

Rohit Singh

RESEARCH INTERESTS: As a computational biologist, I aim to speed up drug discovery using insights generated by integrating biological data from diverse sources. I envision that drug discovery can be made many times faster and cheaper using the vast corpus of data being made available by the genome sequencing revolution. To realize this vision will require computational frameworks to integrate the data across modalities, technologies, and species, unlocking speed-ups in all aspects of the discovery pipeline: identifying the root causes of disease, shortlisting candidate drugs, and minimizing the risk of clinical-trial failure. Building such integrative frameworks is the focus of my research.

EDUCATION

- 2011 **Massachusetts Institute of Technology**
Ph.D., Computer Science (on non-resident status: 2007-2011) (Advisor: Bonnie Berger)
- PhD Thesis: Algorithms for the analysis of protein interaction networks.
- 2002 **Stanford University**
M.S., Computer Science (Advisor: Jean-Claude Latombe)
- MS Thesis: An efficient algorithm for identifying 3-D structural patterns in protein structures.
- 2000 **Indian Institute of Technology, Kanpur**
B.Tech., Computer Science & Engineering

HONORS & AWARDS

- **RECOMB Test of Time Award** (2019): RECOMB is one of the two top-ranked conferences in the field of computational biology. The award recognized the RECOMB 2007 paper that presented some of my PhD research.
- George M. Sprowls Award for the **Best PhD Theses in Computer Science at MIT** (2011).
- Christopher Stephenson Memorial Award for **Best Masters Research in Computer Science at Stanford** (2002).
- World Finalist, ACM Programming Contest (represented IIT Kanpur, the only Indian team) (2000).
- Dean's List, IIT Kanpur (1997).
- Offered admission to the IITs as well as AIIMS (India's premier medical school), ranking 17th (IIT) and 26th (AIIMS) nationwide in the respective entrance tests; **I was the only such admittee India-wide** (1996).
- National Talent Search Scholar, India (1994).

EMPLOYMENT

- 2019-curr. **MIT Computer Science and Artificial Intelligence Lab.** Cambridge, MA
Research Scientist
- Interpretable integration of multimodal single-cell data for understanding cell-level variability and gene regulatory mechanisms.
 - Apply representations learned from millions of protein sequences to predict protein-protein and protein-drug interactions.
- 2012-2019 **Tech Square Trading** Boston, MA
Co-founder, Chief Executive Officer & Portfolio Manager
Tech Square Trading (TST) was a quantitative investment fund, trading stocks algorithmically.
- Researched statistical and machine learning methods for integrating financial data from diverse sources to make accurate stock price forecasts.

- Led a high-caliber, collegial team of quantitative researchers of PhD and Masters-level graduates from MIT, Columbia and Harvard.
- Owned final responsibility for all aspects of the business, including recruitment, budgeting, planning, and marketing.

2011-2018 **Independent Researcher**

After I left the MIT campus, I continued to do research in computational biology (see publications below). I mentored junior graduate students and worked to build community resources around the algorithms we had previously designed.

2008-2011 **Cubist Systematic** *Research Analyst (Quantitative Investing)* **New York, NY**

2007-2008 **Merrill Lynch** *Associate (Quantitative Investing)* **New York, NY**

- Applied statistical and machine learning tools to generate stock trading signals.

2003-2007 **Prof. Bonnie Berger's Research Group, MIT** **Cambridge, MA**

- PhD research on algorithms for analyzing, predicting and aligning protein interaction networks.

2002-2003 **Accelrys Inc. (now part of Dassault Systemes)** **San Diego, CA**

Software Engineer (Bioinformatics Applications Group, Life Sciences R&D Divn.)

- Co-designed the next release of Discovery Studio GCG's, the best-selling bioinformatics software.

PUBLICATIONS

Below, "*" denotes shared first authorship, "\$\$" indicates 300 or more citations, "\$" indicates 75 or more citations (as counted by Google Scholar in Oct 2021).

Manuscripts under review

1. [Singh R*](#), Devkota K*, Sledzieski S, et al. *Combining network and sequence information for PPI prediction*.
2. [Singh R](#), Berger B. *Deciphering the species-level structure of topologically associating domains* ([preprint](#)).
3. Kumar L, Brenner N, Sledzieski S, et al. *Transfer of Knowledge from Model Organisms to Evolutionarily Distant Non-Model Organisms: The Coral Pocillopora damicornis Membrane Signaling Receptome*.

Refereed Conference Publications [RECOMB: 2, PSB: 4, ICML: 1, ISMB: 1 (+ highlights presentation), NeurIPS MLSB: 1]

In computational biology, RECOMB and ISMB have an acceptance rate < 20% and PSB has an acceptance rate < 30%. In machine learning, ICML is a top-tier conference, with an acceptance rate of ~20-25% .

1. Sledzieski S*, [Singh R*](#), Cowen L, Berger B. *Adapting protein language models to DTI prediction*. NeurIPS Workshop on Machine Learning in Structural Biology (**NeurIPS MLSB**), 2021 (*to appear*).
2. Sledzieski S*, [Singh R*](#), Cowen L, Berger B. *Sequence-based prediction of protein-protein interactions: a structure-aware interpretable deep learning model*. Int'l Conf on Research in Mol. Biology (**RECOMB**), 2021.

Pre-2019

3. \$ [Singh R](#), Xu J, Berger B. *Global alignment of multiple protein interaction networks*. Pac Symp Biocomput. (**PSB**) 2008:303-14.
4. [Singh R](#), Berger B. *Influence Flow: Integrating Pathway-specific RNAi data and Protein Interaction Data*. Int'l Symp of Intelligent Sys in Mol Bio (**ISMB**) 2007, PLoS Track.
5. \$\$ [Singh R](#), Xu J, Berger B. *Pairwise Global Alignment of Protein Interaction Networks by Matching Neighborhood Topology*. Proc of Int'l Conf on Research in Comp Mol Bio (**RECOMB**) 2007.

- a. This paper received the **Test of Time Award at RECOMB 2019**. It introduced the *global* network alignment problem for protein interaction networks, now a fundamental task in biological network analysis. Our IsoRank algorithm pioneered the use of spectral theoretic techniques for such analyses.
6. Sontag D*, Singh R*, Berger B. *Probabilistic modeling of systematic errors in two-hybrid experiments*. Pac Symp Biocomput (**PSB**) 2007.
7. **§** Singh R*, Xu J*, Berger B. *Struct2net: integrating structure into protein-protein interaction prediction*. Pac Symp Biocomput (**PSB**) 2006.
8. Singh R, Berger B. *Chaintweak: sampling from the neighbourhood of a protein conformation*. Pac Symp Biocomput (**PSB**) 2005:52-63.
9. Singh R, Palmer N, Gifford D, Berger B, Bar-Joseph Z. *Active Learning for Sampling in Time-Series Experiments With Application to Gene Expression Analysis*. Proc. of the 22nd Int'l Conf on Machine Learning (**ICML**), 2005.
10. Singh R, Saha M. *Identifying structural motifs in proteins*. Pac Symp Biocomput (**PSB**) 2003:228-39.
11. **§** Deshpande H, Nam U, Singh R. *Classification of Music Signals in the Visual Domain*. Proc of Digital Audio Effects Conf (**COST-G6**), 2001.
12. Lahoti A*, Singh R*, Mukerjee A. *Multidimensional interval algebra for describing symmetric block layouts*. Proc of the 3rd Int'l Conf on Graphics Recognition (**GREC**) 1999.

Refereed Journal Publications [Nat. Methods, PNAS, Genome Bio., N.A.R., Cell Systems, Sci. Signaling, Bioinformatics]

13. Sledzieski S*, Singh R*, Cowen L, Berger B. *D-SCRIPT translates genome to phenome with sequence-based, structure-aware, genome-scale predictions of protein-protein interactions*. **Cell Systems**, 2021.
14. Singh R*, Hie B*, Narayan A, Berger B. *Schema: metric learning enables interpretable synthesis of heterogeneous single-cell modalities*. **Genome Biology**, 2021; 22:131 .

Pre-2019

15. **§** Park D*, Singh R*, Baym M, Liao CS, Berger B. *IsoBase: a database of functionally related proteins across PPI networks*. **Nucleic Acids Res**. 2011 Jan;39 (Database issue, top-cited issue NAR publishes):D295-300.
16. **§** Friedman AA, Tucker G, Singh R, et al. *Proteomic and functional genomic landscape of receptor tyrosine kinase and Ras to extracellular signal-regulated kinase signaling*. **Science Signaling** 2011; 4(196).
17. Hosur R, Singh R, Berger B. *Sparse estimation for structural variability*. **Algorithms Mol Biol**. 2011; 6: 12.
18. **§** Singh R*, Park D*, Xu J*, Hosur R, Berger B. *Struct2Net: a web service to predict protein-protein interactions using a structure-based approach*. **Nucleic Acids Res**. 2010 Jul;38 (Web Server issue):W508-15.
19. **§§** Liao CS, Lu K, Baym M, Singh R, Berger B. *IsoRankN: spectral methods for global alignment of multiple protein networks*. **Bioinformatics**. 2009 June 15; 25(12): i253-i258.
20. Kaplow IM*, Singh R*, Friedman A, Bakal C, Perrimon N, Berger B. *RNAiCut: automated detection of significant genes from functional genomic screens*. **Nature Methods**. 2009 Jul;6(7):476-7.
21. **§§** Singh R, Xu J, Berger B. *Global alignment of multiple protein interaction networks with application to functional orthology detection*. **Proc Natl Acad Sci**. 2008 Sep 2;105(35):12763-8.
22. Sterner B, Singh R, Berger B. *Predicting and annotating catalytic residues: an information theoretic approach*. **J Comput Biol**. 2007 Oct;14(8):1058-7.

RESEARCH PRESENTATIONS

- *Invited talks:*
 - Columbia Univ (Guest lecture on protein interaction prediction) (2021)
 - MIT Bioinformatics seminar (2021)
 - Pacific Symp on Biocomputing: Workshop on Bioinformatics of Corals (2021)
- *Computational Biology Conferences:*
 - RECOMB:** Int'l Conf on Research in Comp Mol Bio (< 20% acceptance rate)
 - ISMB:** Int'l Symp of Intelligent Sys in Mol Bio (< 20% acceptance rate)
 - PSB:** Pacific Symp on Biocomputing (< 30% acceptance rate)

- ISMB (2021): Highlights track presentation of my Genome Biology publication.
- RECOMB (2007)
- ISMB (2007)
- PSB (2007)
- PSB (2006)
- PSB (2005)
- PSB (2003)
- *Other Conferences:*
 - Fly Cell Atlas Meeting (2021)
 - Am. Soc. of Human Genetics, Satellite Conf on Spatial Genomics (2020)
 - Int'l Conf on Machine Learning, **ICML** (2005)
 - Int'l Conf on Graphics Recognition, GREC (1999)

INDUSTRY TALKS AND PRESENTATIONS

- *Invited Talks/Panels:*
 - Panel Discussion on Quantitative Investing (organized by MIT Sloan & Bank of America)
 - Guest Speaker (MIT Sloan course on Quantitative Investing)
- *Investment Management Conferences:* Opal (2018), MFA (2018, 2019) and CONTEXT (2019).

MENTORSHIP (ACADEMIA)

I have mentored 12 graduate/undergraduate students in my roles as Research Scientist (now) and senior graduate student (previously). All students below are/were at MIT:

- Graduate students (2019-): Samuel Sledzieski, Alex Wu.
- Undergraduate students (2019-): Chiho Im, Binwei Yan, Stuti Khandavala, Jesse Yang, Ishan Ganguli.
- Graduate students (2006-2013): Daniel Park, Raghavendra Hosur, Sumaiya Nazeen.
- Undergraduate students (2006-2013): Beckett Sterner, Irene Kaplow.

MENTORSHIP (INDUSTRY)

I hired and managed 5 MS/PhD-level researchers, 1 senior non-research staff and ~30 high-caliber interns, as the Chief Executive Officer of Tech Square Trading. While names have been omitted for privacy reasons, references to verify the following are available upon request:

- The firm's Chief Financial Officer was previously an Associate Provost of Research Administration at two Boston-area universities.
- Two researchers with PhD in Applied Mathematics, MIT.
- PhD candidate in Computer Science, Univ of North Carolina, Chapel Hill.
- Multiple employees with MS/BS in Mathematics and Computer Science.
- Mentored approximately 30 interns, primarily undergraduate and graduate students from MIT and Harvard Univ.

PATENTS

- Singh R. and Berger B., *Method for Identifying Network Similarity by Neighborhood Matching*, USPTO # 8000262.
- Chefitz A. and Singh R., *Self-administered, non-invasive, transcutaneous viral detector*, (App. # 17/079,649).

SOFTWARE AND COMMUNITY RESOURCES

- *D-SCRIPT*: D-SCRIPT is a deep learning method for predicting a physical interaction between two proteins given just their sequences (<http://dscript.csail.mit.edu>).

- Schema: Schema is a metric learning method for integrating heterogeneous data modalities (<http://schema.csail.mit.edu>).
- IsoRank and IsoRank-N: executables for the IsoRank/IsoRankN programs, along with test data (<http://isorank.csail.mit.edu>).
- IsoBase: a web-database of functional ortholog predictions, using IsoRank/IsoRankN (<http://isobase.csail.mit.edu>).
- Struct2Net: a web-service for predicting interaction between two proteins, given just sequence data (<http://struct2net.csail.mit.edu>).
- ChainTweak: a program for sampling the neighborhood of a given protein's backbone structure (<http://theory.csail.mit.edu/chaintweak/>).
- RNAiCut: a web-service for choosing the right cut-offs in the results from RNAi gene-perturbation assays (<http://rnaicut.csail.mit.edu>).

TEACHING

- Designed and taught a 4-week course on The Business of Quant Investing (Jan 2020 and Jan 2021, during MIT's Interim Activities Period). The class received positive reviews and a strong enough interest that it needed to be moved to a bigger room.
- Teaching Assistant for 6.867, MIT's main graduate-level course on Machine Learning (2005).
- Designed and taught a 12-week, 3-hours-per-week class on Classical Mechanics for local high school students, as part of MIT's Educational Studies Program, an outreach program (2003).

SERVICE

- **Session co-organizer** for the Special Session on "Genotype to Phenotype in Model and Non-Model Organisms" at the Great Lakes Bioinformatics Conference (2021).
- Have **reviewed manuscripts** for 8 journals and conferences incl. Nature Biotechnology, RECOMB and ICML.

GRANTS: PROPOSAL AND PROGRESS REPORT PREPARATION

Below, "PI" stands for Principal Investigator.

- 2021** Co-Investigator: **NSF** Harnessing the Data Revolution (*under review*).
Focus: ORCHARDS: Open Reef Collaborative Habitat Advancing Reproducible Data Science
PIs: This is a multi-institute proposal spanning eight universities
- 2021** Supporting role: **NIH** MIRA #13155717 (*funded*).
Focus: Manifold Representations and Active Learning for 21st Century Biology
PI: Prof. Bonnie Berger (MIT)
- 2021** Supporting role: **Sanofi Inc.** (*funded*). I lead the implementation of this project.
Focus: Learning the Language of Protein Interactions for Therapeutic Discovery and Design
PIs: Prof. Bonnie Berger (MIT) and Prof. Bryan Bryson (MIT)
- 2021** "Key Person" designation: **NIH** (*under preparation*)
Focus: Systematic Characterization of Genomic Variation on Genome Function and Phenotype
PIs: Dr. Dou Yu (MIT), Prof. Tim Lu (MIT), Prof. Bonnie Berger (MIT)
- 2021** Supporting role: **NIH** (*under preparation*).
Focus: Resources for Functional Studies in *Drosophila*
PIs: Prof. Norbert Perrimon (Harvard Medical School)
- 2019-21** **NIH** R01-GM081871 (*progress report preparation*)
Focus: Structure Based Prediction of the Interactome
PI: Prof. Bonnie Berger (MIT)
- 2019-21** **NIH** R01-HG010959 (*progress report preparation*)
Focus: Privacy-preserving Genomic Medicine at Scale
PI: Prof. Bonnie Berger (MIT)

REFERENCES

1. **Bonnie Berger**, Massachusetts Institute of Technology, bab@mit.edu, 617-253-1827
2. **Norbert Perrimon**, Harvard Medical School, perrimon@receptor.med.harvard.edu, 617-432-7672
3. **Lenore Cowen**, Tufts University, cowen@cs.tufts.edu, 617-627-5134