San Francisco, CA
ish@ihaque.org
ihaque.org
ImranSHaque
ihaque
imranshaque
Updated 10 May 2019

Imran S. Haque, PhD

Positions **Publications** **Patents** **Posters**
Invited Talks **Awards** **Open-Source** **Teaching**

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.

Positions

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

Head of R&D for venture-backed startup developing tests for cancer detection based on next-generation sequencing, machine learning, and functional genomics.

- Hired and led multidisciplinary R&D team of 24+ machine learning scientists and engineers and molecular and computational biologists. Managed both individual contributors and managers.
- Developed Freenome's research program around analyzing systemic and immune response to colorectal cancer. Presented results at major scientific and clinical conferences; publications currently under review and several patent applications filed (currently unpublished).
- Collaborated with engineering team on design, implementation, and scale-up of Freenome 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*

Founded and set direction for Counsyl's research program, focusing on technology development and population-scale genomic data mining.

- Hired and led team of four scientists integrating computational and molecular approaches to develop advanced genomics assays for future Counsyl products, including liquid biopsy assays for cancer therapy selection and recurrence monitoring as well as low cost (best in class by 10x) assays for inherited conditions. Co-inventor on twelve published (and additional unpublished) patent applications.
- Jointly led clinical development research program leveraging Counsyl's genetic screen database, focused on driving guidelines acceptance and reimbursement. Successfully gained 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* on 346,790 individuals.

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

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

- Developed laboratory information management system (LIMS) for automated high-throughput wetlab, including database backend design and workflow management frontend.
- Worked daily with CTO and CSO on projects ranging from team scaling to laboratory methods and process innovation.
- Developed machine-learning based automated assay calling for PCR and genotyping assays, increasing throughput and reliability to reduce genotype call error rate by 1000-fold.
- Managed summer engineering and science internship program for five years.
- Revamped engineering hiring process and managed university recruiting, helping grow engineering team from 10 to 30. Hired and trained replacement team lead to enable creation of research department.

2009 **Vertex Pharmaceuticals**, *Special Project*

Developed algorithms leveraging GPUs and advanced analytics to empower medicinal chemists to rapidly explore new molecular hypotheses and synthetic strategies. Accelerated chemistry search from overnight to interactive; published results in peer-reviewed journal.

Publications

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

2019 Accepted Cecchi A, Vengoechea ES, Kaseniit KE, ..., **Haque IS**, Moyer K, Page PZ, Muzzey D, Grinzaid KA. Accepted at *Molec Genet Genom Med*. (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

2018 URL Wan N, Weinberg D, Liu T, ..., **Haque IS**. *bioRxiv*: 478065. (2018)

Machine learning enables detection of early-stage cancer by whole-genome sequencing of plasma cell-free DNA

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

2018 URL Kaseniit KE, **Haque IS**. *WIPO Patent App*: WO 2018/144135 A1. (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 App*: 15/720,351. (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 App*: 15/619,078. (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 App*: 15/465,553. (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 App*: 14/540,334. (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

Selected Posters

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

PDF Kaseniit KE, Karczewski K, **Haque IS**. Poster at Am Coll Med Genet (ACMG) 2017. *Top Rated Abstract*.
ClinVar submitter list leaderboard obscures extensive variation and bias in submission types

2016 PDF **Haque IS**, Haverty C, Goldberg JD, Evans EA. Poster at Am Soc Hum Genet (ASHG) 2016. *Top 10% Abstract*.
Computing confidence intervals on positive predictive value for non-invasive prenatal screening

Selected Invited Talks

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

Mar 2019 **Molecular Medicine Tri-Con**
(How to fix) the very reasonable ineffectiveness of machine learning in biomarker discovery

Mar 2019 **OpenEye CUP XIX**
blog, slides Thanks, I Hate It: Why your biological machine learning model probably won't work & what to do about it

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

Oct 2018 **Early Detection of Cancer**
slides The Reasonable Ineffectiveness of Biological Data

Jul 2018 **DeepChem User Group Meeting**
slides (What to do) when gradient descent digs too deep, and too greedily

Oct 2017 **(Invited Keynote) Cancer Crosslinks**
slides Embracing heterogeneity: The freenome, information, and early disease detection

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

Aug 2016 **2016 AGBT Precision Health Meeting**
slides Overcoming artificial selection to achieve the promise of inherited cancer screening

Apr 2015 **Stanford ChildX 2015**
video Can/Do: the disconnect between what we can do and what we do in perinatal precision medicine

Nov 2013 **PyData 2013**
slides Beyond the dict: Python tools for data wrangling

May 2010 **Resilience Workshop @ CCGrid 2010**
slides Hard Data on Soft Errors: A Global-Scale Assessment of GPGPU Memory Soft Error Rates

Mar 2010 **OpenEye CUP XI**
slides LINGOs and GPUs

Fellowships and Awards

2006–2009 National Science Foundation Graduate Research Fellowship

2005–2006 UC Berkeley Alumni Scholarship

2004–2006 SRCEA Undergraduate Research Assistant Scholarship

2002–2006 UC Berkeley Chancellor's Scholarship

2002–2006 Robert Byrd Honors Scholarship

Open-source contributions

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

Teaching experience

Winter 2011 Teaching Assistant, Stanford CS 109, Introduction to Probability for Computer Scientists, Mehran Sahami

Fall 2010 Teaching Assistant, Stanford CS 148, Introduction to Computer Graphics and Imaging, Pat Hanrahan

Spring 2010 Teaching Assistant, Stanford BMI214/CS 274, Representations and Algorithms for Computational Molecular Biology, Russ Altman