Salt Lake City, UT
ish@ihaque.org
ihaque.org
ImranSHaque
imranshaque
Updated 05 Jul 2024

Imran S. Haque, PhD

Positions

2023– **Recursion**, *SVP, AI and Digital Sciences*

2019–2023 **Recursion**, *VP, Data Science*

Head of AI/ML and computational sciences at category-defining AI-enabled therapeutics firm

- **Inception Labs**: (2024–) Founding leader of Inception Labs, Recursion's emerging technology R&D group. Hired, managed, and set technical vision for interdisciplinary group of machine learning scientists, engineers, and biologists to advance the leading edge of AI in biology.
- **Corporate**: Drafted and revised Recursion IPO S-1. Recursion technology/AI representative for BD and partnering, including partnership ($12B+) with Roche/Genentech. Technical search, diligence, and integration leader on M&A, including 2023 chemical technology acquisitions of Cyclica and Valence Discovery. Led organization through multinational expansion and through COVID-19 remote transition.
- **Functional**: Built and led 60+-person department (data science, machine learning/AI, computational chemistry, and computational biology) building autonomous discovery engines mapping biological and chemical space and applying them to drug discovery in fibrosis, oncology, neuroscience, and rare disease.
- **Technology**: Key leader defining vision and driving implementation of large-scale (bulk/single-cell) NGS, statistical genetics, and computational chemistry to advance discovery and validation pipelines. Drove data partnerships with Tempus and Helix in oncology and germline omics.
- **Therapeutics**: Co-lead on anti-COVID-19 discovery programs. Joint leadership team member on Bayer and Roche/Genentech partnerships defining computational discovery and ML strategy. Member of Recursion portfolio review committee overseeing entire Recursion discovery pipeline.

2021–2023 **Known Medicine**, *Scientific Advisor*

Member of SAB advising precision oncology startup from founding through acquisition by Pathos.

2016–2018 **Freenome**, *Chief Scientific Officer*

Head of R&D for venture-backed startup developing tests for early cancer detection.

- Hired and led multidisciplinary R&D team of 24+ machine learning scientists and engineers and molecular and computational biologists.
- Developed Freenome's research program around analyzing systemic and immune response to colorectal cancer. Presented results at major scientific and clinical conferences.
- Design, implementation, and scale-up of data and high-performance computing infrastructure for large-scale genomic machine learning.
- Investor relations through fundraising (including US$72M series A) and board of directors meetings.

2016 **Counsyl**, *VP, Scientific Affairs*

2013–2016 **Counsyl**, *Director, Research*

Guideline-changing research in diagnostics development and population-scale genomic discovery.

- Jointly led clinical development research to drive clinical guidelines and reimbursement, gaining guideline acceptance for pan-ethnic expanded carrier screening. Team presented 19 conference talks and 39 posters in 2015-16 and published 16 papers in 2015-17, including landmark paper in *JAMA*.
- Hired and led team of four integrating computational and molecular approaches to develop advanced genomics assays, including oncology liquid biopsy assays and best in class assays for inherited conditions.

2011–2013 **Counsyl**, *Sr. Software Engineer/Sr. Research Scientist*

Engineering team lead managing automated high-throughput clinical genetic testing laboratory.

Education

2006–2011 **PhD, Computer Science**, *Stanford University*

Large-scale machine learning for drug discovery. Scaled chemical machine learning algorithms to operate on large-scale (16M+ molecules) compound and assay databases, reducing computational cost by over 1 million-fold in both time (CPU) and space (storage).
Dissertation: Accelerating Chemical Similarity Search with GPUs and Metric Embeddings
Committee: Vijay Pande, Daphne Koller, Russ Altman

2002–2006 **BS, Hons., Electrical Engineering and Computer Science**, *UC Berkeley*

Coursework in software engineering and algorithms, probability theory, and computational biology.
Research in hardware design for error correcting codes and parallel programming languages.

Selected Invited Talks

Oct 2023 **Broad Institute Machine Learning in Drug Discovery Symposium**
video, slides Decoding Biology with Data Abundance

Mar 2023 **The Royal Society - Machine Learning and AI in Biological Science, Drug Discovery, and Medicine**
video, slides Mapping and navigating biology and chemistry with genome-scale imaging

Dec 2022 **Learning Meaningful Representations of Life (LMRL) @ NeurIPS 2022**
slides Biological Cartography: Building and Benchmarking Representations of Life

Dec 2021 **Learning Meaningful Representations of Life (LMRL) @ NeurIPS 2021**
video, slides Mapping Biology With a Unified Representation Space for Genomic and Chemical Perturbations to Enable Accelerated Drug Discovery

Apr 2021 **GPU Technology Conference (GTC)**
slides Zero to potential COVID-19 treatments in under 4 weeks with deep-learning driven drug screens

Apr 2019 **(Invited Keynote) Cancer Research UK 3rd Int'l Symp. on Oesophageal Cancer**
blog, slides We Are Legion: Statistics and Generalization from Cells to Populations

Oct 2018 **AACR Special Conf. on Convergence: AI, Big Data, and Prediction in Cancer**
blog, slides Making hay of needles: Connecting clinical and physical parameters in the search for early cancer

Jan 2017 **Society for Maternal-Fetal Medicine Annual Meeting**
slides 1 in 550: Using 346,790 expanded carrier screens to estimate the risk of Mendelian conditions

Selected Posters

2024 PDF Neumann C, Shankaran H, Nadella K, ..., **Haque I**, Donnella H, Cuccarese M, Evangelista M. Poster at Am Assoc Cancer Res (AACR) 2024.
Phenomics-enabled discovery and optimization of small-molecule RBM39 degraders as an alternative to CDK12 targeting in high-grade serous ovarian cancer (HGSOC)

2023 PDF Rudnick J, Nadella K, Rajan M, ..., **Haque I**, Donnella H, Cuccarese M, Evangelista M. Poster at Am Assoc Cancer Res (AACR) 2023.
A Phenomics Platform Combining Imaging and Artificial Intelligence for Rapid Validation and Advancement of Novel Oncology Targets

2022 PDF Bhandari A, Cuccarese MF, Fales K, ..., **Haque I**, Alfa R, Rinaldi J. Poster at Am Assoc Cancer Res (AACR) 2022.
Identification and optimization of novel small molecule modulators of immune checkpoint resistance with a unified representation space for genomic and chemical perturbations

2018 PDF Niehaus K, Wan N, Weinberg D, ..., **Haque IS**, Putcha G. Poster at Am Coll Gastroenterol (ACG) 2018.
Early-stage colorectal cancer detection using artificial intelligence and whole-genome sequencing of cell-free DNA

PDF Delubac D, Ariazi E, Berliner J, ..., **Haque IS**. Poster at Am Assoc Cancer Res (AACR) 2018.
Multi-analyte profiling reveals relationships among circulating biomarkers in colorectal cancer

2017 PDF Davison D, Kaseniit KE, **Haque IS**. Poster at Am Coll Med Genet (ACMG) 2017. *Top Rated Abstract*.
Duplication tag SNP g.27134T>G should not be considered diagnostic of SMA carrier status

Publications

Google Scholar statistics: https://scholar.google.com/citations?user=Cp_wDj4AAAAJ

2024 URL Lazar NH, Celik S, Chen L, ..., **Haque IS**. *Nat Genet*: 1–12. (2024)
High-resolution genome-wide mapping of chromosome-arm-scale truncations induced by CRISPR-Cas9 editing

2023 URL Fay MM, Kraus O, Victors M, ..., **Haque IS**, Mabey B. *bioRxiv*: 2023.02.07.527350. (2023)
RxRx3: Phenomics Map of Biology

PDF Sypetkowski M, Rezanejad M, Saberian S, ..., **Haque I**, Earnshaw B. In proceedings of *CVPR 2023: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVMI workshop)*. 4284–4293. (2023)
RxRx1: A dataset for evaluating experimental batch correction methods

2022 URL Celik S, Huetter J, Melo-Carlos S, ..., **Haque I**. *bioRxiv*: 2022.12.09.519400. (2022)
Biological Cartography: Building and Benchmarking Representations of Life

2021 URL Koh PW, Sagawa S, Marklund H, ..., **Haque IS**, Beery S, Leskovec J, Kundaje A, Pierson E, Levine S, Finn C, Liang P. *Proc 38th Intl Conf Mach Learning (ICML)* 139:5637–5664. (2021)
WILDS: A Benchmark of in-the-Wild Distribution Shifts

URL **Haque IS**. *Nat Biomed Eng* 5:490–492. (2021)
Enhanced DNA libraries for methylation analysis (News and Views)

2020 URL Cuccarese MF, Earnshaw BA, Heiser K, ..., **Haque IS**, Chong YT, Gibson CC. *bioRxiv*: 2020.08.02.233064v2. (2020)
Functional immune mapping with deep-learning enabled phenomics applied to immunomodulatory and COVID-19 drug discovery

URL Heiser K, McLean PF, Davis CT, ..., **Haque IS**, Low AS, Gibson CC. *bioRxiv*: 2020.04.21.054387v1. (2020)
Identification of potential treatments for COVID-19 through artificial intelligence-enabled phenomic analysis of human cells infected with SARS-CoV-2

DOI Kaseniit KE, **Haque IS**, Goldberg JD, Shulman LP, Muzzey D. *Genet Med* 22:1694–1702. (2020)
Genetic ancestry analysis on >93,000 individuals undergoing expanded carrier screening reveals limitations of ethnicity-based medical guidelines

2019 DOI Cecchi A, Vengoechea ES, Kaseniit KE, ..., **Haque IS**, Moyer K, Page PZ, Muzzey D, Grinzaid KA. *Molec Genet Genom Med* 7:e836. (2019)
Screening for Tay-Sachs disease carriers by full-exon sequencing with novel variant interpretation outperforms enzyme testing in a pan-ethnic cohort

URL Liu Y, Liu T, Weinberg DE, ..., **Haque IS**. *bioRxiv*: 564773. (2019)
Spatial co-fragmentation pattern of cell-free DNA recapitulates in vivo chromatin organization and identifies tissues-of-origin

DOI Heitzer E, **Haque IS**, Roberts CE, Speicher MR. *Nat Rev Genet* 20:71–88. (2019)
Current and future perspectives of liquid biopsies in genomics-driven oncology

URL Wan N, Weinberg D, Liu T, ..., **Haque IS**. *BMC Cancer* 19:832. (2019)
Machine learning enables detection of early-stage cancer by whole-genome sequencing of plasma cell-free DNA

2018 URL Manghnani K, Drake A, Wan N, **Haque IS**. *ML4H workshop at NeurIPS 2018*: arXiv cs.LG:1812.03188. (2018)
METCC: METric learning for Confounder Control: Making distance matter in high-dimensional biological analysis

DOI Hogan GJ, Vysotskaia VS, Beauchamp KA, ..., **Haque IS**, Mar-Heyming R, Kang HP, Muzzey D. *Clin Chem* 64(7):1063–1073. (2018)
Validation of an Expanded Carrier Screen that Optimizes Sensitivity via Full-Exon Sequencing and Panel-wide Copy Number Variant Identification

DOI Beauchamp KA, Muzzey D, Wong KK, ..., **Haque IS**. *Genet Med* 20(1):55–63. (2018)
Systematic Design and Comparison of Expanded Carrier Screening Panels

DOI Ghiossi C, Goldberg JD, **Haque IS**, Lazarin GA, Wong KK. *J Genet Counsel* 27:616–625. (2018)
Clinical Utility of Expanded Carrier Screening: Reproductive Behaviors of At-Risk Couples

2017 DOI **Haque IS**, Elemento O. *bioRxiv*: 237578. (2017)
Challenges in Using ctDNA to Achieve Early Detection of Cancer

DOI Artieri CG, Haverty C, Evans EA, ..., **Haque IS**, Yaron Y, Muzzey D. *Prenat Diagn* 37(5):482–490. (2017)
Noninvasive Prenatal Screening at Low Fetal Fraction: Comparing Whole-Genome Sequencing and Single-Nucleotide Polymorphism Methods

DOI Vysotskaia VS, Hogan GJ, Gould GM, ..., **Haque IS**. *PeerJ* 5:e3046. (2017)
Development and validation of a 36-gene sequencing assay for hereditary cancer risk assessment

DOI Amorim CE, Gao Z, Baker Z, ..., **Haque IS**, Pickrell J, Przeworski M. *PLoS Genetics* 13(9):e1006915. (2017)
The population genetics of human disease: the case of recessive, lethal mutations

DOI Lazarin GA, **Haque IS**, Evans EA, Goldberg JD. *Prenat Diagn* 37(4):350–355. (2017)
Smith-Lemli-Opitz syndrome carrier frequency and estimates of in utero mortality rates

2016 DOI **Haque IS**, Lazarin GA, Kang HP, Evans EA, Goldberg JD, Wapner RJ. *JAMA* 316(7):734–742. (2016)
Modeled Fetal Risk of Genetic Diseases Identified by Expanded Carrier Screening

DOI Kang HP, Maguire JR, Chu CS, **Haque IS**, Lai H, Mar-Heyming R, Ready K, Vysotskaia VS, Evans EA. *PeerJ* 4:e2162. (2016)
Design and validation of a next generation sequencing assay for hereditary *BRCA1* and *BRCA2* mutation testing

DOI Mehta N, Lazarin GA, Spiegel E, ..., **Haque IS**, Wapner R. *Genet Test Molec Biomarker* 20(9):504–509. (2016)
Tay-Sachs Carrier Screening by Enzyme and Molecular Analyses in the New York City Minority Population

DOI Kaseniit KE, Theilmann MR, Robertson A, ..., **Haque IS**. *Clin Chem* 62(10):1401–1408. (2016)
Group Testing Approach for Trinucleotide Repeat Expansion Disorder Screening

DOI Lazarin GA, **Haque IS**. *Semin Perinatol* 40(1):29–34. (2016)
Expanded carrier screening: A review of early implementation and literature

2015 DOI Vikram S, Rasmussen MD, Evans EA, **Haque IS**. *bioRxiv*: 021527. (2015)
SSCM: A method to analyze and predict the pathogenicity of sequence variants

2014 DOI Lazarin GA, Hawthorne F, Collins NS, ..., **Haque IS**. *PLoS One*: e114391. (2014)
Systematic Classification of Disease Severity for Evaluation of Expanded Carrier Screening Panels

DOI **Haque IS**, Beauchamp KA, Pande VS. *bioRxiv*: 008631. (2014)
A Fast 3 x N Matrix Multiply Routine for Calculation of Protein RMSD

2013 DOI Kearnes SM, **Haque IS**, Pande VS. *J Chem Inf Model* 54(1):5–15. (2013)
SCISSORS: Practical Considerations

DOI Lazarin GA, **Haque IS**, Nazareth S, Iori K, Patterson AS, Jacobson JL, Marshall JR, Seltzer WK, Patrizio P, Evans EA, et al.. *Genet Med* 15(3):178–186. (2013)
An empirical estimate of carrier frequencies for 400+ causal Mendelian variants: results from an ethnically diverse clinical sample of 23,453 individuals

2011 DOI Ready K, **Haque IS**, Srinivasan BS, Marshall JR. *Fertil Steril*: 407–413. (2011)
Knowledge and attitudes regarding expanded genetic carrier screening among women's healthcare providers

DOI **Haque IS**, Pande VS, Walters WP. *J Chem Inf Model* 51(9):2345–2351. (2011)
Anatomy of high-performance 2D similarity calculations

DOI **Haque IS**, Pande VS. *J Chem Inf Model* 51(9):2248–2253. (2011)
Error bounds on the SCISSORS approximation method

Haque IS, Pande VS. In *GPU Computing Gems: Emerald Edition (Ed: W.-M. W. Hwu)*. (2011)
Large-Scale Chemical Informatics on GPUs

DOI Beauchamp KA, Bowman GR, Lane TJ, ..., **Haque IS**, Pande VS. *J Chem Theor Comput*: 3412–3419. (2011)
MSMBuilder2: Modeling Conformational Dynamics at the Picosecond to Millisecond Scale

DOI Pronk S, Larsson P, Pouya I, ..., **Haque IS**, Beauchamp K, Hess B, Pande VS, Kasson PM, Lindahl E. In proceedings of *SC11: 2011 Intl Conf High Perf Comput, Network, Storage and Analysis*. 60. (2011)
Copernicus: A new paradigm for parallel adaptive molecular dynamics

2010 DOI **Haque IS**, Pande VS. In proceedings of *CCGrid 2010: 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing*. 691–696. (2010)
Hard data on soft errors: A large-scale assessment of real-world error rates in GPGPU

DOI **Haque IS**, Pande VS, Walters WP. *J Chem Inf Model* 50(4):560–564. (2010)
SIML: a fast SIMD algorithm for calculating LINGO chemical similarities on GPUs and CPUs

DOI **Haque IS**, Pande VS. *J Chem Inf Model* 50(6):1075–1088. (2010)
SCISSORS: a linear-algebraical technique to rapidly approximate chemical similarities

DOI **Haque IS**, Pande VS. *J Comput Chem* 31(1):117–132. (2010)
PAPER — accelerating parallel evaluations of ROCS

DOI Ponder JW, Wu C, Ren P, ..., **Haque I**, Mobley DL, Lambrecht DS, DiStasio Jr RA, et al.. *J Phys Chem B* 114(8):2549–2564. (2010)
Current status of the AMOEBA polarizable force field

2006 DOI Pitera JW, **Haque I**, Swope WC. *J Chem Phys* 124:141102. (2006)
Absence of reptation in the high-temperature folding of the trpzip2 β-hairpin peptide

Patents and Published Patent Applications

2022 URL Jensen J, Dahlem T, Hugo S, ..., **Haque I**, Lazar N, Gardner A, Banowsky B, Allen A. *US Patent App*: 17/585,660. (2022)
CRISPR guide selection

URL Fogelson BMF, McLean P, **Haque I**, Saunders M, Fish E, Baker C, Rodriguez Vera JS. *US Patent App*: 2022/0027795 A1. (2022)
Techniques for training a classifier to detect executional artifacts in microwell plates

URL Fogelson BMF, McLean P, **Haque I**, Saunders M, Fish E, Baker C, Rodriguez Vera JS. *US Patent App*: 2022/0028061 A1. (2022)
Techniques for analyzing and detecting executional artifacts in microwell plates

2019 URL Drake A, Delubac D, Niehaus K, ..., **Haque I**, Liu T, Wan N, Kannan A, White B. *US Patent* (granted 2023):11,681,953. (2019)
Machine learning implementation for multi-analyte assay of biological samples

URL Liu Y, Delubac D, **Haque IS**. *WIPO Patent App*: WO 2019/191649 A1. (2019)
Methods and systems for analyzing microbiota

URL Delubac D, **Haque IS**, Singer M. *WIPO Patent App*: WO 2019/147663 A1. (2019)
Methods and systems for abnormality detection in the patterns of nucleic acids

2018 URL Kaseniit KE, **Haque IS**. *US Patent* (granted 2022):11,527,304 B2. (2018)
Systems and methods for inferring genetic ancestry from low-coverage genomic data

URL Gould GM, Wang X, Grauman PV, ..., **Haque IS**, Evans EA, Haas KR. *WIPO Patent App*: WO 2018/085779 A1. (2018)
Methods for assessing genetic variant screen performance

URL Haas KR, Wang X, Kang H, ..., **Haque IS**. *WIPO Patent App*: WO 2018/170443 A1. (2018)
Multi-dimensional sample-dependent and batch-dependent quality control

URL Haas KR, Muzzey D, Hogan GJ, ..., **Haque IS**, Kang H. *WIPO Patent App*: WO 2018/119322 A1. (2018)
Variant call processing system and method for review of called variants and quality control metrics

URL Muzzey D, Artieri CG, Evans EA, **Haque IS**. *US Patent* (granted 2023):11,854,666. (2018)
Noninvasive prenatal screening using dynamic iterative depth optimization

2017 URL Evans EA, **Haque IS**, Beauchamp K, Chu C, Artieri CG, Welker N. *US Patent* (granted 2023):11,708,574. (2017)
Nucleic acid sequencing adapters and uses thereof

URL **Haque IS**, Maguire JR, Chu C, Evans EA. *US Patent App*: 15/587,811. (2017)
Noninvasive prenatal diagnostic methods

URL Kaseniit KE, Theilmann MR, Robertson AD, ..., **Haque IS**. *US Patent App*: 15/488,129. (2017)
Group testing approach for a genetic screening assay

URL Maguire JR, Chu C, **Haque IS**, Evans EA, Welker N. *US Patent* (granted 2020):10,597,717. (2017)
Combinatorial DNA screening

2016 URL **Haque IS**, Evans EA, Vikram SM, Rasmussen MD. *US Patent App*: 15/189,957. (2016)
Methods of predicting pathogenicity of genetic sequence variants

URL Gibiansky AL, **Haque IS**, Maguire JR, Robertson AD. *US Patent App*: 14/884,656. (2016)
Variant caller

2015 URL Patterson AS, **Haque IS**, Evans EA, Chu C. *US Patent* (granted 2019):10,497,463. (2015)
Automated nucleic acid repeat count calling methods

2011 URL **Haque I**, Pande V. *US Patent* (granted 2014):8,706,427. (2011)
Method for rapidly approximating similarities

Open-source contributions

Primary author **PeloMon**
Reverse engineering of Peloton data interface to build custom Bluetooth cadence/power meter

Primary author **xorshift**
High-performance vectorized RNGs for Python

Primary author **MemtestG80/MemtestCL**
Tools to check GPU memory for errors

Primary author **SIML**
High-performance LINGO similarity library for GPU and CPU

Primary author **PAPER**
GPU-based 3-D shape comparison for chemical similarity

Primary author **IRMSD**
Optimal SSE2/SSE3-based code for structural RMSD computation