Sarah Gurev

Machine learning for protein evolution and design

(916) 838-9672
sgurev@mit.edu

Education

2020 - Present **Ph.D. Candidate in Electrical Engineering and Computer Science**, MIT.
Master of Science in EECS, MIT 2023.
Advisor: Debora Marks, Harvard Medical School
Takeda, Richard Frazier, and Verena Fellowships

2016 - 2020 **Bachelor of Science in Computer Science**, *Biocomputation Track*, Stanford University.
Tau Beta Pi Engineering Honors Society
Research with Gill Bejerano, Alex Chan, Michael Levitt, and Ravi Majeti

Publications

- N. Thadani*, **S. Gurev***, P. Notin*, N. Youssef, N. Rollins, D. Ritter, C. Sander, Y. Gal, D. Marks (2023), **Learning from pre-pandemic data to forecast viral escape**, *Nature.*
  * indicates equal contribution
- N. Youssef, **S. Gurev**, ... J. Lemieux, J. Luban, M. Seaman, D. Marks (2024), **Protein design for evaluating vaccines against future viral variation**, *BioRxiv.*
- A. Shaw, H. Spinner, J. Shin, **S. Gurev**, N. Rollins, D. Marks (2023), **Removing bias in sequence models of protein fitness**, *BioRxiv.*
- S. Salman, **S. Gurev**, M. Arsalan, F. Dar, A. Chan (2021), **Liver exchange: A pathway to increase access to transplantation**, *Harvard Health Policy Review.*
- M. Linde, A. Fan, T. Köhnke, A. Trotman-Grant, **S. Gurev**, ... R. Majeti (2023), **Reprogramming cancer into antigen presenting cells as a novel immunotherapy**, *Cancer Discovery.*

Short Papers and Extended Abstracts:

- **S. Gurev***, N. Youssef*, N Jain, D. Marks (2024), **Alignment-based and protein foundation models for viral evolution, vaccines and vectors**, *Neurips Workshop on Machine Learning for Structural Biology & Neurips Workshop on AI for New Drug Modalities* (Spotlight).
- **S. Gurev***, N. Youssef*, H. Pierce-Hoffman, D. Marks (2024), **Future-proof vaccine design with a generative model of antibody cross-reactivity**, *ICLR Workshop on Generative and Experimental Perspectives for Biomolecular Design & ICML Workshop on Machine Learning for Life and Material Science.*
- **S. Gurev**, N. Youssef, N. Thadani ... J. Lemieux, J. Luban, M. Seaman, D. Marks (2023), **Learning from pre-pandemic data for the design and testing of variant-proof vaccines**, *Molecular Machine Learning Conference.* (Selected for Contributed Talk - Best 4 Papers)
- N. Thadani, N. Rollins, **S. Gurev**, P. Notin, Y. Gal, D. Marks (2021), **Viral evolution and antibody escape mutations using deep generative models**, *ICML Workshop on Computational Biology.* (Selected for Spotlight Talk)
- M. Linde, C. Dove, **S. Gurev**, P. Phan, F. Zhao, L. Miller, R. Majeti (2019), **Reprogramming leukemia cells into antigen presenting cells as a novel cancer vaccination immunotherapy**, *Blood.*

Funding and Awards

- TIME 100 AI (2024)
- Verena Fellow-in-Residence Award (2024)
- AI for New Drug Modalities Workshop at Neurips - Spotlight (2024)
- Merck-MBG Biologics and Vaccines Symposium - Best Poster Award (2024)
- Molecular Machine Learning Conference - Best Paper/Talk (2023)
- Takeda Fellowship (2022)
- Workshop on Computational Biology ICML - Spotlight (2021)
- Richard H. Frazier Fellowship (2020)
- Stanford Undergraduate Research Major Grant (2018)
- Tau Beta Pi Engineering Honors Society (2019)
- USA Biology Olympiad Semifinalist (2015)
- Science Olympiad National Champion (2016) and Top 5 (2012-2015)

Conferences and Presentations

Talks

- **Deep generative model foreshadows SARS-CoV-2 evolution and facilitates early vaccine evaluation.** *American Society for Virology*, 2024.
- **Using the past to predict the future: unsupervised early warning of viral antibody escape.** *Boston Protein Design and Modeling Club*, 2024.
  `https://www.youtube.com/watch?v=ZgrTrBYZE48&t=150s`
- **Unsupervised viral antibody escape prediction for future-proof vaccines.** *Broad Institute Models, Inference & Algorithms*, 2024.
  `https://www.youtube.com/watch?v=MsSYYc_qZ3U`
- **Future-proof vaccine design using deep generative models of antibody escape.** *Contributed Talk, Andean School on Host-Pathogen Dynamics*, 2024.
- **Chalk talk on vaccine design.** *MIT Computational and Systems Biology*, 2024.
- **Early warning of viral antibody escape from a biologically-informed computational framework & Pandemic surveillance discussion panel** *Invited Talk, Precision Public Health Symposium*, 2023.
- **Learning from pre-pandemic data for the design and testing of variant-proof vaccines.** *Contributed Talk (Best 4 papers), Molecular Machine Learning Conference*, 2023.
- **Using the past to predict the future: unsupervised early warning of viral antibody escape.** *NSF-Simons Center at Harvard*, 2023.
- **Predicting viral antibody escape: An integrated computational and experimental approach.** *Harvard Systems Biology*, 2023.
- **Early warning of viral antibody escape from a biologically-informed computational framework.** *Contributed Talk, Winter q-Bio*, 2023.
- **Learning from pre-pandemic data to forecast viral antibody escape.** *Broad Institute Primer on Medical and Population Genetics*, 2022.
  `https://www.youtube.com/watch?v=NkOAaFcYetU`
- **Learning from pre-pandemic data to forecast viral antibody escape.** *Massachusetts Consortium on Pathogen Readiness*, 2022.

Media

- **A New AI tool that can predict viral variants.** *KCBS Radio*, 2023. https://www.audacy.com/podcast/kcbs-radio-on-demand-011f4/episodes/a-new-ai-tool-that-can-predict-viral-variants-7fe31?

Posters

- **S. Gurev**, N. Youssef, N. Thadani ... J. Lemieux, J. Luban, M. Seaman, D. Marks, *Merck-MBG Biologics and Vaccines Symposium*, Learning from pre-pandemic data for the design and testing of variant-proof vaccines, 2024 (**Best Poster Award**)
- **S. Gurev***, N. Youssef*, D. Marks, *Virus Genomics, Evolution and Bioinformatics Conference*, Tradeoffs of alignment-based and protein language models for viral fitness and escape prediction, 2024
- N. Youssef*, **S. Gurev***, H. Pierce-Hoffman, L. Caldera, A. Cohen, P. Bjorkman, D. Marks, *ICML Workshop on Machine Learning for Life and Material Science*, Future-proof vaccine design with a generative model of antibody cross-reactivity, 2024
- **S. Gurev***, N. Youssef*, H. Pierce-Hoffman, D. Marks, *ICLR Workshop on Generative and Experimental Perspectives for Biomolecular Design*, A future-proof vaccine design method using a deep generative model of antibody escape, 2024
- **S. Gurev**, N. Youssef, N. Thadani ... J. Lemieux, J. Luban, M. Seaman, D. Marks, *Molecular Machine Learning Conference*, Learning from pre-pandemic data for the design and testing of variant-proof vaccines, 2023
- **S. Gurev**, N. Youssef, N. Thadani ... J. Lemieux, J. Luban, M. Seaman, D. Marks, *Vaccines Summit*, Learning from pre-pandemic data for the design and testing of variant-proof vaccines, 2023
- **S. Gurev**, N. Youssef, N. Thadani, ... J. Lemieux, J. Luban, M. Seaman, D. Marks, *Gordon Research Conference on Protein Engineering*, Design and testing of variant-proof vaccines from machine learning models on pre-pandemic data, 2023
- **S. Gurev**, N. Thadani, P. Notin, N. Youssef, N. Rollins, C. Sander, Y. Gal, D. Marks, *Winter q-Bio conference*, Early warning of viral antibody escape from a biologically-informed computational framework, 2023
- N. Thadani, N. Rollins, **S. Gurev**, D. Marks, *CSHL Probabilistic Modeling in Genomics*, Predicting SARS-CoV-2 evolution with deep generative models of natural sequences, 2021
- N. Thadani, N. Rollins, **S. Gurev**, P. Notin, D. Marks, *Atlas of Variant Effect Mutational Scanning Symposium*, Using coronavirus sequences and mutation effects data to predict evolution of SARS-CoV-2, 2021
- **S. Gurev**, J. Rodrigues, M. Levitt, *Stanford Bio-X Symposium*, Understanding determinants of affinity in receptor:chemokine interactions with molecular dynamics, 2018

Service

Peer-Reviewer, *Nature Communications (2) - co-review, PNAS - co-review, ICLR Workshop on Generative and Experimental Perspectives for Biomolecular Design, ICML Workshop on ML for Life and Material Science, Mutational Scanning Symposium*.

2024 **Teaching Assistant**, *Advanced Computational Biology: Genomes, Networks, Evolution*, MIT.

- Wrote new problem set, led office hours and recitations, mentor projects on ML for proteins

2021 - 2024 **Mentor**, *Graduate Application Assistance Program*, MIT EECS.

- Advised many underrepresented PhD applicants each year throughout their grad school applications

2021 - 2024 **Mentor**, *Thriving Stars*, MIT EECS.

- Advise several newly admitted women graduate students

2024 **Lectures**.

- Advanced Computational Biology graduate course, MIT EECS: Created and delivered guest lecture on protein language models.
  `https://www.youtube.com/watch?v=uPoFdCUqBWk`
- Next Epoch, Harvard: Assisted in teaching and developing material for a 3 day machine learning in biology tutorial for primarily first-gen college students.
- Health Disparities Think Tank: Taught and created code tutorial for a undergraduate workshop on exploratory data analysis and data visualization.
- Machine learning for Healthcare graduate course, MIT EECS: Created and delivered guest lectures on genomics in medicine.
- Skype a Scientist

2016 - 2023 **Executive Director, Director & Event Supervisor**, *Golden Gate Science Olympiad*, Stanford University and UC Berkeley.

- Directed 501(c)(3) nonprofit that holds a yearly science competition for 800+ high school students
- Managed a 12-person board of directors and over 150 volunteers, including running weekly meetings as well as coordinating the scientific events, developing timelines, and writing grants
- Expanded community development efforts by founding the Adopt a Team program and coaching one of the first international Science Olympiad teams (team of girls from Peru)
- Continued involvement in Science Olympiad volunteering for other competitions

2019 - 2020 **Founder, President & Teacher**, *Adopt a Science Olympiad Team at Stanford*, Stanford University.

- Founded an organization to create and coach Science Olympiad teams at schools in local underserved communities - a legacy which continues to coach new teams today
- Led team of volunteers to partner with local charter schools and Lauren's House afterschool program (East Palo Alto nonprofit) to prepare students to compete in local competitions
- Raised money for competition fees and engineering materials so student participation would be free
- Taught weekly after school science lessons and weekend all-day-build-events designed around preparing students for Science Olympiad competition (coding in Scratch, balsa wood bridge building, bottle rocket building, anatomy, oceanography, etc.)

2009 - 2017 **Teaching Assistant**, *Diagnostic Preschool Classroom*, Special Education Program, Ralph Richardson Center.

- Individual instruction for special education preschoolers learning to walk and communicate (1000+ hours)

Research Experience

2020 - Present **PhD Student**, *MIT*, Debora Marks Lab, Harvard Medical School.

- Predicting trajectories of future viral evolution due to immune constraints using deep generative models of historic sequence diversity as well as structural and biochemical data available pre-pandemic
- Early warning of SARS-CoV-2 escape variants by iteratively modeling and optimizing selection of emerging and forecasted sequences to be assayed for infectivity and antibody neutralization
- Application of modeling of antibody escape to both design and test variant-proof vaccines
- Modeling viral escape across disease-threat viral families

2020 - 2021 **Undergraduate Researcher**, *Liver Exchange Project*, Stanford University.

- Independently operationalized an optimal liver exchange with balanced risk algorithm
- Helped finalize algorithm, converted algorithm into codebase, and ran simulations
- Used matching algorithm to find previously missed matches for liver organ exchanges in Pakistan

2019 - 2020 **Undergraduate Researcher**, *Gill Bejerano Lab*, Stanford University.

- Independently developed an automated abstraction NLP tool that can identify patients undergoing a diagnostic odyssey from their clinical notes

2019 **Computational Biology Intern**, *Clinical Virology*, Gilead Sciences.

- Evaluated machine learning tools for peptide-MHC binding and presentation prediction and built a pipeline to investigate HIV peptide and HLA allele combinations for the HIV Vaccine project
- Created MongoDB research database of HIV peptide and mutant data
- Developed method to select mutation combinations critical to antibody binding to select subjects

2018 - 2019 **Undergraduate Researcher**, *Michael Levitt Lab*, Stanford University.
- Awarded competitive Stanford Major Grant based on research proposal surrounding the use of homology modeling and molecular dynamics simulations to probe determinants of affinity in receptor:chemokine interactions
- Analyzed molecular modeling data using Python to work towards proposing mutations on CCL5 (chemokine with anti-HIV properties) that increase binding affinity for CCR5 (receptor)

2017 - 2018 **Undergraduate Researcher**, *Ravi Majeti Lab*, Stanford University.
- Reprogrammed leukemia cells into antigen presenting cells by C/EBPα-induced transdifferentiation
- Gained experience with plasmid design, tissue culture, cloning, FACS, and lentiviral transduction
- Designed and executed experiments to analyze metabolic profiles throughout transdifferentiation

2016 - 2017 **Undergraduate Researcher**, *Stanford Space Initiative Biology Team*, Stanford University.
- Researched synthesis chemistry for solid-phase enzymatic DNA synthesis with TdT

2015 - 2016 **Research Assistant**, *Marjorie Solomon Lab*, UC Davis MIND Institute.
- Analyzed data and assisted with MRI scans as part of Autism Spectrum Disorder studies

Research Mentorship

PhD Students
- Abigail Jackson, Harvard-MIT
- Fiona Qu, Harvard Medical School
- Navami Jain, Harvard Medical School
- Tomas Grudny, MIT

Research Associates
- Ben Kotzen, Massachusetts General Hospital

Undergraduate Students
- Aarushi Mehrotra, MIT
- Seojean Kim, Wellesley
- Hailey Pan, MIT
- Sahil Sood, Harvard
- Omolivie Eboreime, Harvard
- Sage Widder, Wellesley

Skills and Coursework

Programming Languages
- Advanced Experience: Python
- Experience: Bash, R, C, C++, SQL, Spark, MongoDB

Libraries & Tools
- PyTorch, Sklearn, Git, Pandas, Numpy, Scipy, SLURM

Selected Coursework
- ML/NLP, Algorithms, Statistical Models in Biology, Matrix Theory, Systems, Databases
- Genetics, Cell and Molecular Biology, Organic Chemistry, Bioethics, Genomics