

GOWRI NAYAR

(+1) 561-603-0075 | gnayar@stanford.edu | www.linkedin.com/in/gowrinayar

I am a PhD candidate at Stanford University, under the supervision of Dr. Russ Altman. My research focuses on developing computational methods to decode complex protein pathways and predict context-dependent function at scale, through novel graph-based algorithms, multi-modal training and foundational AI models, aiming to address the unsolved challenges of proteomic science. With nine years of combined research experience, I have established a proven track record of defining high-impact research proposals, securing funding, and teaching graduate-level classes in AI for bioinformatics. Completing my PhD in Summer 2026, I seek a faculty role where I can lead a robust research program dedicated to bridging machine learning and proteomics while fostering the next generation of scientific talent.

EDUCATION

Stanford University

Sept 2021 - current

Ph.D. in Biomedical Informatics

Advised by Dr. Russ Altman. My research focuses on functional proteomics, developing machine learning methods to predict protein functional sites and functional groups through network methods. I have led teams of students to develop state-of-art algorithms using transformer and graph attention network methods. I have collaborated with wet-lab scientists to evaluate the functional predictions.

Georgia Institute of Technology

Sept 2016 - Dec 2017

M.Sc. in Computer Science

Strong foundations in algorithms, complexity theory, graph theory, linear algebra, statistics, and Machine Learning. Rigorous experience with proof-based classwork. Semester project in modeling language acquisition and accent detection to prove neuroscience hypothesis through Machine Learning techniques. Used by neuroscience laboratory to develop new language acquisition theories for experimental testing.

University of Pennsylvania

Sept 2012-May 2016

B.A. in Cognitive Science and Computer Science

Core courses in computer science, mathematics, logic, algorithms, complexity theory. Won best use of technology in PennApps Hackathon - application to match music characteristics of different genres to create optimal mixes.

RESEARCH AND PROFESSIONAL EXPERIENCE

Helix Lab, Stanford University

Sept 2021 - current

Graduate Researcher

Pathway-Level Representation Learning for Disease Mechanism Discovery

Developed a transformer-based model that treats proteins as tokens, pretrained with masked language modeling on random walks over protein interaction networks and fine-tuned on structured pathway graphs (Reactome DAGs) with hard negative sampling. Learned pathway embeddings are aligned with Human Phenotype Ontology representations to enable zero-shot prediction of disease-pathway associations.

Heterogeneous Protein Representation for Functional Discovery

Developed a graph-based framework using multi-relational networks and attention-based message passing to integrate protein-protein interactions, gene expression, and sequence features into unified embeddings. Implemented dynamic modality weighting, degree-controlled sampling, and inductive evaluation splits to improve generalization to understudied proteins and enable discovery of novel functional relationships. First author publication (conference proceedings).

Interpreting Functional Residues from Protein Language Models

Analyzed transformer-based protein language models by probing attention matrices and embedding structure across layers and heads to identify residue-level functional signals. Demonstrated that attention patterns encode catalytic and binding sites without supervision, enabling recovery of functionally important residues directly from sequence embeddings. First authored publication.

Predicting Subcellular Localization with Sequence-Interaction Embeddings and Matrix Decomposition

Developed models to predict protein localization as continuous Subcell vectors using sequence embeddings (e.g., ESM) and interaction-derived features. Applied matrix factorization against learned compartment centroids to infer mixture weights over cellular compartments, with a residual component to capture out-of-distribution classes such as secreted and transport proteins. Manuscript in preparation.

IBM Research, Almaden

Jan 2018 - Sept 2021

Research Engineer

Scalable Machine Learning on Large-Scale Bacterial Genomic Data

Designed and optimized machine learning algorithms on a 2 TB bacterial genomic dataset, developing scalable pipelines for processing and analyzing large-scale sequence data. Implemented efficient data handling and model training strategies to enable high-throughput genomic analysis. Co-first authored publication.

SARS-CoV-2 Mutation Dynamics and Diagnostic Impact Analysis

Analyzed SARS-CoV-2 genomic evolution to estimate a 3% monthly mutation rate and quantified its impact on diagnostic accuracy by modeling PCR primer mismatches. Predicted a 22% decrease in test efficacy over 19 months. First-author publication.

TEACHING

Mentorship, Stanford University

2025 - current

- Mentored second-year Ph.D student (Mohini Misra) for a protein benchmark project evaluating PLM embedding spaces' representation of fine-grained functional changes. Resulted in publication at ICLR 2026. (Fall 2025)
- Mentoring first-year Ph.D. student (Kae Nurge) in identifying proteins that have different localization patterns in different cell lines and benchmark how existing predictive models perform (Winter 2026 - current)
- Mentoring medical student (Sophia Pribus) in using PLM attention matrices to find druggable residues on protein with known cancer associations. Resulted in presentation at AACR and manuscript in preparation. (Winter 2025 - current)
- Mentoring medical student (Helen Xu) in developing a variant predictor model for CYP proteins to determine how a mutation affects abundance and function (Fall 2025 - current)
- Mentoring masters student (Riya Narain) in evaluating the overlap between PLM embedding features (principle components) and functional domains. Manuscript in preparation. (Fall 2025 - current)

Instructor for Computational Applications for High-Throughput Protein Data, Stanford University

Spring 2026

Developed and taught 3-week course designed for wet-lab biologists. The course serves as an introduction to computationally processing, analyzing, and training models with high-throughput proteomic data, including protein sequence data, expression and interaction data, and localization data. I developed the lectures to provide an introduction to the key computational methods and 3 coding homework assignments designed to serve as reference tutorials.

Guest Lecturer for Representations and Algorithms for Computational Molecular Biology, Stanford University

Fall 2024, Fall 2025

Held the lecture on Hidden Markov Models and their applications in representing protein families. Lectured to a class of 100+ students, mostly graduate level.

TA for Representations and Algorithms for Computational Molecular Biology, Stanford University

Fall 2024

Held weekly office hours to address students questions from lecture and on homework assignments. Worked with the professor to redesign the syllabus and structure the order of the lectures. Updated coding project assignments to reflect modern machine learning algorithms commonly used in the field. Redesigned the projects to teach using tools like HuggingFace, ESM-2, AlphaFold, and Graph Neural Networks.

TA for Linear Algebra, Georgia Institute of Technology

Fall 2017

Planned and taught weekly recitation classes of 15+ students. Held small group tutoring sessions and open office hours to provide extra assistance with material. Grade quizzes, homework, and staff-wide exam grading

TA for Mathematical Foundations of Computer Science, University of Pennsylvania 2014-2016

Planned and taught weekly recitation classes of 15+ students. Held small group tutoring sessions and open office hours to provide extra assistance with material. Grade quizzes, homework, and staff-wide exam grading

PUBLICATIONS

- Nayar G, Altman RB. **GATSBI: Improving context-aware protein embeddings through biologically motivated data splits.** ISMB Proceedings in Bioinformatics, Oxford University Press, July 2026. [\[Paper\]](#).
- Misra, Mohini, Nayar, Gowri, and Altman Russ B. **TripleThreat: Benchmarking Functional Sensitivity in Protein Representations with Paralog-Ortholog Triplets.** Learning meaningful representation of life workshop at ICLR, Apr. 2026. [\[Paper\]](#).
- Pribus, Sophia J., Nayar, Gowri, et al. **Interpreting PLMs for Cancer Discovery: High Attention Hotspots Predict Pathogenic Mutation Positions and Novel Drug Binding Sites.** Cancer Research, vol. 86, no. 7Supplement, American Association for Cancer Research, Apr. 2026. [\[Paper\]](#).
- Nayar, Gowri, Alp Tartici, et al. **Paying Attention to Attention: High Attention Sites as Indicators of Protein Family and Function in Language Models.** PLoS Computational Biology, vol. 21, no. 9, Public Library of Science (PLoS), Sept. 2025, p. e1013424. [\[Paper\]](#).
- Nayar, Gowri, Kristy A. Carpenter, et al. **Exposing the Molecular Reaction Blind Spots of LLMs with PathwayQA.** bioRxiv, Aug. 2025, (*in review at Bioinformatics, Oxford University Press*) [\[Paper\]](#).
- Tartici, Alp, Nayar, Gowri, Altman, RB. **Pool PaRTI: A PageRank-Based Pooling Method for Identifying Critical Residues and Enhancing Protein Sequence Representations.** Bioinformatics (Oxford, England), Oxford University Press (OUP), June 2025. [\[Paper\]](#).
- Nayar, Gowri, and Russ B. Altman. **Heterogeneous Network Approaches to Protein Pathway Prediction.** Computational and Structural Biotechnology Journal, Dec. 2024. [\[Paper\]](#).
- Yang, L, Bolton, E, Nayar, G, Altman, RB **Longitudinal Modeling of Multimorbidity Trajectories Using Large Language Models.** medRxiv, Oct. 2024. [\[Paper\]](#).
- Nayar, Gowri*, Cousins, Henry C.*, Altman, Russ B. **Computational Approaches to Drug Repurposing: Methods, Challenges, and Opportunities.** Annual Review of Biomedical Data

Science, Annual Reviews, Aug. 2024. [\[Paper\]](#).

- **Nayar, Gowri**, Ignacio Terrizzano, et al. **ggMOB: Elucidation of Genomic Conjugative Features and Associated Cargo Genes across Bacterial Genera Using Genus-Genus Mobilization Networks**. Frontiers in Genetics, Frontiers Media SA, Dec. 2022. [\[Paper\]](#).
- Agarwal, Akshay,..., **Nayar, Gowri**,..., Kaufman James H. **Predicting Epitope Candidates for SARS-CoV-2**. Viruses, MDPI AG, Aug. 2022. [\[Paper\]](#).
- Beck, Kristen L., ..., **Nayar, Gowri**,..., Kaufman James H. **Semi-Supervised Pipeline for Autonomous Annotation of SARS-CoV-2 Genomes**. Viruses, MDPI AG, Dec. 2021. [\[Paper\]](#).
- **Nayar, Gowri**, Edward E. Seabolt, et al. **Analysis and Forecasting of Global Real Time RT-PCR Primers and Probes for SARS-CoV-2**. Scientific Reports, Springer Science and Business Media LLC, Apr. 2021. [\[Paper\]](#).
- **Nayar, Gowri***, Seabolt, Edward E.*, et al. **IBM Functional Genomics Platform, A Cloud-Based Platform for Studying Microbial Life at Scale**. arXiv [q-Bio.QM], Nov. 2019. [\[Paper\]](#).

INVITED TALKS

- **GATSBI: Heterogeneous Protein Embeddings**, ISMB Conference, Washington, D.C., 2026
- **Data Splits Dictate the Utility of Models**, HAI Annual Conference - AI+Science: Accelerating Discovery, Stanford, CA, 2026
- **Predicting Gain-of-Function Variants in Fibroblast Growth Factor Receptors**, Critical Assessment of Genome Interpretation 7 (CAGI7), Boston, MA, 2025
- **Paying Attention to Attention**, Keystone Symposia: AI in Molecular Biology, Santa Fe, NM, 2025 (Poster)
- **Paying Attention to Attention**, PepTalk and Biologic Summit, San Diego, CA, 2024

AWARDS

Stanford Data Science Fellowship	2025-2027
NIH F31 Fellowship	2024-2026
TA recognition award, Stanford Department of Biomedical Data Science	Spring 2025
Patent US12456542B2 - Material representation in computational systems	2025
Donald V. Jackson Fellowship, Georgia Institute of Technology	2017

ACADEMIC AND COMMUNITY SERVICE

Manuscript Peer Review

Nature Biomedical Engineering; Journal of Proteome Research

Departmental Service

- Student Representative on Ph.D. Admissions Panel (2024)
- Graduate Student Representative for Department of Biomedical Data Science (2023-2024)
- Member of DEI Committee for Department of Biomedical Data Science (2022-current)
- Peer2Peer Mentor: Mentor unrepresented applicants in writing statement of purpose (2022-current)

SKILLS

Programming Languages and Frameworks

Python (PyTorch, PyTorch Geometric, Pandas NumPy, Scikit-learn. etc.), training on multi-GPU cluster, data/model parallelism, PyTorch DDP, Weights & Biases, Slurm, HPC clusters, multi-process, multi-threaded pipelines, SQL, Google Cloud, Azure, AWS, Git

Soft Skills

Leadership, Mentoring, Public speaking, Communication, Adaptability, Self-motivation