Chirag Lakhani

New York, NY | chirag (dot) lakhani (at) gmail.com | clakhani (at) nygenome (dot) org | cmlakhan.github.io
linkedin.com/in/chiraglakhani/ | github.com/cmlakhan

Positions Held

Senior Research Scientist, New York Genome Center 2021 – Present

- Principal Investigators: David Knowles (Columbia University) and Towfique Raj (Mt. Sinai School of Medicine)
- Research Scientist for an artificial intelligence U01 grant funded through the NIH Alzheimer's Disease Sequencing Project (ADSP).
- Led projects related to the functional interpretation of rare and common AD risk variants using genomic deep learning models.
- Collaborated closely with the ADSP Functional Genomics Consortium (FunGen-AD) on projects related to the prediction of molecular QTLs using genomics deep learning models.

T32 Training Fellow, Harvard Medical School 2018 – 2020

- Co-Advisors: Chirag Patel (Harvard Medical School) and James Meigs (Massachusetts General Hospital)
- Funded by an NIH Training Grant aimed at training postdoctoral fellows with a quantitative background to solve problems in human metabolic health, such as obesity and Type 2 diabetes.

Postdoctoral Research Fellow, Harvard Medical School - Department of Biomedical Informatics 2015 – 2020

- Advisor: Chirag Patel (Harvard Medical School)
- Led the development of "Exposome Data Warehouse," a database of aggregated publicly available environmental data (socioeconomic factors, pollution, and climate) that links to electronic health records.
- Conducted an analysis where we repurposed insurance claims data to conduct the largest twin study in the United States in order to understand the role of genetics and the environment for 560 phenotypes. We leveraged "Exposome Data Warehouse" to also understand the contribution of environmental exposure to these phenotypes.

Data Scientist, HelloWallet, Inc 2014 – 2015

- Led initiative to prototype and productionize a machine learning-based transaction classification system in the HelloWallet app. Work ranged from data cleaning and prototyping algorithms using scikit-learn to developing a technical approach for productionizing the app in a Java-based app stack.
- Led data engineering initiative to migrate analytics database from MySQL environment to a Hadoop/Spark-based environment.

Data Scientist, Zaloni, Inc 2013 – 2014

- Led new data science initiative at Zaloni, implementing large-scale machine learning solutions to meet client needs (such as fraud detection or scalable clustering algorithms for customer segmentation). Served as liaison between clients and development teams in the US and India to create data science products of value. Developed "Introduction to Data Science" training course focusing on applying Mahout to data science problems in the real world; taught at several Fortune 500 companies.

Independent Researcher, Duke University and SAMSI 2012 – 2013

- Advisor: Mauro Maggioni (Duke University)
- Focused on developing scalable solutions to traditional nonlinear dimensionality reduction algorithms. Collaborated with Duke Prof. Mauro Maggioni and his group on leveraging geometric multi-resolution analysis to compute manifold learning algorithms efficiently.

Postdoctoral Research Associate, North Carolina State University 2011 – 2011

- Advisor: Hamid Krim (North Carolina State University)
- Collaborated with Prof. Hamid Krim and his group on applying computational topology to problems in data mining. Primarily interested in the use of persistent cohomology for nonlinear dimensionality reduction. Also considered the use of Morse theory for detecting community structures in complex networks.

Education

North Carolina State University, PhD in Mathematics — December 2010

- Thesis: The GIT Compactification of Quintic Threefolds
- Advisor: Amassa Fauntleroy

North Carolina State University, B.S. in Mathematics — December 2003

- Minor: Physics
- Magna Cum Laude

Publications

[1] Anjali Das, Lakhani, Chirag M, Chloé Terwagne, Jui-Shan T Lin, Tatsuhiko Naito, Towfique Raj, and David A Knowles. "Leveraging functional annotations to map rare variants associated with Alzheimer's disease with gruyere". In: *medRxiv* (2025). accepted for publication in *American Journal of Human Genetics*, pp. 2024–12.

[2] Lakhani, Chirag M, Jui-Shan T Lin, Anjali Das, Tatsuhiko Naito, Towfique Raj, and David A Knowles. "Integration of Deep Learning Annotations with Functional Genomics Improves Identification of Causal Alzheimer's Disease Variants". In: *medRxiv* (2025), pp. 2025–03.

[3] Tatsuhiko Naito, Kosei Hirata, Beomjin Jang, Lakhani, Chirag M, Alice Buonfiglioli, Wan-Ping Lee, Otto Valladares, Li-San Wang, Yukinori Okada, Hong-Hee Won, et al. "Mosaic chromosomal alterations in blood are associated with an increased risk of Alzheimer's disease". In: *medRxiv* (2025), pp. 2025–05.

[4] Bryan Andrews, Chirayu Wongchokprasitti, Shyam Visweswaran, Lakhani, Chirag M, Chirag J Patel, and Gregory F Cooper. "A new method for estimating the probability of causal relationships from observational data: Application to the study of the short-term effects of air pollution on cardiovascular and respiratory disease". In: *Artificial intelligence in medicine* 139 (2023), p. 102546.

[5] Sara J Cromer, Lakhani, Chirag M, Josep M Mercader, Timothy D Majarian, Philip Schroeder, Joanne B Cole, Jose C Florez, Chirag J Patel, Alisa K Manning, Sherri-Ann M Burnett-Bowie, et al. "Association and interaction of genetics and area-level socioeconomic factors on the prevalence of type 2 diabetes and obesity". In: *Diabetes Care* 46.5 (2023), pp. 944–952.

[6] Rockwell J Weiner, Lakhani, Chirag M, David A Knowles, and Gamze Gürsoy. "LDmat: efficiently queryable compression of linkage disequilibrium matrices". In: *Bioinformatics* 39.2 (2023), btad092.

[7] Devi Sai Sri Kavya Boorgu, Shruthi Venkatesh, Lakhani, Chirag M, Elizabeth Walker, Ines M Aguerre, Claire Riley, Chirag J Patel, Philip L De Jager, and Zongqi Xia. "The impact of socioeconomic status on subsequent neurological outcomes in multiple sclerosis". In: *Multiple sclerosis and related disorders* 65 (2022), p. 103994.

[8] Andrew Deonarine, Genevieve Lyons, Lakhani, Chirag M, and Walter De Brouwer. "Identifying communities at risk for COVID-19–related burden across 500 US cities and within New York City: unsupervised learning of the coprevalence of health indicators". In: *JMIR Public Health and Surveillance* 7.8 (2021), e26604.

[9] Yixuan He, Lakhani, Chirag M, Danielle Rasooly, Arjun K Manrai, Ioanna Tzoulaki, and Chirag J Patel. "Comparisons of polyexposure, polygenic, and clinical risk scores in risk prediction of type 2 diabetes". In: *Diabetes Care* 44.4 (2021), pp. 935–943.

[10] Sara J Cromer, Lakhani, Chirag M, Deborah J Wexler, Sherri-Ann M Burnett-Bowie, Miriam Udler, and Chirag J Patel. "Geospatial analysis of individual and community-level socioeconomic factors impacting SARS-CoV-2 prevalence and outcomes". In: *MedRxiv* (2020).

[11] Lakhani, Chirag M, Braden T Tierney, Arjun K Manrai, Jian Yang, Peter M Visscher, and Chirag J Patel. "Repurposing large health insurance claims data to estimate genetic and environmental contributions in 560 phenotypes". In: *Nature genetics* 51.2 (2019), pp. 327–334.

[12] Lakhani, Chirag M. "The GIT compactification of quintic threefolds". In: *arXiv preprint arXiv:1010.3803* (2010).

Academic Talks

MLxQTL: Machine Learning Based Prediction of Single Cell eQTLs in Neuronal Cell Types May 2025

NIH Alzheimer's Disease Sequencing Project xQTL Workshop, San Francisco, CA

Learning the Regulatory Code of Alzheimer's Disease Genomes October 2024

NIH Alzheimer's Disease Sequencing Project AI/ML External Advisory Board Meeting, Bethesda, MD

Learning the Regulatory Code of Alzheimer's Disease Genomes August 2024

NIH Alzheimer's Disease Sequencing Project AI/ML Working Group, Online

Deep Learning Approaches for Functional Genomic Variant Prediction using xQTL Data March 2024

NIH Alzheimer's Disease Sequencing Project xQTL Workshop, New York, NY

Deep Learning Analysis of Genetic Variants in PICALM/EED Locus March 2023

NIH Alzheimer's Disease Sequencing Project Functional Genomics Meeting, Online

Repurposing Large Health Insurance Claims Data and Publicly Available Exposomic Data to Estimate Genetic and Environmental Contributions in 560 Phenotypes September 2019

Chan-Zuckerberg Initiative Workshop, CU Anschutz, Aurora, CO

Building a Search Engine to Find Environmental and Phenotypic Factors Associated with Health and Disease March 2017

NSF Big Data Spokes Meeting, Washington, DC

Building an Exposome API February 2016

Environmental Statistics Seminar, Harvard School of Public Health, Boston, MA

Systematic and Large-Scale Investigation of Twin and Sibling Concordance of 1723 Traits in a Nationally Representative Health Claims Cohort October 2015

Platform Talk, American Society of Human Genetics, Baltimore, MD

Approximation of Kernel Matrices November 2012

SAMSI Working Group, SAMSI, Research Triangle Park, NC

Tensor Decompositions in Signal Processing October 2011

Tensor Seminar, North Carolina State University, Raleigh, NC

Schemes and Varieties July 2005

Visiting Student Seminar, Tata Institute of Fundamental Research, Mumbai, India

Poster Presentations

Integration of Deep Learning Annotations with Functional Genomics Improves Identification of Causal Alzheimer's Disease Variants July 2025

Gordon Research Conference - Alzheimer's Disease, Venture Beach, CA

Localization of Polygenic Signal in Alzheimer's Disease through the Integration of Cell-Type Functional Annotations and Deep Learning Models May 2024

Biology of Genomes, Cold Spring Harbor, NY

Localization of Polygenic Signal in Alzheimer's Disease through the Integration of Cell-Type Functional Annotations and Deep Learning Models March 2024

Broad Institute - Variant To Function Meeting, Cambridge, MA

Localization of Polygenic Signal in Alzheimer's Disease through the Integration of Cell-Type Functional Annotations and Deep Learning Models October 2023

American Society of Human Genetics, Washington, DC

Honors and Awards

Stellar Abstract Award - Harvard PQG Conference	November 2018
Cloud Computing Credits Award ($100,000) (Microsoft Azure)	May 2017
AMS Travel Grant - MRC (Computational and Applied Topology)	June 2011
NSF Travel Grant - MEGA 2011 Conference, Stockholm, Sweden	May 2011
NSF Travel Grant - Aspects of Moduli Conference, Pisa, Italy	June 2008
Department Undergraduate Award for Research	December 2003
PAMS College Undergraduate Award for Service	December 2003

Computational Skills

Programming Languages: C/C++, Python, Java

Databases: Hive, MySQL, PostgreSQL, HAWQ, Impala, Microsoft SQL Server, Amazon Aurora, Spark

Machine Learning Libraries: R, scikit-learn, Stan, PyTorch, Pyro, Tensorflow, Weights and Biases, Optuna, Hugging Face

Cloud Services: Amazon Web Services, Microsoft Azure, Google Cloud Platform