BIOGRAPHICAL SKETCH

NAME: Berger, Bonnie

eRA COMMONS USER NAME (credential, e.g., agency login): BABERGER

POSITION TITLE: Simons Professor of Mathematics and Professor of Electrical Engineering and Computer Science

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.)*

EDUCATION/TRAINING

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	Completion Date MM/YYYY	FIELD OF STUDY
Brandeis University, Waltham, MA	AB	06/1983	Computer Science
Massachusetts Institute of Technology	SM	01/1986	Computer Science
Massachusetts Institute of Technology	Ph.D.	06/1990	Computer Science
Massachusetts Institute of Technology	Postdoc	06/1992	Applied Mathematics

A. Personal Statement

Advances in modern biology revolve around automated data collection and sharing of the large resulting datasets. I am considered a pioneer in the area of bringing computer algorithms to the study of biological data, and a founder in this community that I have witnessed grow so profoundly over the last 26 years. I have made major contributions to many areas of computational biology and biomedicine, largely, though not exclusively through algorithmic innovations, as demonstrated by nearly twenty thousand citations to my scientific papers and widely-used software. In recognition of my success, I have just been elected to the National Academy of Sciences and in 2019 received the ISCB Senior Scientist Award, the pinnacle award in computational biology. My research group works on diverse challenges, including Computational Genomics, High-throughput Technology Analysis and Design, Biological Networks, Structural Bioinformatics, Population Genetics and Biomedical Privacy. I spearheaded research on analyzing large and complex biological data sets through topological and machine learning approaches; e.g. my lab has played integral roles in MPEG (biological data compression standard). I collaborate closely with biologists, MDs, and software engineers, implementing our new techniques in order to design experiments to maximally leverage the power of computation for biological exploration.

I have trained more than 100 undergrad, MEng, PhD students and postdoctoral fellows, many of whom now hold top academic positions. Among my PhD students are: Serafim Batzoglou – Stanford University tenured and Insitro; Phil Bradley – Fred Hutchinson and U. Washington; Manolis Kellis – MIT tenured; Lior Pachter – UC Berkeley and Caltech tenured; Mona Singh – Princeton University tenured; and Russell Schwartz – CMU tenured;. Recent PhD's who are Assistant Profs include: Michael Baym – Harvard Medical School (HMS); Alan Bryan – U Alabama Birmingham Medical School; Leonid Chindelevitch – Imperial College London; Po-Ru Loh – HMS/Broad Institute Member; Michael Schnall-Levin – 10X Genomics VP; and Y. William Yu – U.Toronto.

I continually and actively engage in community service and advocating diversity, through my roles as Vice President of ISCB, Head of the RECOMB Steering Committee, and a recent member of the NIH NIGMS Advisory Council. I have served as both Proceedings and Conference Chairs for the two top conferences in my field— RECOMB and ISMB. I have headed a workshop at ISMB 2016 on Gender Equality, and been instrumental in including women and diversity as ISCB Fellows Chair (2015-2020). I serve as an Executive Editor for Journal of Computational Biology, Associate Editor for Bioinformatics, and as member of editorial boards of Annual Reviews for Biomedical Data Science, IEEE/ACM TCBB, Genome Biology, and Cell Systems. I have given keynote addresses and distinguished lectures at RECOMBs, ISMB/ECCBs, ACM-BCB, RECOMB-RSG; RECOMB-Bioinformatics Education, Norway's Volterra Lecture, Gordon Research

Conference, CSHL, NIH Pi Day, Israeli Bioinformatics Symposium, GenoPri, AMS Joint Meetings and UCSD's Rosenblatt Lecture.

B. Positions and Honors

Positions and Employment

1990-1992 NSF Mathematical Sciences Postdoctoral Research Fellowship.
1992-1993 Radcliffe Bunting Institute, Science Scholar.
1992- Member of Computer Science & Artificial Intelligence Laboratory (CSAIL), MIT.
1992-1997 Assistant Professor of Applied Mathematics, MIT.
1997-1999 Associate Professor of Applied Mathematics, MIT.
1999-2002 Associate Professor of Applied Mathematics, tenured, MIT.
2002- Professor of Applied Mathematics, MIT.
2004-2012 Affiliated Faculty, Harvard-MIT Health Sciences and Technology (HST).
2004- Affiliated Faculty, Computational and Systems Biology (CSBi) at MIT.
2008- Beth Israel Deaconess Board of Overseers and Medical Advisory Committee.
2010- Joint Appointment, Dept. of Electrical Engineering and Computer Science, MIT.
2010- Associate Member, Broad Institute of MIT and Harvard.
2012- Affiliated Faculty, Harvard Medical School.
2014- Faculty Member, Harvard-MIT Health Science and Technology.
2015- Member, Center for Microbiome Informatics and Therapeutics.
2016- Simons Professor of Mathematics, MIT.
2018-2019 Interim Head of Applied Mathematics, MIT.

Other Experience and Professional Memberships

1995 Organizer for DIMACS Workshop: Sequence-based methods for protein folding.
1996-2003 BOD for Program in Mathematics and Molecular Biology (PMMB).
1998 NSF selection panel for the Protein Data Bank (PDB).
2001- Creator and organizer of MIT Math/CSAIL Bioinformatics Seminar.
2002- HST Graduate; Bioinformatics & Integrative Genomics; and Curriculum Committees.
2003-2006 ACM Nominating Committee.
2003-2014 NIH Scientific Review Group: Comparative modelling, BCMB & BDMA, ad-hoc member.
2004-2014 Brandeis University Science Advisory Council.
2006-2015 NIH NCBI Board of Scientific Counselors, 4 time ad-hoc member.
2008-2014 Beth Israel Deaconess Board of Overseers and Medical Advisory Committee.
2009-2014 NIH NIGMS Advisory Council, 3 time ad-hoc member.
2010 RECOMB 2010 Program Chair.
2010-2016 ISMB Proceedings Chair (2012), Conference Chair (2013) & Steering Committee (2012-13); Area Chair (2010, 2012-16).
2011-2017 Sloan Fellowship Selection Committee, Computational & Evolutionary Molecular Biology.
2014 Senior Member of ISCB.
2015-2018 FASEB Excellence in Science Award Committee.
2015- RECOMB Steering Committee Chair, and member since 2009.
2015.2016 NIH NIGMS Advisory Council.
2015-2021 ISCB Vice President; Member, Board of Directors; Cmte. Chair: Awards, Fellows & Diversity.
2016 ISMB 2016 Gender Equality Workshop Leader.
2016 Cold Spring Harbor Lab's Biological Data Sciences Program Organizer (with 2 others).
2018-2021 Barcelona Supercomputing Center Science Advisory Board.
2018-2020 NIH-NIGMS BDMA Panel.
2019 USC Biology Evaluation Committee.
2019-2022 AAAS Member-at-Large (Mathematics).

Selected Awards and Honors

1990 Ph.D. thesis won MIT George M. Sprowls Prize for best research in computer science.
1995-1998 NSF Career Award.
1999 Biophysical Society's Dayhoff Award for research (1 award per year).

1999 Technology Review's Inaugural TR100 Award for 100 top young innovators for the 21st century.
2004 Elected as a Fellow of the Association for Computing Machinery (ACM).
2010 RECOMB Test of Time Award for "Protein folding in the HP model is NP-complete."
2012 NIH Margaret Pittman Award for Outstanding Scientific Achievement & Lectureship.
2012 Elected as a Fellow of the International Society for Computational Biology (ISCB).
2013 Elected to the American Academy of Arts and Sciences.
2013 Brandeis University Alumni Achievement Award.
2015 École Polytechnique Fédérale de Lausanne (EPFL) Honorary Doctorate.
2016 Elected to the American Institute for Medical and Biological Engineering (AIMBE).
2019 RECOMB Test of Time Award for Isorank algorithm.
2019 Elected as a Fellow of the American Mathematical Society (AMS).
2019 ISCB Senior Scientist Award (1 per year since 2003).
2020 AWM / SIAM Sonya Kovalevsky Lecture Prize.
2020 Elected as a member of the National Academy of Sciences (NAS).

C. Contributions to Science (* for corresponding author or † for my student is 1st author)

1. Computational genomics. The last two decades have seen exponential increases in genomic and genetic data that will soon outstrip advances in computing power. Extracting new science from these massive datasets will require not only faster computers, but algorithms that scale sublinearly in the size of the data. I therefore introduced 'compressive genomics', a novel class of algorithms able to take advantage of data redundancy to compress biological data (ref *d*). This compression permits operations and computations directly on the compressed data, thereby enabling algorithms that scale sublinearly with the size of the data. These algorithms address seminal challenges in large-scale genomics (refs *b, c*), metagenomics and chemogenomics (*Cell Systems* 2015). There has been keen interest in this work by the computational biology research community, including an ISMB 2016 workshop. Importantly, the work was an invited contribution to *Nature Reviews Genetics* (2013), "Voices of Biotech" in *Nature Biotech* (May 2016 20th Anniversary Issue), *Communications of the ACM* (Cover, 2016) and *IEEE Transactions on Information Theory* (Levenshtein Honorary Issue, 2020). My student Ariya Shajii received the Best Student Paper Award at RECOMB/*Cell Systems* 2018 for his work on statistical binning of barcoded reads to improve downstream accuracy. With novel linear-time privacy-preserving algorithms for population stratification through compressive strategies, we made these techniques for GWAS realizable (ref *a*). Earlier, I founded and developed conservation-based methods for comparative genomics, together with my PhD students (S. Batzoglou, L. Pachter, and M. Kellis). Using these methods in collaboration with Dr. Eric Lander, we performed the first whole-genome alignments for human and mouse, as well as comparisons to detect exonic regions. We also performed the first comparisons of yeast genomes to identify genes and regulatory regions (nearly 2000 combined citations).

a) H. Cho †, D.J. Wu, and B. Berger *. "Secure genome-wide association analysis using multiparty computation." *Nature Biotechnology* **36** (2018): 547-551. Highlighted in *Science* (June 8, 2018). 60 citations
b) Deniz Yorukoglu †, Yun William Yu, Jian Peng, and Bonnie Berger *, **"**Compressive Mapping for Next-Generation Sequencing." *Nature Biotech* **4** (2016): 374-376.
c) Y.W. Yu, D. Yorukoglu, J. Peng and B. Berger *. "Quality Score Compression Improves Downstream Genotyping Accuracy." *Nature Biotech* **33** (2015): 240-3. 60 citations
d) P-R. Loh †, M. Baym †, and B. Berger *. "Compressive Genomics." *Nature Biotech* **30** (2012): 927-930. Most downloaded *Nat Biotech*, July 2012. 120 citations

2. Single-cell genomics. We have recently turned our attention to developing geometric and machine learning algorithms for visualizing (*Cell Systems*, 2018), integrating (ref *f*) and analyzing single-cell RNA-seq (scRNA-seq) data, which have led to important biomedical insights (ref *h*, *q*). A key insight of my lab has been that single-cell datasets are also highly constrained to lie along low-dimensional manifolds within high-dimensional spaces. We can make use of this property to better sample these datasets to construct compact representations of them (refs *e-g*). Our work receives great interest from the community (Ann Revs, 2020).

e) A. Narayan †, B. Berger *, and H. Cho *. "Assessing single-cell transcriptomic variability through density-preserving data visualization." *Nature Biotech* (Jan 18, 2021).

f) B. Hie †, B. Bryson, B. Berger *. "Efficient Integration of Heterogeneous Single-cell Transcriptomes using Scanorama." *Nature Biotech* **37** (2019):685–691. Earlier bioRxiv. 111 citations

g) B. Hie †, H. Cho †, B. DeMeo, B. Bryson and B. Berger *. "Geometric Sketching Compactly Summarizes the Single-Cell Transcriptomic Landscape." *Cell Systems* **8**, 6 (2019):483-493.e7 Cover image. 30 cites

h) J. Ordovas-Montanes, D. F. Dwyer, S. Nyquist †, K. M. Buchheit, M. Vukovic, C. Deb, M. H. Wadsworth II, T. K. Hughes, S. W. Kazer, E. Yoshimoto, K. N. Cahill, N. Bhattacharyya, H. R. Katz, B. Berger, T. M. Laidlaw, J. A. Boyce, N. A. Barrett and A. K. Shalek. "Allergic Inflammatory Memory in Human Respiratory Epithelial Progenitor Cells." *Nature* **560** (2018):649-654. 119 citations

3. Population genetics. My students and I have spearheaded recent exciting work in population genomics. In response to a challenge from Nick Patterson, David Reich, and later Alkes Price, we capitalized on our observations about the structure of population data as well as statistical and algorithmic advances to improve the power and speed of GWAS and population genetics inferences. Our BOLT-LMM (ref *i*) is the current state-of-the-art and software of choice for performing large-scale association studies. Our methods, Alder (ref *l*) and MixMapper, have newly enabled inference of population flow with admixture. As an example of practical utility, we used MixMapper to analyze genome-wide data from 56 populations and showed that all sampled Southeast Asian Austronesian groups harbor ancestry that is more closely related to aboriginal Taiwanese than to any present-day mainland population, thereby resolving a controversial question (*Nat Comm*, 2014). Our methods have also allowed us to uncover population flow in Roma, Indian and African populations. In addition, we have made significant progress on haplotype phasing, allowing haplotype reconstruction of a single sequenced individual using NGS data and applying these methods to Autism datasets (e.g. ref *k, Recomb* 2015, *Nature Comm* 2020). Most recently, we have extended haplotype phasing methods to uncover the role of mosaic copy number variation in ASD (*Nature Neuroscience*).

i) P-R. Loh †, G.J. Tucker †, B. Berger †, N. Patterson, and A. Price. "Efficient Bayesian Mixed-model Analysis Increases Association Power in Large Cohorts." *Nature Gen* **47** (2015): 284–290. 721 citations

j) I. Lazaridis, N. Patterson, A. Mittnik, G. Renaud, S. Mallick, K. Kirsanow, P. H. Sudmant, J. G. Schraiber, S. Castellano, M. Lipson, B. Berger, et al. "Ancient Human Genomes Suggest Three Ancestral Populations for Present-Day Europeans." *Nature* **513** (2014): 409-413. 981 citations

k) E. Berger, D. Yorukoglu, J. Peng, and B. Berger *. "HapTree: A Novel Bayesian Framework for Single Individual Polyplotyping using NGS Data." *PLoS Computational Biology* **10**, 3 (2014): e1003502. Also RECOMB 2014 & 2015. 74 citations

l) P. R. Loh †, M. Lipson †, N. Patterson, P. Moorjani, J. K. Pickrell, D. Reich *, and B. Berger *. "Inferring Admixture Histories of Human Populations Using Linkage Disequilibrium." *Genetics* **193**, 4 (2013): 1233-1254; including full cover. 320 citations

4. Biological discovery and functional annotation. I pioneered the highly active and rapidly evolving field of global network alignment. I introduced global biological network alignment (over 1000 citations)—a critical step in the transfer of functional knowledge across species—and set the standard for its use in functional orthology prediction, primarily through our Isorank suite of programs based on a novel Eigenvalue formulation of the product graph of networks (*US Patent 8000262 B2,* 2011). This work received the RECOMB 2019 Test of Time Award. Our Isorank algorithm and Isobase tools have been incorporated into numerous external webservers including the Perrimon lab's DiOPT. I have also developed approaches to characterize the prevalence of microRNAs and suggested additional function of transcribed coding regions, a finding later experimentally verified. We were one of first to bring representation learning to biology, integrating *heterogeneous* sources of information through networks, and successfully applied this to functional annotation (ref *o, p*). I have devised the first model to perform secure drug-target interaction prediction, newly leveraging neural networks for this problem (ref *n*). We have introduced a novel neural language processing model that newly combines both grammaticality (fitness) and semantics (antigenic capacity) to flag viral sequences that have a high amount of change for additional experiments to determine whether or not the immune system can still recognize these sequences. Our model accurately predicts mutant strains of Influenza HA, HIV Env and SARS CoV-2 proteins that have high escape potential, including new mutant strains of Covid that should be investigated for further escape (ref. *m*).

m) B. Hie †, E. Zhong †, B. Berger * and B. Bryson *. "Learning the Language of Viral Evolution and Escape." *Science* **371**, 6526 (Jan. 15, 2021) 284-288. Also NeurIPS 2020.

n) B. Hie †, H. Cho † and B. Berger *. "Realizing Private and Practical Pharmacological Collaboration." *Science* **362**, 6412 (2018): 347-350. Highlighted in *Science* issue. 22 citations

o) H. Cho †, B. Berger *, and J. Peng *. "Compact Integration of Multi-Network Topology for Functional Analysis of Genes." *Cell Systems* **3**, 6 (2016): 540-548. F1000 Prime recommended. 164 citations with *RECOMB 2015* conference version.

p) V. Khurana †, J. Peng †, C.Y. Chung †, P. K. Auluck, D. F. Tardiff, S. Fanning, T. Bartels, M. Koeva, S. W. Eichhorn, H. Benyamini, Y. Lou, A. Nutter-Upham, V. Baru, Y. Freyzon, N. Tuncbag, M. Costanzo, B. J. San Luis, D. C. Schöndorf, M. I. Barrasa, S. Ehsani, N. Sanjana, Q. Zhong, T. Gasser, D. P. Bartel, M. Vidal, M. Deleidi, C. Boone, E. Fraenkel *, B. Berger B * and S. Lindquist *. "Genome-Scale Networks Link Neurodegenerative Disease Genes to α-Synuclein through Specific Molecular Pathways." *Cell Systems* **4**, 2 (2017): p157-170.e14. Cover article; highlighted in *Science Trans Med*. Received Bishop Dr. Karl Golser Research Prize for ground-breaking work on Parkinson's. 78 citations

5. High-throughput technology analysis and design. Long-running experimental collaborations with Norbert Perrimon (HMS, HHMI), Isaac Kohane (HMS), Vik Khurana (HMS) and the late Susan Lindquist (WI, HHMI) have served to shed light on the genetics of disease though the development of methods to analyze RNAi in *Sci Signaling* (2011), Mass Spec, Lumier, and CLiP-seq data, in *Cell* 2014 (ref *t*), 2015. Moreover, we have designed experiments for measuring and designing RNA-protein binding (*Cell Systems* 2017), and developed the first algorithm to infer RNA structure binding preferences from experimental data. We introduced novel methods for particle picking (ref *s*) and heterogeneous reconstruction of cryo-EM structures (*Nat Meth*). We performed a meta-analysis of human, non-human primate, and mouse single-cell RNA-seq datasets for putative SARS-CoV-2 targets and identified important new targets (ref *r*). We introduced the use of uncertainty to predict *any* biomolecular binding interaction and demonstrate a computation-experimentation iterative loop to improve generative design of small molecule based on high biochemical affinity for a target (ref *q*). My group is solely responsible for the computational analyses in all joint collaborations.

q) B. Hie †, B. Bryson *, B. Berger *. "Leveraging Uncertainty in Machine Learning Accelerates Biological Discovery and Design." *Cell Systems* (2020) **11**, 5: 421-546.

r) Carly G. K. Ziegler, Samuel J. Allon, Sarah K. Nyquist †, …, Bonnie Berger †, Robert W. Finberg, Leslie S. Kean, Manuel Garber, Aaron G. Schmidt, Daniel Lingwood, Alex K. Shalek * Jose Ordovas Montanes *, HCA Lung Biological Network. "SARS-CoV-2 Receptor ACE2 is an Interferon-stimulated Gene in Human Airway Epithelial Cells and is Detected in Specific Cell Subsets Across Tissues." *Cell* **181**, 5 (2020):1016-1035. 646 citations

s) T. Bepler †, A. Morin, J. Brasch, L. Shapiro, A. Noble * and B. Berger *. "Positive-Unlabeled Convolutional Neural Networks for Particle Picking in Cryo-Electron Micrographs." *Nature Methods* **16** (2019): 1153–1160. 65 citations

t) M. Taipale, G. Tucker †, J. Peng †, I. Krykbaeva, Z. Y. Lin, B. Larson, H. Choi, B. Berger †, A. C. Gingras * and S. Lindquist *. "A Quantitative Chaperone Interaction Network Reveals the Architecture of Cellular Protein Homeostasis Pathways." *Cell* **158**, 2 (2014): 434-448. 279 citations

Overall, over half of my 200 papers are as a result of collaborations with experimental groups. Full bibliography is available at: https://scholar.google.com/citations?user=bYjKaowAAAAJ&hl=en&oi=ao

Ongoing Research Support

NIH 1-R01-GM081871-12 Berger (PI) April 1, 2008 – May 15, 2021
Structure-Based Prediction of the Interactome
Devise algorithms for inferring structural binding and investigating whether such data enhances systems-level analysis in genome-scale protein-protein, protein-RNA and protein-small molecule interactions.

NIH 1 R01 HG010959-01A1 Berger (PI), Amarasinghe, Cho (Co-I) Sept. 18, 2020 – June 30, 2024
Privacy-preserving Genomic Medicine at Scale
Develop computational methods for biomedical data integration, analysis, and interpretation in a privacy-preserving and highly scalable manner.

NIH 1U01CA250554-01 T. Lu, O. Yilmaz, Berger (PDs/PIs) July 8, 2020 – June 23, 2023
Developing High-throughput Genetic Perturbation Strategies for Single Cells in Cancer
Bo Biden Moonshot to Cancer Grant.

Charles E. Ross (1917) Seed Fund for Science Innovation Berger, Yilmaz (PIs) July 1, 2020 – June 30, 2021
Detecting Non-coding Driver Mechanisms of Cancer with Deep Learning
To introduce ML algorithms towards a high-resolution map of expected mutation density for a given cancer.

J-Clinic/Sanofi Collaborative Seed Grant Berger, Bryson (PIs) Jan. 1, 2021 – Dec. 31, 2021
Learning the language of Protein Interaction
To extend our uncertainty prediction approaches to study antibody-antigen interactions and prioritize high-affinity interactions to eventually incorporate these techniques into pharmacological workflows.