for biomedical data, graph representation and generation for drug discovery

Graduate Students:

Name of Trainee: Chen Liu
Position and period of mentorship: Graduate student in Computer Science, Fall 2023-present
Research Project: Biomedical imaging and segmentation, EHR data forecasting

Name of Trainee: Danqi Liao
Position and period of mentorship: Graduate student in Computer Science, Fall 2023-present
Research Project: Biomedical imaging and segmentation, RNA-degradation prediction

Name of Trainee: Xingzhi Sun
Position and period of mentorship: Graduate student in Computer Science, Fall 2023-present
Research Project: Single cell dynamics, hypergraph scattering

Name of Trainee: Siddharth Visawanath
Position and period of mentorship: Graduate student in Computer Science, Fall 2023-present
Research Project: Geometric Scattering for molecule generation, metabolic network modeling
Publications: RECOMB, PMLR

Name of trainee: Kevin Bijan Givenchian
Research Project: Cancer vaccine optimization, and immunogenicity prediction from sequence
Position and period of mentorship: M.D./Ph.D. student, 2020-present

Undergraduates:

Name of trainee: Kincaid Macdonald
Position and period of mentorship: Undergraduate Student in Applied Math; May 2018-present
Research project: Geometric scattering in neural networks
Publications: MLSP 2021

Name of trainee: Andrew Benz
Position and period of mentorship: Undergraduate Student in Computer Science; August 2018-present
Research project: Neural networks for data anomaly detection, FMRI data analysis

Name of trainee: Katherine Du
Position and period of mentorship: Undergraduate Student in Applied Math; August 2016-present
Research project: Modeling cell-cell communication

Name of trainee: Michal Girasimiuk
Position and period of mentorship: Undergraduate Student in Applied Math; August 2016-present
Research project: Modeling patient data with manifold forests

Name of trainee: Tom Wallenstein
Position and period of mentorship: Undergraduate Student in Applied Math; August 2016-present
Research project: Reinforcement learning in FMRL

<u>Former Lab Members</u>

Former postdoctoral trainees:

Name of trainee: Guy Wolf, PhD
Position and period of mentorship: Associate Research Scientist; 2018-2019, Gibbs Fellow; 2015-2018
Research focus: Data geometry and harmonics
Current Position: Associate Professor, Department of Mathematics/MILA, University of Montreal, CA

Name of trainee: David van Dijk , PhD
Position and period of mentorship: Associate Research Scientist December 2017-September 2019
Research focus: Manifold Learning for single-cell data
Current Position: Assistant Professor, Department of Medicine, Yale University

Name of trainee: Kevin Moon
Position and period of mentorship: Postdoctoral Researcher; August 2016-August 2018
Research focus: Data Visualization with Manifold Learning
Current Position: Assistant Professor, Department of Mathematics and Statistics, Utah State University, UT

Name of trainee: Feng Gao
Position and period of mentorship: Postdoctoral Researcher; September 2019-present
Research project: Analysis of Spatial Transcriptomics and Imaging Data
Current Position: Postdoctoral Researcher at Columbia University, New York, NY

Name of trainee: Jiexi Huang
Position and period of mentorship: Postdoctoral Researcher; September 2019-present
Research project: Modeling Dynamics in Data with Neural Networks, multiscale visualization of clinical data
publications: ICML 2020, Nature Biotechnology 2021
Current position:

Name of trainee: Holly Steach
Position and period of mentorship: Postdoctoral Fellow; October 2021-May 2024
Research project: Directed graph neural networks for metabolics and cellular communication
Publications: Applied and Harmonic Analysis, PMLR, RECOMB
Current Position: Senior Scientist at Arsenal Biosciences

Name of trainee: Edward De Brouwer
Position and period of mentorship: Postdoctoral Researcher March 2023-April 2024
Research project: Data visualization, EHR data analysis
Publications: NeurIPS, under review at Journal of Surgery
Current position: Senior Scientist at Genentech

Name of trainee: Scott Youlten
Position and period of mentorship: Postdoctoral Researcher January 2023-2024
Research project: Computational Cancer and Developmental biology
Publication: In revision at Cancer Discovery
Current position: Associate Research Scientist in the Giraldez Lab at Yale University

Former doctoral trainees:

Name of Trainee: Aarthi Venkat
Position and period of mentorship: Graduate student in Computational Biology 2020-present
Research Project: Unsupervised learning of cellular data manifolds
Awards: Gruber Fellowship, Cancer Biology Training Grant
Publications: ICASSP, Genome Research, Trends in Immunology, Under review in Cancer Discovery

Current Position: Eric and Wendy Schmidt Fellow at Harvard University/Broad Institute

Name of Trainee: Daniel Sumner Magruder
Position and period of mentorship: Graduate student in Computer Science Fall 2021-present
Research project: Neural network design based on insights from neuroscientific data
Publications: NeurIPS, LOG, Developmental Cell
Current position: Postdoctoral Researcher, Yale Dept of Neurology

Name of trainee: Egbert Castro
Position and period of mentorship: Graduate Student. In Computational Biology; Fall 2018-present
Research project: Graph Convolutional Neural Networks for Protein Data
Publications: IEEE Big Data, Nature Scientific Reports, Nature Machine Intelligence,
Current Position: CEO of Ascent Bio, Incubated at the Merck Accelerator

Name of trainee: Dennis Shung
Position and period of mentorship: Graduate Student in Investigative Medicine/Clinical Fellow; Fall 2019-present
Research project: Neural networks for dynamic risk modeling, patient representation, and prediction
Publications: Nature Scientific Reports, Gastroenterology
Awards: K-23 Diabetes, Digestive, and Kidney Diseases Extramural Research, American Gastroenterology Association Research Scholar Award, Dostanovic Award
Current Position: Faculty at Yale University

Name of trainee: Matt Amodio
Position and period of mentorship: Graduate Student in Computer Science; September 2017-present
Research project: Generative Neural Networks
Publications: ICML 2018, Arthritis and Immunology 2018, Journal of Immunology 2018, CVPR 2019, Nature Methods 2020, Cell Patterns 2021, IJCAI 2021
Awards: Jeopardy Champion
Current Position: Eric and Wendy Schmidt Fellow, Broad Institute of Harvard and MIT

Name of trainee: Manik Kuchroo
Position and period of mentorship: Medical Student; April 2018-present, Graduate Student in Neuroscience: May2019-present
Research project: Coarse graining and multiscale clustering, applications to cancer biology
Publications: IEEE Big Data 2019, Nature Biotechnology, MLSP 2021
Awards: NIH F32 Award
Current Position: Resident at Mass General Hospital (Physician-Scientist Track)

Name of trainee: Alexander Tong
Position and period of mentorship: Graduate Student in Computer Science; June 2018- present
Research project: Learning latent data structures with Deep Learning
Publications: Intelligent Data Analysis 2019, ICML 2020, ICML 2021, MLSP 2020, MLSP 2021
Awards: Best student paper award MLSP 2020
Current Position: Postdoctoral Researcher at MILA/Montreal with Prof. Yoshua Bengio, to be Assistant Professor at Duke University

Name of trainee: William Chen
Position and period of mentorship: Medical Student; October 2015-2020
Research project: Modeling Sample State Space with Earth Mover's Distance
Publications: Nature Methods 2020
Current Position: Resident at University of California, San Francisco Hospital

Name of trainee: Scott Gigante
Position and period of mentorship: Graduate Student in Computational Biology; October 2017-2021
Research project: Visualizing data using data geometry
Publications: NeurIPS 2019, SAMPTA 2019, SAMPTA 2019, Nature Biotechnology 2019
Awards & honors: Gruber Fellowship, First Place Data Hack Yale, Honorable Mention IEEE Viz
Current position: Immune AI

Name of trainee: Daniel Burkhardt
Position and period of mentorship: Graduate Student; April 2016-2021
Research project: Single cell compositional and distributional analysis
Publications: Current Opinions in Systems Biology 2016, Nature Biotechnology 2019, IEEE Data Science Workshop 2019, Nature Biotechnology 2021
Awards & Honors: F32 NIH Graduate Research Fellowship, Data Science Workshop Travel Award, Best Poster Award at Genetics Departmental Retreat, Carolyn Slayman Prize in genetics
Current Position: Cellarity

Name of trainee: Jay Stanley
Position and period of mentorship: Graduate Student; April 2017- 2020
Research project: Using Graph Signal Processing in Biomedical Data
Current Position: Postdoctoral Researcher, Program for Applied Mathematics, Yale University

Former undergraduate trainees:

Name of trainee: Krishnan Srinivasan
Position and period of mentorship: Undergraduate Student in Computer Science; August 2017-January 2018
Research project: Deep Autoencoders for Data Processing
Current Position: Graduate Student, Stanford

Name of trainee: Alex Strzalkowski
Position and period of mentorship: Undergraduate Student in Applied Math; August 2017-May 2018
Research project: Processing Image-based Biomedical Data
Current Position: Graduate Student, Princeton

Name of trainee: David Darrow
Position and period of mentorship: High School Student; August 2017-present
Research project: Batch normalization of data using Signal Processing
Current Position: Undergraduate Student, MIT

Name of trainee: Gabriel Dolston
Position and period of mentorship: Undergraduate Student in Applied Math; August 2018-May 2019
Research Project: Vertex Frequency Clustering of Graphs with Signals
Current Position: Graduate Student, Princeton

Thesis Committees

Molecular Cellular and Developmental Biology: Gaddareth Higgs

Interdisciplinary Neuroscience Program: Kailong Peng, Emily Siff

Clinical Psychology Program: Erica Busch

Genetics: Danielle Miyagishima, Mark Noble, Mattheo Morales

Biomedical Engineering: Zhuo Chen, Amanda Alexander

Computer Science: Xiangru Tang

Computational Biology/Bioinformatics: Jake Sumner, Anna Su

Qualifying Committees

Computational Biology Bioinformatics: Vimig Socrates

Interdisciplinary Neuroscience Program: Kailong Peng, Antonio Fonesca

Genetics: Syndi Barish, Nicholas Dias, Sydney Muchnik, Atreyo Pal, Evan Geller, Mattheo Morales, Mackenzie Noon

Senior (Undergraduate) Thesis Advising

Computer Science Students: Tucker Hart, Tom Wallenstein, Jeffrey Wang, Jeffrey Zhou, Lucas Corey, Anie Gao, Leonardo Gomez, Anushka Acharya
Applied Math Students: Nikita Klemenko, Adam Eriksen, Abhinav Godhavarthi, Sam Leone, Marcus Hodgson
Math Students: Andrew Benz, Ellie Shang, Charles Xu
Molecular Cellular Developmental Biology: Katherine Du

PROFESSIONAL SERVICE

NIH

2022	Panelist for National Institute of Neurological Disorders and Stroke (NINDS) R01 on Leveraging Existing ADRD Data Resources Special RFA
2019	Ad-Hoc Member, Study Section on Genomics Computational Biology and Technology (GCAT)
2019-2022	External Program Consultant, Extracellular RNA Communication, National Institute for Drug Abuse (NIDA)
2021	Ad-Hoc Member, National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) DDK-C Subcommittee.
2010	Co-chair of Post-CMOS VLSI Track, Technical Program Committee, Great Lakes Symposium on VLSI Design
2011	Co-chair of Technical Program Committee, International Workshop on Bio-Design Automation
2010-2015	Technical Program Committee, Design Automation and Test in Europe
2013-2015	Technical Program Committee, Cancer Panomics Conference
2013	Organizer and TPC co-chair, Modeling of Biological Systems Workshop

2014-2015	Technical Program Committee, Design Automation Conference
2014-2016	Technical Program Committee, International Conference on Computer-Aided Design
2016-present	Technical Program Committee, RECOMB, Regulatory Genomics
2016-present	Technical Program Committee, Neural Information Processing Systems (NeurIPS)
2016-present	Technical Program Committee, International Conference on Machine Learning (ICML)
2018-pesent	Technical Program Committee, Reviewer, International Conference on Representation Learning (ICLR)
2020-2021	Technical Program Committee, AAAI Artificial Intelligence Conference
2020-2021	Technical Program Committee, Area Chair, Machine Learning for Computational Biology (MLCB)
2022-2023	Technical Program Committee, Area Chair, Learning on Graphs (LOG)
2023	General chair, Sampling Theory and Applicaitons (SAMPTA)
2024	Program Chair, Learning on Graphs (LOG)

Peer Review Groups for Conferences

Journal Service:
Reviewer for Nature Methods, BMC Bioinformatics, Elife, PLOS Computational Biology, Nature Communications, Cell, Science, Nature Biotechnology

Editorial Board Member:
Genome Research, Cell Patterns (2020-2023), E-Life

Professional Organizations:
2008-2010 IEEE Council on Design Automation (CEDA) Technical Activities Committee
2008-2012 ACM Special Interest Group on Design Automation (SIGDA)
2001-present Eta Kappa Nu honor society
2016-present Human Cell Atlas Consortium Member

Meeting Planning/Participation

2013	Co-organizer and Program Chair, Modeling of Biological Systems (MOBS), co-located with the Design Automation Conference, Austin, Texas
2019-2020	Organizing Committee of Learning Representations of Life Workshop, NeurIPS
2019-2020	Organizer and Instructor of Machine Learning for Single Cell Analysis Workshop, Sponsored by Yale SEAS, Genetics Dept.
2020	Organizing Committee of Topological Data Analysis and Beyond Workshop, NeurIPS
2021	Organizing committee of ACM-BCB, Workshop on Methods for Single Cell Analysis
2021	Organizing committee of Geometric and Topological Representation Learning, ICLR
2021	Co-organizer of ICERM Workshop on Data Geometry and Topology
2023	General Chair of Sampling Theory and Applications
2023	Co-organizer of New Frontiers in Graph learning, NeurIPS 2023
2024	Co-organizer of Joint Math Meetings Special Session on Geometry and Topology in Biomedical data
2024	Program Chair of Learning on Graphs

Research Organization Efforts

2021-2022	Organizer of NeurIPS competition on Multi-modal single cell data integration

2021 Started the Open problems in single cell analysis effort with CZI (openproblems.bio)

***Service - Yale University**:*
University Committees:

2017-2018	Member of CS Advisory Panel for University Science Strategy Committee
2017, 2021	Computational Biology Bioinformatics program Admissions Committee
2019	Faculty Search Committee, Microbial Sciences Institute, Department of Genetics
2021	Yale-Boehringer Ingelheim Fellowship Selection Committee
2021	Computational Biology Bioinformatics program Admissions Committee
2021	Wu-Tsai Institute Membership Committee
2021-2022	Data Science advisory committee
2021	Wu-Tsai Institute Working group for Neurocomputation and Machine Learning
2021-2022	Faculty Search Committee, Wu-Tsai Institute
2021	Faculty Search Committee, Department of Biomedical Engineering, Yale University
2021-present	Educational Committee, Dept. Genetics
2021-Present	Yale-BI Partnership Steering Committee
2022	Steering Committee, Yale Institutes for Foundations of Data Science
2023	Faculty Search Committee, CS Dept.
2023	Faculty search committee, Biostatistics-Data Science, YPH
2023	Computational Biology Bioinformatics, Faculty Search Committee
2023	Dept. CS, Teaching and Curriculum Committee
2023	Basic Science Panel for YTAI, YSM

Medical School Committees:

2016-present	Education Committee for the Department of Genetics, Yale University
2016-2018, 2021	Faculty Search Committee, Department of Genetics, Yale University
2017-2018	Faculty Search Committee, Department of Immunobiology, Yale University
2021-2022	Faculty Search Committee for Associate Dean of Biomedical Informatics