Max W. Shen

maxwshen@mit.edu | www.maxwshen.com | U.S. Citizen

My research uses applied machine learning and statistical methods for fundamental scientific discovery and high-impact applications. As a computationalist at heart, I believe in interdisciplinarity, getting one's feet wet, and that better solutions to real-world problems arise by marrying each problem's unique structure with thoughtful modeling and inference design.

EDUCATION

Ph.D. Candidate, Computational & Systems Biology
Massachusetts Institute of Technology, 2015 – present. *Expected 2021.*

B.S. in Computer Science with Specialization in Bioinformatics
Summa Cum Laude
University of California, San Diego, 2015.

RESEARCH EXPERIENCE

Broad Institute of M.I.T. and Harvard
Doctoral Research, 2018 – present
Advisors: David R. Liu & Aviv Regev
- Design of high-throughput biological experiments for investigations of the relationships between DNA sequence context and genome editing
- Development of statistical and computational methods for directed evolution
- Design of machine learning models combining deep learning with biological domain knowledge

Computer Science & Artificial Intelligence Laboratory, Massachusetts Institute of Technology
Doctoral Research, 2015 – 2018
Advisor: David K. Gifford
- Investigation of the relationship between DNA sequence context and CRISPR editing outcomes
- Design of high-throughput CRISPR experiments for functional genomics
- Statistical analysis and data processing of high-throughput sequencing data
- Design of machine learning models combining deep learning with biological domain knowledge
- Development of user-friendly data visualizations

Computer Science Department, University of California, San Diego
Undergraduate Research, 2014 – 2015
Advisor: Pavel A. Pevzner

Radiology Imaging Laboratory, University of California, San Diego, School of Medicine
Undergraduate Research, 2011 – 2013
Advisor: Mingxiong Huang

PUBLICATIONS & PROJECTS

Research interests: High-throughput biology, genome editing, directed evolution, statistical modeling.

Arbab, M.*, **Shen, M.W.***, Mok, B., Wilson, C., Matsuzek, Z., Cassa, C.A., Liu, D.R. "Determinants of Base Editing Outcomes from Target Library Analysis and Machine Learning." *Cell* (2020), in press.

Shen, M.W.*, Arbab, M.*, Hsu, J.Y., Worstell, D., Culbertson, S.J., Krabbe, O., Cassa, C.A., Liu, D.R., Gifford, D.K., Sherwood, R.I. "Predictable and precise template-free CRISPR editing of pathogenic variants." *Nature* 563, 646-651 (2018).

Miller, S., Wang, T., Randolph, P.B., Arbab, M., **Shen, M.W.**, Huang, T.P., Matuszek, Z., Newby, G.A., Rees, H.A., Liu, D.R. "Continuous evolution of SpCas9 variants compatible with non-G PAMs." *Nature Biotechnology* (2020).

Antipov, D., Hartwick, N., **Shen, M.W.**, Raiko, M., Lapidus, A., Pevzner, P.A. "plasmidSPADES: assembling plasmids from whole genome sequencing data." *Bioinformatics* 32 (22), 3380-3387 (2016).

Lin, Y., Yuan, J., Kolmogorov, M., **Shen, M.W.**, Chaisson, M., Pevzner, P.A. "Assembly of long error-prone reads using de Bruijn graphs." *Proceedings of the National Academy of Sciences* 113 (52), E8396-E8405 (2016).

Huang, Ming, Huang, C. W., Robb, A., Angeles, A., Nichols S. L., Baker D. G., Song T., Harrington D. L., Theilmann R. J., Srinivasan R., Heister D., Diwakar M., Canive J. M., Edgar J. C., Chen Y., Ji Z., **Shen M.W.**, El-Gabalawy F., Levy M., McLay R., Webb-Murphy J., Liu T. T., Drake A., Lee R. R. "MEG source imaging method using fast L1 minimum-norm and its applications to signals with brain noise and human resting-state source amplitude images." *NeuroImage* 84, 585-604 (2014).

PRESENTATIONS, AWARDS, & PRESS

Invited talk: "Predicting and optimizing base editing outcomes". Models, Inference, and Algorithms; Broad Institute, May 2020 (postponed).

Invited talk: "Predicting and optimizing base editing outcomes". MIT Computational & Systems Biology Ph.D. Program Retreat; Oct. 2020.

Invited talk: "Predictable and precise template-free CRISPR editing of pathogenic variants." Arbor Biotechnologies, Jun. 2019.

Invited talk: "Predictable and precise template-free CRISPR editing of pathogenic variants." Synthego, Jan. 2019.

Press feature (< 1% selected): Recognition for data visualization
https://plot.ly/newsroom/indelphi/, Nov. 2018.

Selected talk (~15% selected): "Predictable and precise template-free CRISPR editing of pathogenic variants."
Biology of Genomes Conference at Cold Spring Harbor, May, 2018.

Best poster award (top 1 out of 20): "Predictable and precise template-free CRISPR editing of pathogenic variants." Computational & Systems Biology Ph.D. Program Retreat, Oct. 2017.

NSF graduate research fellowship (~15% acceptance), 2015.

SERVICE

Models, Inference, and Algorithms (MIA) at Broad Institute
    Steering committee member (2020 – present)

TEACHING EXPERIENCE

MIT independent activities period 2018: "Causal inference & deep learning."
    Prepared and co-taught with a postdoc this short graduate level class (4 classes, 6 total hours) from recent papers in the field of deep learning and causal inference. Typical attendance: 20 students.

MIT independent activities period 2017: "Applied probabilistic programming & Bayesian machine learning."
    Prepared and co-taught with two graduate students this short upper-undergraduate level class (6 classes, 9 total hours) on Bayesian machine learning and Stan. First class attendance: 100 students. Typical attendance: 25 students.

Teaching assistant, MIT. "Quantitative & Computational Biology." (2016).

Rosalind (online bioinformatics textbook) contributor, 2014 – 2015.

With co-founder Phillip Compeau (Teaching professor at Carnegie Mellon University), designed and wrote online problem sets with automatic code validation, and proofread and advised on textbook chapters.

Teaching assistant, UC San Diego. "Introduction to Artificial Intelligence: Search and Reasoning" (2015) and "Basic Data Structures & Object-Oriented Design" (2013).

MANAGEMENT EXPERIENCE

FUSION dance competition XIV & XV coordinator, 2013 – 2015.
Managed 9 teams of 110 students
Financially responsible for $35,000 in revenue and expenses
Organized an annual show with 1,800 audience members

INDUSTRY EXPERIENCE

Bioinformatics intern, Human Longevity, Inc. San Diego, CA. 2015.

Software engineering intern, Illumina, Inc. San Diego, CA. 2014.

Software engineering intern, Qualcomm, Inc., research & development department. Seoul, South Korea, 2013.

REFERENCES

David R. Liu
Professor of Chemistry and Chemical Biology
Harvard University
(617) 714-8650, liu@chemistry.harvard.edu

Richard I. Sherwood
Assistant Professor of Medicine, Division of Genetics
Brigham and Women's Hospital and Harvard Medical School
(650) 799-9363, rsherwood@rics.bwh.harvard.edu

Aviv Regev
Professor of Biology
Massachusetts Institute of Technology
(617) 714-7021, aregev@broadinstitute.org