

RESEARCH INTERESTS: We are currently witnessing transformational advances in biology and computation happening in parallel: the availability of massive single-cell genomic datasets and the advent of deep learning. As a computational biologist, I aim to speed up the development of precision diagnostics and therapeutics by applying machine learning to biological data.

SELECTED PUBLICATIONS (SEE FULL LIST BELOW):

1. [Singh R*](#), Wu A*, Mudide A*, Berger B. *Unraveling causal gene regulation from the RNA velocity graph using Velorama*. Int'l Conf on Res. in Mol. Biology (**RECOMB**) 2023
2. [Singh R*](#), Sledzieski S, et al. *Contrastive learning in protein language space predicts interactions between drugs and protein targets*. **Proc Natl Acad Sci**. 2023 June 8;120(24): e2220778120.
3. 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.
4. [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.

HONORS & AWARDS

- George M. Sprowls **Award for the Best PhD** Theses in Computer Science at **MIT**.
- Christopher Stephenson Memorial **Award for Best Masters** Research in Computer Science at **Stanford**.
- **RECOMB Test of Time Award**: RECOMB is one of the two top-ranked conferences in the field of computational biology. The award recognized my paper on the IsoRank algorithm for network alignment.
- **World Finalist, ACM Programming Contest** (represented IIT Kanpur, the only Indian team).
- Dean's List, IIT Kanpur.
- 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 in my year**.
- National Talent Search Scholar, India.

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

EMPLOYMENT

- 2023-curr *Asst Professor* **Duke University (Dept of Biostatistics and Bioinformatics)** **Durham, NC**
Asst Professor **Duke University (Dept of Cell Biology)** **Durham, NC**
- My lab focuses on conceptual, mathematical and algorithmic innovations that enable faster and more precise discovery of disease mechanisms, identification of druggable targets, and design of novel small-molecule and antibody therapeutics.

- 2019-2023. *Research Scientist* **MIT Computer Science and Artificial Intelligence Lab.** **Cambridge, MA**
 2023-2023. *Instructor* **MIT Dept. of Mathematics** **Cambridge, MA**
- Discovering causal mechanisms of gene regulation from single-cell genomics data and the interpretable integration of multimodal single-cell data for understanding cell-level heterogeneity.
 - Scalable and accurate prediction of protein-protein interactions
 - Design and application of protein language models for drug-target interaction prediction and de novo design of monoclonal antibody therapeutics.
- 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 research team 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 my PhD, 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 Systems)** **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 2022).

Preprints and Manuscripts under review

1. [Singh R*](#), Li J*, Tattikota S* et al. *Optimal transport analysis of single-cell transcriptomics directs hypotheses prioritization and validation.* **Submitted.**
2. [Singh R*](#), Wu A*, Berger B. *Unveiling causal regulatory mechanisms through cell-state parallax.* **Submitted.**
3. [Singh R](#), Berger B. *Deciphering the species-level structure of topologically associating domains.* **Submitted.**
4. Louis M, Liao H, [Singh R](#), et al. *Using Machine Learning to Identify Metabolite Spectral Patterns that Reflect Outcome after Cardiac Arrest.* **Submitted.**

Refereed Conference Publications [RECOMB: 3, PSB: 4, ICML: 1, ICLR: 1 ISMB: 2 (+ 1 highlights), NeurIPS MLSB: 2]

In computational biology, RECOMB and ISMB have an acceptance rate < 20% and PSB has an acceptance rate < 30%.

In machine learning, ICML and ICLR are top-tier conferences, with acceptance rates of ~20% .

1. [Singh R*](#), Wu A*, Mudide A*, Berger B. *Unraveling causal gene regulation from the RNA velocity graph using Velorama*. Int'l Conf on Res. in Mol. Biology (**RECOMB**) 2023. **Invited by both Cell Systems and Genome Research for consideration**.
2. Sledzieski S*, [Singh R*](#), Cowen L, Berger B. *Contrasting drugs from decoys*. NeurIPS Workshop on Machine Learning in Structural Biology (**NeurIPS MLSB**) 2022 (to appear).
3. [Singh R*](#), Devkota K*, Sledzieski S, et al. *Topsy-Turvy: integrating a global view into sequence-based PPI prediction*. Int'l Symp of Intelligent Sys in Mol Bio (**ISMB**) 2022.
4. [Singh R*](#), Wu A*, Berger B. *Granger causal inference on DAGs identifies genomic loci regulating transcription*. Int'l Conf on Learning Representations (**ICLR**) 2022 (**scored in the top ~1% of ~3,300 submissions**).
5. 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.
6. 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.
7. **§** [Singh R](#), Xu J, Berger B. *Global alignment of multiple protein interaction networks*. Pac Symp Biocomput. (**PSB**) 2008:303-14.
8. [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.
9. **§§** [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.
10. Sontag D*, [Singh R*](#), Berger B. *Probabilistic modeling of systematic errors in two-hybrid experiments*. Pac Symp Biocomput (**PSB**) 2007.
11. **§** [Singh R*](#), Xu J*, Berger B. *Struct2net: integrating structure into protein-protein interaction prediction*. Pac Symp Biocomput (**PSB**) 2006.
12. [Singh R](#), Berger B. *Chaintweak: sampling from the neighbourhood of a protein conformation*. Pac Symp Biocomput (**PSB**) 2005:52-63.
13. [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.
14. [Singh R](#), Saha M. *Identifying structural motifs in proteins*. Pac Symp Biocomput (**PSB**) 2003:228-39.
15. **§** Deshpande H, Nam U, [Singh R](#). *Classification of Music Signals in the Visual Domain*. Proc of Digital Audio Effects Conf (**COST-G6**), 2001.
16. 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](#)]

1. [Singh R*](#), Sledzieski S*, Bryson B, Cowen L, & Berger B. *Contrastive learning in protein language space predicts interactions between drugs and protein targets*. **Proc Natl Acad Sci**. 2023, 120(24), e2220778120.
2. Ewen-Campen B*, Luan H*, Xu J*, [Singh R*](#) et al. *split-intein Gal4 provides intersectional genetic labeling that is repressible by Gal80*. **Proc Natl Acad Sci**. 2023 Jun 13;120(24):e2304730120.
3. Wu A, Markovich T, ... , [Singh R*](#). *Causally-guided Regularization of Graph Attention Improves Generalizability*. **Transactions on Machine Learning Research** 2023; 2835-8856.
4. [Singh R*](#), Devkota K*, Sledzieski S, et al. *Topsy-Turvy: integrating a global view into sequence-based PPI prediction*. **Bioinformatics** 2022; 38:i264-272.

5. Kumar L, Brenner N., et al. *Transfer of Knowledge from Model Organisms to Evolutionarily Distant Non-Model Organisms: The Coral Pocillopora damicornis Membrane Signaling Receptome*. **PLoS One** 2022.
6. 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.
7. [Singh R*](#), Hie B*, Narayan A, Berger B. *Schema: metric learning enables interpretable synthesis of heterogeneous single-cell modalities*. **Genome Biology** 2021; 22:131.
8. **§** 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.
9. **§** 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).
10. Hosur R, [Singh R](#), Berger B. *Sparse estimation for structural variability*. **Algorithms Mol Biol.** 2011; 6: 12.
11. **§** [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.
12. **§§** 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.
13. 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.
14. **§§** [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.
15. 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:*
 - MIT Bioinformatics seminar (2021 & 2022)
 - Columbia Univ (Guest lecture on protein interaction prediction) (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, 2021)
 - PSB (2003, 2005, 2006, 2007)
- *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 15 graduate/undergraduate/high-school students in my roles as Research Scientist and senior graduate student. All students below are/were at MIT or were visiting MIT:

- Graduate students: Samuel Sledzieski, Alex Wu, Daniel Park, Raghavendra Hosur, Sumaiya Nazeen.
- Undergraduate students: Chiho Im, Binwei Yan, Stuti Khandavala, Jesse Yang, Ishan Ganguli, Beckett Sterner, Irene Kaplow.
- High-school students: Anish Mudide (S.T. Yau Science Contest, National Gold Medal), Lynn Tao (Regeneron STS, Scholar), Akiva Kohane.

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*, (Applied: # 17/079,649).

SOFTWARE AND COMMUNITY RESOURCES

- *Haystack*: software to support a hybrid computational-biological protocol for discovering the lineage-determining regulators of a differentiation process (<http://haystack.csail.mit.edu>).
- *Velorama & Grid-Net*: graph neural network approaches to infer causal regulatory links from single-cell genomic data (<http://gridnet.csail.mit.edu>, <http://cb.csail.mit.edu/cb/velorama>).
- *D-SCRIPT & Topsy-Turvy*: deep learning methods 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 & IsoRank-N*: programs for global network alignment problem of protein interaction networks (<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

- **Program/Proceedings Committee member:** RECOMB 2023, ISMB/ECCB 2023.
- **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.

- 2022** **Key Person: NIH** (*under consideration*)
Focus: Cancer Target Discovery and Development
PIs: Prof. Bonnie Berger (MIT), Dr. Dou Yu (MIT), Prof. Omer Yilmaz (MIT)
- 2022** **Key Person: Sanofi Inc.** (*funded*). I led the application and negotiation process.
Focus: Protein language models for de novo antibody design and function prediction
PIs: Prof. Bonnie Berger (MIT) and Prof. Bryan Bryson (MIT)
- 2021** Supporting role: **NIH MIRA #13155717** (*funded*)
Focus: Manifold Representations and Active Learning for 21st Century Biology
PI: Prof. Bonnie Berger (MIT)
- 2021** **Key Person: Sanofi Inc.** (*completed*). I led 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** Supporting role: **NIH** (*funded*)
Focus: Resources for Functional Studies in *Drosophila*
PI: 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
4. **Bryan Bryson**, Massachusetts Institute of Technology, bryand@mit.edu, 617-258-7641