BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. DO NOT EXCEED FIVE PAGES.

NAME: Singh, Mona

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

POSITION TITLE: Professor of 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.)*

INSTITUTION AND LOCATION	DEGREE (if applicable)	END DATE MM/YYYY	FIELD OF STUDY
Harvard University, Cambridge, MA	AB	06/1989	Computer Science
Harvard University, Cambridge, MA	MS	06/1989	Computer Science
Massachusetts Institute of Technology, Cambridge, MA	PHD	09/1995	Computer Science
Princeton University	Postdoctoral Fellow	1997	Computational Biology
Whitehead Institute for Biomedical Research	Postdoctoral Fellow	1999	Computational Biology

A. Personal Statement

For more than 20 years, I have led a computational biology research group at Princeton University. I have a broad background in computer science, with formal training in algorithms, machine learning, and computational biology, along with postdoctoral training in a structural biology lab. My group has pioneered innovative data-driven approaches for uncovering cancer-relevant genes, predicting protein interactions, identifying functional and interaction sites within proteins, and analyzing biological networks. We are currently developing predictive methods to identify somatic alterations in cancer genomes that play a role in tumor initiation or progression, or that help shape the immune response to cancer. A key aspect of our current work is a focus on developing methods that work well across diverse populations.

Over the years, I have had many leadership positions in the computational biology community, including serving as Proceedings Chair for many major conferences (ISMB, RECOMB, WABI, and ACM-BCB) and serving as Editor-in-Chief of the Journal of Computational Biology

To date, 26 students have obtained PhDs under my supervision, of whom three are from racial groups underrepresented in science. Former students have gone on to successful careers in academia (including tenure-track or tenured faculty members at the University of California-San Francisco, Carnegie Mellon University, Princeton University, Duke University, and the NIH) and industry (including at Genentech, Janssen Research & Development, Google, 10X Genomics and GV20 Oncotherapy).

In summary, I have the background, expertise, and experience necessary to successfully carry out the proposed work.

Ongoing and recently completed projects that I would like to highlight include:

NIH R01CA208148
Singh (PI)
07/01/2016 – 06/30/2022
Interaction-based Computational Methods for Analyzing Cancer Genomes

NIH R01GM76275
Singh (PI)

04/01/2019 – 02/28/2023
Predicting and Analyzing Variation in Cellular Interactomes

Citations:

1. Berger B, Peng J, Singh M. Computational solutions for omics data. Nat Rev Genet. 2013 May;14(5):333-46. PubMed Central PMCID: PMC3966295.
2. Przytycka TM, Singh M, Slonim DK. Toward the dynamic interactome: it's about time. Brief Bioinform. 2010 Jan;11(1):15-29. PubMed Central PMCID: PMC2810115.

B. Positions, Scientific Appointments and Honors

Main Appointments

2018 - present	Wang Family Professor in Computer Science, Princeton University
2012 - present	Member, Rutgers Cancer Institute of NJ
2011 - present	Professor of Computer Science, Princeton University
2006 - 2011	Associate Professor of Computer Science, Princeton University
2001 - present	Department of Molecular Biology Graduate Faculty, Princeton University
2000 - present	Faculty, Lewis-Sigler Institute for Integrative Genomics, Princeton University
1999 - 2006	Assistant Professor of Computer Science, Princeton University

Other Positions and Scientific Appointments

2022	Conference Co-Chair, International Conference on Intelligent Systems for Molecular Biology
2021 - present	Editor-in-Chief, Journal of Computational Biology
2021 - present	Steering Committee, Workshops on Algorithms in Bioinformatics
2016	Proceedings Chair, Annual International Conference on Research in Computational Molecular Biology (RECOMB)
2012 - 2014	Chairperson, NIH MABS Study Section
2012	Proceedings Co-Chair, ACM Conference on Bioinformatics, Computational Biology, and Biomedicine
2010	Proceedings Co-Chair, Workshop on Algorithms in Bioinformatics
2010 - 2011	Steering Committee and Proceedings Committee Chair, International Conference on Intelligent Systems for Molecular Biology
2010 - 2011	Judge, Siemens Westinghouse National Competition
2009 - 2014	Member, NIH MABS Study Section
2008 - 2011	Board of Directors, International Society of Computational Biology
2000 - 2004	Steering Committee, DIMACS (NSF Center for Discrete Math and Theoretical Computer Science) Special Focus on Computational Molecular Biology

Honors

2019	Fellow, Association for Computing Machinery
2018	Fellow, International Society of Computational Biology
2003	Rheinstein Junior Faculty Award, Engineering and Applied Science, Princeton University
2001	NSF Presidential Early Career Award for Scientists & Engineers (PECASE), National Science Foundation
1998	Program in Mathematics and Molecular Biology Fellowship, Harvard University
1998 - 1999	Medical Foundation and Charles A. King Trust Fellowship, Charles A. King Trust Postdoctoral Research Fellowship Program
1989 - 1992	National Science Foundation Graduate Fellowship, National Science Foundation
1986 - 1989	John Harvard and Elizabeth Cary Agassiz Awards for Academic Achievement, Harvard University
1985 - 1986	National Merit Scholarship, National Merit Scholarship Program

C. Contribution to Science

1. Network approaches to identify cancer driver genes.

   A major goal of cancer genomics is to pinpoint which of the numerous alterations observed in cancer cells are relevant for tumor initiation or progression. Typically, numerous somatic alterations are observed in each cancer genome, only a subset of which are cancer-relevant, and very few of these alterations are observed across large numbers of individuals. Building upon our decades of work in biological network analysis, we have recently pioneered methods that consider somatic alterations in the context of molecular networks. We have shown how to uncover modules within networks comprised of genes that, while not individually frequently mutated, comprise pathways that are altered across a large fraction of individuals. We have introduced guided network propagation techniques that consider genes that are newly observed to be somatically altered in the context of prior information about known cancer genes; this framework is generally applicable to discover disease genes. We have launched a line of work that uses differential allele-specific expression to uncover genes that are dysregulated in cancer due to somatic mutations in cis. In all cases, we have rigorously shown that these methods are highly effective for uncovering cancer-driver genes, even those genes that are altered at low frequencies across tumors.

   a. Hristov BH, Chazelle B, Singh M. uKIN Combines New and Prior Information with Guided Network Propagation to Accurately Identify Disease Genes. Cell Syst. 2020 Jun 24;10(6):470-479.e3. PubMed Central PMCID: PMC7821437.
   b. Przytycki PF, Singh M. Differential Allele-Specific Expression Uncovers Breast Cancer Genes Dysregulated by Cis Noncoding Mutations. Cell Syst. 2020 Feb 26;10(2):193-203.e4. PubMed Central PMCID: PMC7457951.
   c. Hristov BH, Singh M. Network-Based Coverage of Mutational Profiles Reveals Cancer Genes. Cell Syst. 2017 Sep 27;5(3):221-229.e4. PubMed Central PMCID: PMC5997485.

2. Integrative approaches to identify cancer driver genes.

   A key principle driving much of our work in cancer genomics is that somatic mutations within coding regions should be considered in the context of the wealth of functional information known about protein sequences. Our earliest work in cancer genomics considered somatic mutations in the context of sites within proteins that participate in interactions. We developed the first method (with co-I Ghersi) that uncovers cancer-relevant genes by identifying those that have an enrichment of mutations in their interaction sites. We built upon this initial approach by incorporating other types of functional sites as well, and developed an integrative framework that identifies cancer-relevant genes by pinpointing those with an enrichment of mutations in any type of functional or interaction site, within domains, or within the gene as whole. A major technical accomplishment is that we newly derived analytical calculations that enable us to avoid time-prohibitive permutation-based significance tests, making it computationally feasible to simultaneously consider multiple measures of protein site functionality; this approach is generally applicable for other enrichment calculations. We also demonstrated how cancer-driver genes can be identified by integrating information about natural variation within genes, as the amount of natural variation a gene has reflects how mutable it is while maintaining function, and how cancer-relevant mutations can be identified by aggregating mutations over protein domains.

   a. Kobren SN, Chazelle B, Singh M. PertInInt: An Integrative, Analytical Approach to Rapidly Uncover Cancer Driver Genes with Perturbed Interactions and Functionalities. Cell Syst. 2020 Jul 22;11(1):63-74.e7. PubMed Central PMCID: PMC7493809.
   b. Munro D, Ghersi D, Singh M. Two critical positions in zinc finger domains are heavily mutated in three human cancer types. PLoS Comput Biol. 2018 Jun;14(6):e1006290. PubMed Central PMCID: PMC6040777.
   c. Przytycki PF, Singh M. Differential analysis between somatic mutation and germline variation profiles reveals cancer-related genes. Genome Med. 2017 Aug 25;9(1):79. PubMed Central PMCID: PMC5574113.
   d. Ghersi D, Singh M. Interaction-based discovery of functionally important genes in cancers. Nucleic Acids Res. 2014 Feb;42(3):e18. PubMed Central PMCID: PMC3919581.

3. Computational methods for predicting functionally important residues within proteins from sequence and structure.

   Proteins perform nearly all of their functions by interacting with other molecules. Knowledge of the "functional sites" which participate in these interactions is crucial not only for understanding the molecular mechanisms by which proteins carry out their functions but also for understanding how mutations (e.g., as observed in cancer) can affect their functions. Over the past three decades, a large number of methods for predicting functional sites have been developed. We have introduced computational approaches, based on geometric analysis and information theory, that vastly outperform these previous methods for identifying functional sites in proteins; this was shown not only in our work but also in subsequent follow up work by other research groups. Our contributions include an information-theoretic approach to identify residues interacting with either small molecule ligands or macromolecules; an approach to uncover specificity determining positions, that is, amino acids that are important for functional specificity; and a structural approach that identifies ligand-binding sites by directly integrating information about sequence conservation. Most recently, we have introduced a powerful new structural approach to identify ligand-binding sites in proteins by aggregating structural data at the level of domains, and then transferring knowledge about ligand binding from domains to the sequences that contain them. Using this approach, we have accurately identified sites that interact with DNA, RNA, peptides, ions or small molecules in the protein products of approximately 2/3rds of human genes; this vastly expands the number of proteins for which we have annotations of ligand binding.

   a. Kobren SN, Singh M. Systematic domain-based aggregation of protein structures highlights DNA-, RNA- and other ligand-binding positions. Nucleic Acids Res. 2019 Jan 25;47(2):582-593. PubMed Central PMCID: PMC6344845.
   b. Capra JA, Laskowski RA, Thornton JM, Singh M, Funkhouser TA. Predicting protein ligand binding sites by combining evolutionary sequence conservation and 3D structure. PLoS Comput Biol. 2009 Dec;5(12):e1000585. PubMed Central PMCID: PMC2777313.
   c. Capra JA, Singh M. Characterization and prediction of residues determining protein functional specificity. Bioinformatics. 2008 Jul 1;24(13):1473-80. PubMed Central PMCID: PMC2718669.
   d. Capra JA, Singh M. Predicting functionally important residues from sequence conservation. Bioinformatics. 2007 Aug 1;23(15):1875-82. PubMed PMID: 17519246.

4. Computational methods for predicting and characterizing protein interactions and protein interaction specificity.

   In general, predicting the molecular interactions of a protein is a difficult problem. From its founding, my group has been well-known for developing structural bioinformatics approaches for predicting protein interactions and protein interaction specificity by focusing on well-characterized structural domains. Our approaches incorporate both structural knowledge about binding specificity determinants as well as high-throughput experimental data. We have developed some of the best methods for predicting protein-protein interactions mediated by coiled coil domains and protein-DNA interactions mediated by Cys2His2 zinc fingers (the largest class of transcription factors in metazoans). The software corresponding to these methods have been accessed by thousands of unique IP addresses, and have been used by others to perform a range of interesting biological analyses, from analyzing the evolution of transcription factor networks (Pinney et al., PNAS 2009) to implicating a protein as a regulator of meiotic hotspots in human (Myers et al., Science 2010). Most recently, we have developed structural bioinformatics approaches to predict specificity for homeodomains, the second largest class of transcription factors in metazoans. Our proposed work on predicting MHC-peptide interactions builds upon our decades of experience in predicting protein interactions.

   a. Wetzel JL, Zhang K, Singh M. Learning probabilistic protein-DNA recognition codes from DNA-binding specificities using structural mappings. Genome Res. 2022 Sep 19;32(9):1776-86. PubMed Central PMCID: PMC9528988.

b. Persikov AV, Wetzel JL, Rowland EF, Oakes BL, Xu DJ, Singh M, Noyes MB. A systematic survey of the Cys2His2 zinc finger DNA-binding landscape. Nucleic Acids Res. 2015 Feb 18;43(3):1965-84. PubMed Central PMCID: PMC4330361.
c. Persikov AV, Singh M. De novo prediction of DNA-binding specificities for Cys2His2 zinc finger proteins. Nucleic Acids Res. 2014 Jan;42(1):97-108. PubMed Central PMCID: PMC3874201.
d. Fong JH, Keating AE, Singh M. Predicting specificity in bZIP coiled-coil protein interactions. Genome Biol. 2004;5(2):R11. PubMed Central PMCID: PMC395749.

5. Graph-theoretic algorithms for analyzing large-scale cellular networks.

Cellular networks provide a view into the workings of the cell. However, these interaction maps do not come with a key for interpreting them, so it is necessary to develop methods that shed light on their functioning and organization. We introduced network flow techniques as a means for analyzing cellular networks; this work was one of the first to propagate functional information through interaction networks. We introduced network schemas as a general formalism to represent topological patterns that integrate the wealth of prior information known about individual proteins (e.g., their predicted domain structures, functional annotations, or other attributes), and developed fast algorithms and software for querying networks using them. We further showed that it is possible to automatically infer, from all possibilities, which of these patterns are recurrent and over-represented, and demonstrated that these patterns are biologically important. We have also helped enable new cross-interactome analyses by developing a fast network clustering algorithm, particularly suited for dense networks, and have demonstrated when and how network clustering should be used to annotate protein function.

a. Jiang P, Singh M. SPICi: a fast clustering algorithm for large biological networks. Bioinformatics. 2010 Apr 15;26(8):1105-11. PubMed Central PMCID: PMC2853685.
b. Song J, Singh M. How and when should interactome-derived clusters be used to predict functional modules and protein function? Bioinformatics. 2009 Dec 1;25(23):3143-50. PubMed Central PMCID: PMC3167697.
c. Banks E, Nabieva E, Peterson R, Singh M. NetGrep: fast network schema searches in interactomes. Genome Biol. 2008;9(9):R138. PubMed Central PMCID: PMC2592716.
d. Nabieva E, Jim K, Agarwal A, Chazelle B, Singh M. Whole-proteome prediction of protein function via graph-theoretic analysis of interaction maps. Bioinformatics. 2005 Jun;21 Suppl 1:i302-10. PubMed PMID: 15961472.

Complete List of Published Work in My Bibliography:
https://www.ncbi.nlm.nih.gov/myncbi/mona.singh.1/bibliography/public/