Garyk Brixi

gbrixi@stanford.edu • github • scholar

Education

Stanford University 2023 – present
Ph.D. student, Genetics. Brian Hie Lab. Stanford, CA
Awards: NSF GRFP Fellow, Hertz Finalist

Harvard University 2019 – 2023
B.A. in Applied Math, secondary in Computer Science Cambridge, MA
Awards: Presidential Scholar, Davidson Fellow, Barron Prize, Harvard i-Lab Impact Fellow

Experience

Hie Lab, Arc Institute and Stanford University March 2024 – Present

- Training frontier scale language models for genomics with domain informed recipes.
- Developing pre and post-training strategies for discovery using genome language models.

PhD Research Intern, Microsoft Research June - August 2024

- Meta-learning protein language models using supervised data to improve fitness prediction.
- Advised by Kevin Yang and Ava Amini.

Ovchinnikov Lab, Harvard University Jan 2023 – 2024

- Developed energy model adapter for protein language models with inherent interpretability.
- Probing and interpreting protein language models, connecting LMs with statistical models.

Chatterjee Lab, Duke University & Church Lab, Wyss Institute Jan 2022 – Dec 2022

- Created method for designing protein binders from natural interactions.
- Our method has enabled collaborators to degrade disease proteins.

Liang Lab, Harvard School of Public Health Jan 2021 – Dec 2022

- Discovered new genetic loci associated with PCOS.
- Used genetics of related traits to improve prediction of PCOS.

Vision Science Lab, Harvard University May 2020 - Jan 2021

- Comparing deep learning vision model activations with visual neurons in mammal brains.
- Contributed to development and use of DeepDive package for neural probing.

Manuscripts

Genome modeling and design across all domains of life with Evo 2 **Garyk Brixi***, Matthew G Durrant*, Jerome Ku*, Michael Poli*, Greg Brockman, Daniel Chang, Gabriel A Gonzalez, Samuel H King, David B Li, Aditi T Merchant, Mohsen Naghipourfar, Eric Nguyen, Chiara Ricci-Tam, David W Romero, Gwanggyu Sun, Ali Taghibakshi, Anton Vorontsov, Brandon Yang, ..., Dave P Burke, Hani Goodarzi, Patrick D Hsu, Brian L Hie. bioRxiv. 2025.02.18.638918. 2025.

Systems and Algorithms for Convolutional Multi-Hybrid Language Models at Scale Jerome Ku*, Eric Nguyen*, David W. Romero*, **Garyk Brixi**, Brandon Yang, Anton Vorontsov, Ali Taghibakhshi, Amy X Lu, Dave P Burke, Greg Brockman, Stefano Massaroli, Christopher Ré, Patrick D Hsu, Brian L Hie, Stefano Ermon, Michael Poli. Arxiv. 2025.

Protein language models learn evolutionary statistics of interacting sequence motifs Z. Zhang*, H. Wayment-Steele*, **G. Brixi**, H. Wang, D. Kern, M. D. Peraro, S. Ovchinnikov. Proceedings of the National Academy of Science. 2024.

Sequence modeling and design from molecular to genome scale with Evo E. Nguyen*, M. Poli*, M. G Durrant*, B. Kang*, D. Katrekar*, D. B Li*, L. J Bartie, A. W Thomas, S. H King, **G. Brixi**, J. Sullivan, M. Y Ng, A. Lewis, Aaron Lou, S. Ermon, S. A Baccus, T. Hernandez-Boussard, C. Ré, P. D Hsu, B. L Hie. Science. 2024.=

SaLT&PepPr is an interface-predicting language model for designing peptide-guided protein degraders **G. Brixi***, T. Ye*, L. Hong*, T. Wang, C. Monticello, N. Lopez-Barbosa, S. Vincoff, V. Yudistyra, L. Zhao, E. Haarer, T. Chen, S. Pertsemlidis, K. Palepu, S. Bhat, J. Christopher, X. Li, T. Liu, S. Zhang, L. Petersen, M. P. DeLisa & P. Chatterjee. Communications Biology. 2023.

Transferrin Receptor Targeting Chimeras (TransTACs) for Membrane Protein Degradation D. Zhang, J. Duque-Jimenez, F. Facchinetti, **G. Brixi**, K. Rhee, W. W. Feng, P. A. Jänne, & X. Zhou. Nature, 2024.

Understanding and Predicting Polycystic Ovary Syndrome through Shared Genetics with Testosterone, SHBG, and Chronic Inflammation. L. Petersen*, **G. Brixi***, J. Li, J. Hu, Z. Wang, X. Han, A. Yaskolka Meir, Jaakko Termi, S. Mahalingaiah, T. Piltonen, L. Liang. MedRxiv. Under peer-review.

Efficient Human Germ Cell Specification from Stem Cells via Combinatorial Expression of Transcription Factors. C. Kramme, M. Smela, B. Wolf, P. Fortuna , **G. Brixi**, K. Palepu, E. Dong , J. Adams, S. Bhat, S. Koseki, Emma Tysinger, T. Stan, R. Kohman, S. Liu , M. Kobayashi, T. Shioda, G. Church, P. Chatterjee. 2022. bioRxiv and under peer-review.

Directed Differentiation of Human iPSCs to Functional Ovarian Granulosa-Like Cells via Transcription Factor Overexpression. M. Smela , C. Kramme , P. Fortuna, J. Adams, E. Dong, M. Kobayashi, **G. Brixi**, E. Tysinger, R. Kohman, P. Chatterjee, T. Shioda, G. Church. 2022. eLife.

Design of Peptide-Based Degraders via Contrastive Learning. K. Palepu, M. Ponnapati, S. Bhat, E. Tysinger, T. Stan, **G. Brixi**, S. R.T. Koseki, P. Chatterjee. 2022. bioRxiv.

Ready-to-Use Therapeutic Food (RUTF) Containing Low or No Dairy Compared to Standard RUTF for Children with Severe Acute Malnutrition: A Systematic Review and Meta-Analysis. I. Potani, C. Spiegel-Feld, **G. Brixi**, J. Bendabenda, N. Siegfried, R. HJ Bandsma, A. Briend, and A. I. Daniel. Advances in Nutrition. 2021.

Developing Local Treatments for Acutely Malnourished Children in Sub-Saharan Africa: An Optimization Approach Automatically Ensuring Protein Quality. **Garyk Brixi**. Current Developments in Nutrition, 2019.

Innovative optimization of ready to use food for treatment of acute malnutrition. **Garyk Brixi**. Maternal and Child Nutrition. 2018.

Conferences

Molecular Machine Learning Conference. Poster, 2022: Design of Peptide-Guided Protein Degraders with Structure-Agnostic Language Models.

American Society for Human Genetics. Poster, 2022: Cross-trait meta-analysis reveals shared genetic architecture between PCOS and chronic inflammation markers.

American Society for Nutrition. Lightning talk. 2019: Developing Local Treatments for Acutely Malnourished Children in Sub-Saharan Africa: An Optimization Approach Automatically Ensuring Protein Quality.

Additional Projects

Natural Language Processing for Gene Expression Prediction, Inari Spring 2022

- Developed deep learning models to predict plant gene expression from DNA sequence.
- Identified and fixed a domain-specific data leak in SOTA model, presented company wide.

Data Augmentation for Machine Translation, Dr. Chris Tanner Fall 2022

- Developed syntax based data augmentation methods for translation, benchmarked with FAIRSEQ.

Explainable ML to analyze Epigenetic Markers, Prof. Jason Buenrostro Spring 2022

- Using statistical inference and explainable machine learning to characterize epigenetic markers in hematopoiesis, identifying state lineage signatures.

Volunteering

Malnutrition Treatment Development, Valid Nutrition 2017-2023

- Developed optimization tool for acute malnutrition treatment, reducing ingredient costs by 50%.
- Coordinated and raised money for piloting and testing of new therapeutic foods in Kenya and Malawi.

Acute Malnutrition Treatment Evidence Review Team, World Health Organization 2020 - 2021

- Analysis informed the current WHO international guidelines for acute malnutrition treatment.

Relevant Coursework

Computer Science

- Deep Learning for NLP
- Systems Programming
- Artificial Intelligence
- Machine Learning with Graphs
- Artificial Intelligence
- Intro to Theoretical CS
- Data Structures and Algorithms
- Foundations of Parallel Computing
- Capstone in Applied Computation

Math and Statistics

- Linear Algebra
- Multivariable Calculus
- Probability
- Statistical Inference
- Differential Equations

Biology

- Molecular Medicines
- Chromosomes
- Genetic Association Studies
- Virology
- Epigenetics and Gene Regulation
- Experimentation and Imaging