MOONSEONG (BRONSON) JEONG

Los Angeles, CA 90025 | (310) 701 8190 | bronsonj@cs.ucla.edu
bronsonj98.github.io | Google Scholar | ORCID 0009-0009-3836-2676

Research Focus

Ph.D. candidate in Computer Science at UCLA developing scalable statistical and machine-learning methods for genomics, medical imaging, and electronic health records. Current research centers on privacy-preserving summary-statistics methods for genomic inference across genetic ancestries (SUMMIT); the recognition of disease in health records and its consequences for case finding and genetic analysis; compute-efficient analysis of 3D medical images; and secure, interpretable analytics for prescription-monitoring systems. Across these areas, my work emphasizes practical use of sensitive, high-dimensional data under real computational and data-governance constraints.

Education

University of California, Los Angeles (UCLA) — Los Angeles, CA
Doctor of Philosophy in Computer Science — September 2022 – June 2028 (expected)
Advisor: Sriram Sankararaman
Master of Science in Computer Science — March 2025
Bachelor of Science in Mathematics of Computation; Bachelor of Arts in Physics — December 2020

Research Experience

Summary-Statistics Genetics and Privacy-Preserving Biobank Analysis — September 2022 – Present
UCLA; advisor: Sriram Sankararaman

- Developed SUM-RHE, a summary-statistics implementation of randomized Haseman–Elston regression that attains accuracy comparable to individual-level analysis while substantially reducing computational cost and avoiding routine access to individual-level genotype data (*Genome Research*, 2024).
- Led SUMMIT, a framework that combines genome-wide LD scores, estimated with random projections, with GWAS summary statistics to estimate genome-wide, annotation-specific, and cross-trait heritability and genetic correlation; in UK Biobank simulations across European, South Asian, and admixed African-European cohorts it was more accurate and less biased than prior methods, with the largest gains outside European ancestry (bioRxiv, 2026; selected for a platform presentation at ASHG 2026).
- Built a high-performance Python/C++ pipeline for heritability and genetic-correlation estimation across hundreds of traits and diverse biobank-scale cohorts; extended the framework to multivariate inference, gene–environment interaction, and partitioned genetic architecture, contributing to a 2026 *Genome Research* article (ENGINE), a 2026 medRxiv preprint (REGENA), and a 2025 bioRxiv preprint on rare coding-variant heritability (FLEX).
- Extending SUMMIT to keep related and admixed participants in the analysis, removing the phenotypic resemblance of relatives from the estimating equations and modeling shared local ancestry with a second kernel, so that small ancestry groups lose no participants (manuscript in preparation); composing the LD of admixed cohorts from ancestral reference panels and summary statistics (manuscript in preparation).
- Applying summary-level inference to retinal OCT biomarkers to study the genetic architecture of age-related macular degeneration across diverse US cohorts in collaboration with Dr. Maryam Zekavat (Harvard Medical School), Dr. Oren Avram (Optum Labs), and Dr. Adrian Au (UCLA Stein Eye Institute).

Recognition of Disease in Electronic Health Records — 2025 – Present
UCLA; advisor: Sriram Sankararaman

- Using UK Biobank participants whose enrollment blood assays were in the disease range for diabetes, chronic kidney disease, or anemia, measured how often disease-range values had no diagnosis or treatment record, how slowly they were recognized in primary-care and hospital records, and how case-finding models trained on recorded diagnoses compare with models trained on the measured assays (ML4H 2026 Proceedings, first author).
- Extending this work to the heritability of being diagnosed and its effect on genetic correlations estimated between diagnosis codes (manuscript in preparation).

Compute-Efficient 3D Medical Imaging and Imaging Genomics — October 2024 – Present
UCLA; advisors: Sriram Sankararaman and Yuzhe Yang

- Co-developed RAPTOR, a train-free framework that uses frozen two-dimensional foundation models and random projections to construct compact representations of three-dimensional medical volumes without training a new volumetric model.
- Evaluated RAPTOR across 10 medical-imaging tasks; it achieved the strongest performance among evaluated methods in 7 datasets and processed a full volume in approximately 6.5 seconds on a single 11 GB graphics card (ICML 2025 Spotlight).
- Applied volumetric embeddings to UK Biobank brain MRI data, identifying 126 novel genetic associations while replicating 111 of 118 previously reported image-derived phenotype associations; now extending this work to retinal OCT biomarkers and age-related macular degeneration in governed biobank settings.

Secure Prescription-Monitoring Analytics in California CURES September 2022 – Present
UCLA / California Department of Justice Research Center; advisors: Sriram Sankararaman and Siamak Rahman

- Serve as technical lead on an annually renewed, California Department of Justice-approved collaboration involving more than 30 million individuals, approximately 400,000 prescribers, and over 200 million controlled-substance prescriptions.
- Built the analysis-ready pipelines, feature-construction workflows, and statistical and machine-learning analyses required to study prescribing patterns under strict privacy, security, and on-premises operational constraints.
- Analyzed geographic, demographic, specialty, product, and policy-related variation in controlled-substance exposure and developed interpretable monitoring outputs; preliminary findings were presented at the International Association for the Study of Pain in 2024, and a manuscript is in preparation.

Electronic Health Record Modeling and Clinical Prediction June 2023 – February 2025
UCLA; advisors: Sriram Sankararaman and Jeffrey Chiang

- Co-developed DK-BEHRT, which incorporates disease-description embeddings from biomedical language models into a BERT-based architecture for electronic health records, yielding improvements of up to 62% in AUPR and 21% in AUROC over baseline models (AAAI Bridge Program on AI for Medicine and Healthcare, 2025).
- Developed a model combining electronic health record and census information to predict patient demographic features (AUROC 0.89) and performed ablation studies to quantify the contribution of each feature group.

Coalescent Limits of Local Ancestry Inference July 2022 – Present
Joint project with Xinzhu (April) Wei, Cornell University

- Derived a coalescent-theoretic framework for when reference-based local ancestry inference is reliable, as a function of admixture timing relative to source divergence, reference-panel size, proxy references, linkage along the genome, and gene flow.
- Implemented the framework in C++, MATLAB, and Python; validated it against coalescent simulations and existing local ancestry software; presented the work at the Cold Spring Harbor Probabilistic Modeling in Genomics meeting in 2023; manuscript in preparation.

MOSDEF Survey: High-Redshift Galaxy Spectroscopy June 2018 – September 2020
UCLA Physics and Astronomy; advisor: Alice Shapley

- Applied Monte Carlo methods, Gaussian mixture models, and emission-line analysis to characterize neon emission in high-redshift galaxies, resulting in a first-authored paper in *The Astrophysical Journal Letters*.

Publications

Peer-Reviewed Journal Articles

1. Liu, Z., Ramteke, A., Anand, A., Gorla, A., **Jeong, M.**, and Sankararaman, S. "A biobank-scale method for learning modulators of gene-environment interaction underlying human complex traits from multiple environmental exposures." *Genome Research* (advance online publication), 2026. DOI: 10.1101/gr.282105.126.
2. **Jeong, M.**, Pazokitoroudi, A., Liu, Z., and Sankararaman, S. "Scalable summary-statistics-based heritability estimation method with individual genotype level accuracy." *Genome Research* 34(9):1286–1293, 2024. DOI: 10.1101/gr.279207.124.
3. **Jeong, M.**, Shapley, A. E., Sanders, R. L., et al. "The MOSDEF Survey: Neon as a Probe of ISM Physical Conditions at High Redshift." *The Astrophysical Journal Letters* 902:L16, 2020. DOI: 10.3847/2041-8213/abba7a.

Peer-Reviewed Conference and Proceedings Papers

4. **Jeong, M.**, Anand, P., Lee, S. A., Han, S., Liu, Z., and Sankararaman, S. "The Label Is Not the Disease: Recognition Delay in Diagnosis Codes and Its Cost to EHR Case Finding." *Proceedings of the Machine Learning for Health*

Symposium (ML4H 2026), PMLR 346, 2026 (accepted; to appear). Code: github.com/bronsonj98/recognition-delay.

5. An, U.*, **Jeong, M.***, Lee, S. A., Gorla, A., Yang, Y., and Sankararaman, S. "RAPTOR: Scalable Train-Free Embeddings for 3D Medical Volumes Leveraging Pretrained 2D Foundation Models." *Proceedings of the 42nd International Conference on Machine Learning*, PMLR 267:1462–1482, 2025. [link] Spotlight paper (top 2.6% of submissions); *equal contribution.
6. An, U.*, **Jeong, M.***, Lee, S. A., Gorla, A., and Sankararaman, S. "Train-free unsupervised volumetric embeddings using image foundation models enable improved predictions and genomic discoveries." RECOMB-Genetics, 2025. *Equal contribution.
7. An, U., Lee, S. A., **Jeong, M.**, Gorla, A., Chiang, J. N., and Sankararaman, S. "DK-BEHRT: Teaching Language Models International Classification of Disease (ICD) Codes Using Known Disease Descriptions." *Proceedings of the First AAAI Bridge Program on AI for Medicine and Healthcare*, PMLR 281:133–143, 2025. [link]

Preprints

8. **Jeong, M.**, Liu, Z., Ni, J., An, U., Han, S., Anand, A., Anand, P., Boffelli, M., and Sankararaman, S. "Genome-wide LD scores enable robust heritability and genetic correlation estimation from summary statistics across genetic ancestries." bioRxiv, 2026. DOI: 10.64898/2026.10.20.760508. Software: github.com/sriramlab/SUMMIT.
9. Liu, Z., Anand, A., Qie, X., Costantino, M., **Jeong, M.**, Gorla, A., Zaitlen, N., Dahl, A., and Sankararaman, S. "Disentangling phenotype scale, amplification, and effect heterogeneity underlying polygenic GxE architecture of complex traits." medRxiv, 2026. DOI: 10.64898/2026.10.15.26365681.
10. Liu, Z., Fu, B., **Jeong, M.**, Anand, P., Anand, A., Jang, S.-K., Gorla, A., Zhu, J., Pajukanta, P., Palamara, P. F., Zaitlen, N., Border, R., and Sankararaman, S. "Comprehensive gene heritability estimation reveals the genetic architecture of rare coding variants underlying complex traits." bioRxiv, 2025. DOI: 10.1101/2025.10.07.681018.

Manuscripts Under Review

11. Liu, Z.*, Anand, A.*, **Jeong, M.**, Wang, R., Gorla, A., and Sankararaman, S. "multiGENIE reveals the exposome architecture of gene–environment interaction at biobank scale." Under review. *Equal contribution.
12. Anand, P., An, U., **Jeong, M.**, Gorla, A., and Sankararaman, S. "Neural networks complement hidden Markov models in an ensemble for genotype imputation." Under review. Code: github.com/prateekanand2/CrossBlend.
13. Anand, P., **Jeong, M.**, and Sankararaman, S. "Base Composition Confounds Classifier-Free Guidance in DNA Diffusion Models." Under review.

Manuscripts in Preparation

14. **Jeong, M.**, Han, S., Anand, P., and Sankararaman, S. "Heritability from summary statistics of admixed and related biobank participants."
15. **Jeong, M.** and Sankararaman, S. "Joint estimation of gene-by-context covariance across many exposures in biobanks."
16. **Jeong, M.** and Sankararaman, S. "Being diagnosed is heritable and inflates genetic correlations between diseases."
17. **Jeong, M.** and Sankararaman, S. "Conventions for treated participants change genetic analyses of cardiometabolic traits."
18. **Jeong, M.** and Sankararaman, S. "Linkage disequilibrium of admixed cohorts from ancestral reference panels and GWAS summary statistics."
19. **Jeong, M.**, Sankararaman, S., and Wei, X. "Coalescent limits on the accuracy of reference-based local ancestry inference."
20. **Jeong, M.**, Chiu, A. M., An, U., Avram, O., Zekavat, S., and Sankararaman, S. "Heritability analysis of automatically extracted OCT biomarkers in the UK Biobank."
21. **Jeong, M.**, Fu, B., Rahman, S., and Sankararaman, S. "Describing California's controlled opioid prescription landscape with the CURES Dataset: Descriptive statistics and machine-learning explorations."
22. An, U., Gorla, A., **Jeong, M.**, and Sankararaman, S. "Gaussian process-based neural topic model for series-specific topic dynamics in time series datasets."

Selected Conference Presentations

Oral presentations

- "Genome-wide LD scores allow robust heritability and genetic correlation from summary statistics across ancestries,"

American Society of Human Genetics (ASHG) 2026 Annual Meeting, platform session "From Labels to Landscapes: Leveraging Local Ancestry and the Ancestry Continuum to Maximize Cross-Ancestry Association," Montréal, October 22, 2026.

- "Train-free unsupervised volumetric embeddings using image foundation models enable improved predictions and genomic discoveries," RECOMB-Genetics, Seoul, 2025.
- "Scalable summary-statistics-based heritability estimation method with individual genotype level accuracy," RECOMB, Boston, 2024.

Co-authored presentations

- "Comprehensive gene heritability estimation reveals the genetic architecture of rare coding variants underlying complex traits," RECOMB-Genetics 2026, Thessaloniki, May 2026 (presented by first author Zhengtong Liu).

Poster presentations

- "The Label Is Not the Disease: Recognition Delay in Diagnosis Codes and Its Cost to EHR Case Finding," Machine Learning for Health (ML4H) Symposium, Sydney, December 2026 (forthcoming).
- "RAPTOR: Scalable Train-Free Embeddings for 3D Medical Volumes Leveraging Pretrained 2D Foundation Models," ICML Spotlight, Vancouver, 2025.
- "DK-BEHRT: Teaching Language Models International Classification of Disease Codes Using Known Disease Descriptions," AAAI Bridge Program on AI for Medicine and Healthcare, Philadelphia, 2025.
- "Epidemiology of Continuous Opioid Use in California Over the Past Decade," International Association for the Study of Pain, Amsterdam, 2024.
- "Scalable summary-statistics-based heritability estimation method with individual genotype level accuracy," American Society of Human Genetics, Washington, DC, 2023.
- "The theoretical limit of local ancestry inference," Cold Spring Harbor Probabilistic Modeling in Genomics, Cold Spring Harbor, NY, 2023.

Peer Review and Professional Service

Ad hoc and invited reviewer for *Cell Genomics*, *Nature Communications*, *Bioinformatics*, *BMC Genomics*, RECOMB 2026, ML4H 2026, and ICLR 2027.

Teaching Experience

Teaching Assistant, UCLA Computer Science 2023 – Present

- C121/C221: Probabilistic Models in Computational Genomics (Fall 2026).
- CS31: Introduction to Computer Science I (Fall 2025).
- M146: Introduction to Machine Learning (Winter 2024 and Winter 2025).
- CM124/224: Machine Learning Applications in Genetics (Fall 2023).

Awards and Honors

Platform (oral) presentation, ASHG 2026 Annual Meeting 2026

- Abstract selected for an oral presentation in a platform session of the American Society of Human Genetics Annual Meeting, Montréal.

ICML Spotlight selection for RAPTOR 2025

- Selected as a Spotlight paper (top 2.6% of submissions).

UCLA Computer Science Departmental Fellowship 2022 – Present
UCLA Dean's Honors List Winter 2019; Spring 2020

Technical Expertise

Programming and systems: Python, C++, Bash, SQL, MATLAB, and R; PyTorch and TensorFlow; high-performance computing and Slurm; large-scale experimentation, profiling, and performance optimization.

Research methods: statistical genetics, variance-component estimation, genome-wide association analysis, gene–environment interaction, biobank-scale computation, machine learning, three-dimensional medical-image analysis, electronic health record modeling, and secure on-premises health-data pipelines.