

# Xinyu (Brian) Guo

(314) 680-8961 | xinyug@usc.edu | <https://www.xinyuguo.com>

## PROFESSIONAL SUMMARY

AI Researcher and Applied Scientist with Ph.D. in Computational Biology (USC) and experience in large language models (LLMs), self-supervised learning (SSL), and graph neural networks (GNNs). Developed models spanning biological data modeling, natural language processing, and computer vision. Passionate about building interpretable, efficient foundation models and AI systems that bridge foundational research and real-world impact.

## TECHNICAL EXPERTISE

**Machine Learning:** Representation Learning, Self-supervised learning (BYOL, DINO), Contrastive Learning, Transformers, GNNs, Retrieval-Augmented Generation (RAG)

**Languages:** Python, R, SQL, Java, HTML/CSS, JavaScript/TypeScript

**Frameworks:** PyTorch, PyTorch Geometric, TensorFlow, HuggingFace Transformers, CUDA, OpenCV

**Tools:** Git, Docker, AWS/GCP, Supabase, Next.js, Tailwind CSS , Linux

**Domains:** NLP, Computer Vision, Bioinformatics, AI Systems Engineering

## RESEARCH & PROJECTS

### LLM Training from Scratch (2025)

- Trained a transformer-based decoder-only large language model (~560M parameters) from scratch using Byte Pair Encoding (BPE) tokenization and distributed PyTorch (DDP) training.
- Implemented scalable multi-GPU training across 8×A100 GPUs, optimizing gradient synchronization and memory efficiency.
- Fine-tuned conversational behavior using a mixture of synthetic dialogue and curated domain datasets, improving response coherence and factuality in evaluation benchmarks.

### Self-Supervised Learning Evaluation for Spatial Transcriptomics (2025)

- Benchmarked 8 SSL methods across 3 GNN architectures on spatial transcriptomics data.
- Evaluated contrastive and non-contrastive SSL approaches (SimCLR, BYOL, Barlow Twins, etc.) for both brain specific datasets and across developmental tissue datasets.
- Quantified model clustering performance and consistency using spatial auto correlation and embedding variance metrics.

### Legal Patent Copilot (2024)

- Built an LLM-powered patent search system integrating RAG and vector embeddings for efficient prior-art retrieval.
- Deployed sentence transformers over full patent corpus (~10M+ documents) with FAISS-based semantic indexing.
- Reduced search latency by 42% and improved recall quality via fine-tuned reranking layers.

### Intelligent Tennis Coach (2024)

- Developed a tennis analytics platform detecting ball and player trajectories using YOLO + optical flow tracking.
- Built temporal sequence model for action detection, enabling real-time stroke classification and player feedback.
- Processed 100,000+ rally videos for multi-modal learning and real-world evaluation.
- Integrated fine-tuned LLM to generate personalized coaching insights from match statistics.

### DualViewGNN: TWAS-guided Spatial Representation Model (2023)

- Developed a dual-view GNN integrating spatial transcriptomics with GWAS/TWAS-derived genetic risk signals.
- Introduced TWAS-weighted message-passing mechanism improving spatial coherence of risk embedding clusters.
- Results contributed to new insights on tissue-specific disease risk regions.

## **EXPERIENCE**

### **Research Assistant, Statistical Genetics & Genomics Studies | University of Southern California**

Advisor: Dr. Liang Chen | Aug 2022 - Present

- Designed GNN-based representation learning models for gene expression and regulatory network inference.
- Analyzed multi-omics datasets to map genetic variants to cell-type-specific expression dynamics.
- Identified novel potential genetic drivers of nonsense-mediated mRNA decay with downstream regulatory implications.

### **Research Assistant, Statistical Genetics Studies | Johns Hopkins University**

Advisors: Dr. Nilanjan Chatterjee & Dr. Diptavo Dutta | Oct 2020 - Aug 2022

- Developed CSTWAS, a subset-based cross-tissue meta-analysis method for transcriptome-wide association studies.
- Performed transcriptome-wide and single cell level association studies to identify genes linked to renal cell carcinoma.
- Improved interpretability and power for multi-tissue gene-trait associations.

### **Research Assistant, Statistical Genomics Studies | Johns Hopkins University**

Advisors: Dr. Hongkai Ji & Dr. Eneda Toska | Feb 2021 - Aug 2022

- Created RNA-seq/ChIP-seq/ATAC-seq pipelines supporting differential analysis and motif discovery.
- Elucidated how transcriptional and epigenetic regulators mediate hormone-driven cancer tumor resistance to therapeutics, with findings published in *Cell Reports* (2024), *Oncogene* (2023), and *Cancer Research* (2023).

### **Research Assistant, Respiratory Failure Prediction | Washington University in St. Louis**

Advisor: Dr. Andrew Michelson | Jun 2020 - Dec 2020

- Built predictive models (lasso/logistic regression) to identify respiratory failure in COVID-19 patients.
- Model implemented in hospital EHR system, improving early risk detection and timely ICU resource deployment.

## **SELECTED PUBLICATIONS**

- Guo, X., Chen, L. From G1 to M: a comparative study of methods for identifying cell cycle phases. *Briefings in Bioinformatics*.
- Bianchi, A., ..., Guo, X., ..., Merchant, N. Cell-autonomous Cxcl1 sustains tolerogenic circuitries and stromal inflammation via neutrophil-derived TNF in pancreatic cancer. *Cancer Discovery*.
- Dutta, D., Guo, X., et al. Transcriptome-and proteome-wide association studies identify genes associated with renal cell carcinoma. *The American Journal of Human Genetics*.
- Guo, X., Chatterjee, N., Dutta, D. Subset-based method for cross-tissue transcriptome-wide association studies improves power and interpretability. *Human Genetics and Genomics Advances*.

## **EDUCATION**

### Ph.D., Computational Biology and Bioinformatics

University of Southern California | GPA: 3.95/4.0 | Aug 2022 – Present

- Andrew J. Viterbi Fellowship for outstanding Ph.D. students in computational sciences

### M.S., Biostatistics

Johns Hopkins University | GPA: 3.97/4.0 | Aug 2020 – May 2022

- Delta Omega Honorary Society in Public Health

### B.A., Mathematics and Computer Science

Washington University in St. Louis | GPA: 3.89/4.0 | Aug 2018 – May 2020

- Cum Laude, Dean's List