# Max W. Shen

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

My research uses applied machine learning and statistical methods. During my Ph.D., I have focused on advancing fundamental science in genome editing and directed evolution. 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. My work has been invited for a talk at a NeurIPS workshop for machine learning in biology, and published in *Nature* and *Cell*. I have taught graduate-level MIT IAP classes on Bayesian machine learning, causal inference, and deep learning.

Expertise: Applied machine learning, Scientific thinking and experimental design, Data visualization
Skills and tools: Python, Pytorch, Jupyter, GitHub, scikit-learn, Dash plotly, Photoshop/Illustrator, LaTeX, Pandas

## 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

- Designed efficient inference algorithm for a non-linear dynamical system, introducing a trick to bypass an intractable maximum likelihood optimization. Devised novel sparsity regularization loss with theoretical motivation (Invited talk at NeurIPS 2020, LMRL workshop)
- Designed deep conditional autoregressive models to solve sequence-to-sequence problems in genome editing. Led the development of machine learning models and built interactive web app with 300 monthly users (first-author work published in *Cell*, 2020)
- Designed large-scale experiments, and processed and cleaned large datasets

Computer Science & Artificial Intelligence Laboratory, Massachusetts Institute of Technology
*Doctoral Research,* 2016 – 2018
Advisor: David K. Gifford

- Led a collaborative project including development of novel machine learning models, designing experiments at 10x scale of prior work, culminating in a first-author paper in *Nature* (2018)
- Designed a jointly-trained multitask deep learning model with sister networks to learn noisy genome editing activity

- Developed an interactive web app data visualization with 500 users monthly (Recognized by Dash Plotly)

Computer Science Department, University of California, San Diego
*Undergraduate Research*, 2014 – 2015. Advisor: Pavel A. Pevzner
*Undergraduate Research*, 2011 – 2013. Advisor: Mingxiong Huang

## PUBLICATIONS & PROJECTS

**Shen, M.W.,** Zhao, K. T., Liu, D. R. Reconstruction of genes, evolutionary trajectories, and fitness from short sequencing reads of laboratory evolution experiments using machine learning.

- Invited talk at NeurIPS 2020: Learning Meaningful Representations of Life workshop
- Under review at *Nature Chemical Biology* (2021)

**Shen, M.W.*,** Arbab, M.*, 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* 182, 2, 463-480.e30 (2020).

- Featured on the cover of the 7/23/20 issue of Cell with my cover art.
- Co-first author order is listed differently from publication order with permission from all co-first authors.

**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).

Koblan, L.W.*, **Shen, M.W.***, Arbab, M.*, Hussmann J.A., Anzalone, A.V., Doman, J.L., Newby, G.A., Yang, D., Mok, B., Replogle, J.M., Xu, A., Sisley, T.A., Weissman J.S., Adamson, B., Liu, D.R. "Development of C•G-to-G•C transversion base editors from CRISPRi screens, target-library analysis and machine learning." *Nature Biotechnology* (2021, in press).

Yeo, G.H.T., Juez, O., Chen, Q., Banerjee, B., Chu, L., **Shen, M.W.**, Sabry, M., Logister, I., Sherwood, R.I., Gifford, D.K. "Detection of gene cis-regulatory element perturbations in single-cell transcriptomes." *PLOS Computational Biology* 17, 3 (2021).

Hsu, J.Y., Grünewald, J., Szalay, R., Shih, J., Anzalone, A.V., Lam, K.C., **Shen, M.W.**, Petri, K., Liu, D.R., Joung, J.K., Pinello, L. "PrimeDesign software for rapid and simplified design of prime editing guide RNAs." *Nature Communications* 12 (1), 1-6 (2021).

Louie, W., **Shen, M. W.**, Tahiry, Z., Zhang, S., Worstell, D., Cassa C.A., Sherwood R.I., Gifford D.K. "Machine learning based CRISPR gRNA design for therapeutic exon skipping." *PLOS Computational Biology* 17, 1 (2021).

Lin, L., Holmes, B., **Shen, M.W.**, Kammeron, D., Geijsen N., Gifford D.K., Sherwood R.I. "Comprehensive Mapping of Key Regulatory Networks that Drive Oncogene Expression." *Cell Reports* 33, 8 (2020)

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).

## 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).

## 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.

## PRESENTATIONS, AWARDS, & PRESS

Invited talks

- NeurIPs, Learning Meaningful Representations of Life Workshop; December 2020. (Top 10%)
- Models, Inference, and Algorithms; Broad Institute, December 2020.
- MIT Computational & Systems Biology Ph.D. Program Retreat; Oct. 2020.
- Arbor Biotechnologies, Jun. 2019.
- Synthego, Jan. 2019.
- Biology of Genomes Conference at Cold Spring Harbor, May, 2018 (Top 15%)

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

NeurIPs 2020: Learning Meaningful Representations of Life Workshop. Co-organizer and submission committee member

Reviewer for *The CRISPR Journal*

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

## 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
Oversaw marketing, logistics, media, hiring contractors, finances, etc

## 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