Sarah Gurev

Machine learning for viral and immune proteins

(916) 838-9672
sgurev@mit.edu
sarahgurev.github.io

Employment

Sept 2025 - 2026 **FutureHouse Independent Postdoctoral Fellow**.
Advisors: Sergey Ovchinnikov (MIT) and Aaron Schmidt (Massachusetts General Hospital)

Education

2020 - 2025 **Ph.D. in Electrical Engineering and Computer Science**, MIT.
Master of Science in EECS, MIT 2023.
Advisor: Debora Marks, Harvard Medical School, Dept. Systems Biology
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 (2025), **Computationally designed proteins mimic antibody immune evasion in viral evolution**, *Immunity.*
- **S. Gurev***, N. Youssef*, N. Jain, D. Marks (2025), **Variant effect prediction with reliability estimation across priority viruses**, *BioRxiv*.
- A. Shaw, H. Spinner, J. Shin, **S. Gurev**, N. Rollins, D. Marks (2023), **Removing bias in sequence models of protein fitness**, *BioRxiv.*
- R. Arora, M. Angelo , C. Choe, C. Shearer, ... **S. Gurev**, E. Xie, D. Marks, P. Notin (2024), **RNAGym: Large-Scale Benchmarks for RNA Fitness and Structure Prediction**, *BioRxiv.*
- 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:

- A. Mehrotra, N. Jain, N. Youssef, **S. Gurev**†, D. Marks† (2025), **Real-time Forecasting of Influenza Evolution**, *Machine Learning for Structural Biology & ICML Workshop on Generative AI and Biology & Molecular Machine Learning (MoML) Conference.*
  † indicates Senior authorship
- F. Qu, N. Youssef, **S. Gurev**†, D. Marks† (2025), **Recurring antibody motifs reveal germline-encoded interactions**, *MoML.*
- **S. Gurev***, N. Youssef*, H. Pierce-Hoffman, D. Marks (2024), **Future-proof vaccine design with a generative model of antibody cross-reactivity**, *ICLR and ICML Workshops*.
- **S. Gurev***, N. Youssef*, N Jain, D. Marks (2024), **Alignment-based and protein foundation models for viral evolution, vaccines and vectors**, *ICLR Workshop & Molecular Machine Learning Conference & Neurips Workshops* (Spotlight).

- **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)
- S. Salman, **S. Gurev**, M. Arsalan, F. Dar, A. Chan (2021), **Liver exchange: A pathway to increase access to transplantation**, *Harvard Health Policy Review*.

Funding and Awards

- FutureHouse AI-for-Science Independent Postdoctoral Fellowship (2025)
- 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

- Leveraging Global Data-Sharing Platforms: The Role of AI in the Future of Data Sharing. *Biosafety Level 4 Zoonotic Laboratory Network*, 2025.
- Benchmarking Models on Pandemic-Threat Viruses. *BioML Seminar at Berkeley*, 2025. `https://www.youtube.com/watch?v=so-0HBVpZJM`
- Variant Effect Prediction with Reliability Estimation Across Priority Viruses. *Gladstone Virology Institute*, 2025.
- Reliable mutation effect prediction across pandemic-threat viruses. *RosettaCon*, 2025.
- Reliable fitness prediction for priority viruses and benchmarks for RNA. *Merck*, 2025.
- Towards reliable mutation effect prediction across pandemic-threat viruses. *AI in Infectious Disease: Broad Public Health Symposium*, 2025.
- AI for Impact: Pandemic Prediction. *MIT Sloan AI Conference*, 2025.
- 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`
- Deep generative model foreshadows SARS-CoV-2 evolution and facilitates early vaccine evaluation. *American Society for Virology*, 2024.
- Learning from pre-pandemic data for the design and testing of variant-proof vaccines. *Contributed Talk (Best 4 papers)*, *Molecular Machine Learning Conference*, 2023.
- Early warning of viral antibody escape from a biologically-informed computational framework & Pandemic surveillance panel *Invited Talk*, *Precision Public Health Symposium*, 2023.

- 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.
- 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.

Posters

- **S. Gurev***, N. Youssef*, N. Jain, D. Marks, *Keystone Symposium on Predicting and Responding to Emerging Viral Infections*, Variant effect prediction with reliability estimation across priority viruses, 2025
- **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
- **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
- **S. Gurev**, J. Rodrigues, M. Levitt, *Stanford Bio-X Symposium*, Understanding determinants of affinity in receptor:chemokine interactions with molecular dynamics, 2018

Service

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

- Wrote new protein language model problem set, led office hours and recitations, mentor projects on ML for proteins, co-led writing of the exam, graded problem sets, exam, and projects
- Overall student rating 6.7/7
- "Was very clear during her lecture, gave us explanations that had been glossed over by other instructors for which I am very grateful. The PSET she wrote was also the most clear."

2024 **Teaching**, *Summer Antibody and Viral Initiative (SAVI) Workshop*, Harvard.

- Gave undergraduates at Harvard and Wellesley their first computational research experience
- Created and taught a summer-long, twice per week workshop to provide first-hand experience in using machine learning for viral mutation effect prediction
- Practical introduction where students learned by choosing their own pandemic-risk virus based on a current outbreak or area of interest to apply, refine, and evaluate the models

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

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

2021 - 2024 **Mentor**, MIT EECS and Harvard Systems Biology.

- Advise several newly admitted women graduate students (Thriving stars) and underrepresented undergraduates in research (New England Science Symposium)

2024 - 2025 **Lectures**.

- Advanced Computational Biology graduate course, MIT EECS: Created and delivered guest lectures on protein language models (2024 & 2025).
  `https://www.youtube.com/watch?v=uPoFdCUqBWk`
- One Health: Disease in an Interconnected World course, Tulane: Created and delivered guest lecture on AI for pandemic sequence prediction (2025).
- Next Epoch, Harvard: Assisted in teaching and developing material for a 3 day machine learning in biology tutorial for primarily first-gen college students (2024).
- Health Disparities Think Tank: Taught and created code tutorial for a undergraduate workshop on exploratory data analysis and data visualization (2024).
- Machine learning for Healthcare graduate course, MIT EECS: Created and delivered guest lectures on genomics in medicine (2024).

2023-2025 **Peer-Reviewer**, *Nucleic Acids Research, Nature Communications (2x) - co-review, PNAS - co-review, ICLR Workshop on Generative and Experimental Perspectives for Biomolecular Design (2x), ICML Workshop on ML for Life and Material Science, Mutational Scanning Symposium*.

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

- 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

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

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

Research Mentorship

PhD Student

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

RA
- Ben Kotzen, Massachusetts General Hospital

Undergraduate
- Aarushi Mehrotra, MIT
- Sarrah Leung, UC Berkeley
- Seojean Kim, Wellesley
- Hailey Pan, MIT

Research Experience

2020 - 2025 **PhD Student**, *MIT*, Debora Marks Lab, Harvard Medical School.
- Created the first model for early warning of pandemic viral variants that uses only pre-pandemic data
  - Model is more predictive of SARS-CoV-2 variants than high-throughput mutation scans measuring binding to patient pandemic antibodies, yet trained on data available more than **a year earlier**
- Early warning of SARS-CoV-2 variants: iteratively modeling and selecting top escape variants **at first sighting** to assay for infectivity/neutralization (biweekly reports for international health orgs)
  - Flagged many variants **months before** designation by WHO as variants of concern
- Called for industry shift to **testing vaccines not just on past variants** but also a panel of computationally designed variants that mimic the neutralizability of true future variants
- Developed the first model for **vaccine design that focuses antibody responses** to conserved regions more likely to be relevant for future variants

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