WUXINHAO (TIM) CAO

Boston, MA | +1-857-361-6682 | tim_cao@hsph.harvard.edu | LinkedIn | ResearchGate

---

EDUCATION

HARVARD UNIVERSITY   Boston, MA
- Master of Science, Computational Biology and Quantitative Genetics | GPA 3.86   May. 2026

UNIVERSITY OF TORONTO   Toronto, ON (Canada)
- Bachelor of Science (Honors), Statistics, Biochemistry & Immunology | GPA 3.91   May. 2024

PUBLICATIONS

[1] **Cao W**, Kalish B. Pregnancy-Driven Neural Stem-Cell Niche Remodeling. In preparation.   2025
[2] Kukreja B.*, *Jeon S.*, **Cao W**, et al., Kalish B.T.# Maternal Gut–Immune Axis Programs the Neuro-Immune Landscape of the Developing Brain. *Nature Neuroscience* (Accepted)   2025
[3] Feng M.Y., **Cao W**, Li G., et al. Molecular Cartography of the Human and Mouse Down-Syndrome Brain. *Nature Communications*   2025
[4] Yang Y., Pang K., Li G., **Cao W**, Li X., Zhang Z. Deciphering 3′UTR-Mediated Gene Regulation Using Interpretable Deep Representation Learning. *Advanced Science*   2024
[5] Tahmasian N., Kukreja B., **Cao W**, et al. Perinatal Brain Injury Triggers Niche-Specific Changes to Cellular Biogeography. *eNeuro*   2024

AWARDS AND HONORS

[1] 1st Place – UTSW Healthcare Case Competition, $1500, UT Southwestern   2025
[2] Medical Biophysics Program Entrance Award, $5000, University of Toronto   2024
[3] SickKids Summer Research Award, $7500, The Hospital for Sick Children   2024
[4] Dorothy Helen McRobb Scholarship, $2000~$3000, University of Toronto   2022-2024
[5] University of Toronto Excellence Award (UTEA), $7500, University of Toronto   2023
[6] Undergraduate Research Opportunity Program, $7500, University of Toronto   2022

RESEARCH EXPERIENCE

BETH ISRAELDEACONESS MEDICAL CENTER   Boston, MA
Graduate Researcher, Vaccine Research   Nov.2025 – Present
Supervisor: Sizun Jiang, PhD
- Analyzing multiplexed spatial proteomics (Akoya CODEX) and single-cell spatial transcriptomics (CosMx/Xenium) from virus-associated lymphomas and primary CNS lymphoma (PCNSL) to map tumor–immune microenvironment architecture.
- Building pipelines that couple cell typing, spatial neighborhood features, and clinical variables (e.g., EBV status, treatment response) to identify immune-escape signatures and niches enriched for exhausted or dysfunctional T cells.
- Contributing computational methods for cross-domain fusion of imaging and gene-expression features in spatial multi-omics datasets, aiming to improve cell population delineation and discover targetable microenvironmental states.

BOSTON CHILDREN'S HOSPITAL   Boston, MA
Graduate Researcher, Newborn Medicine   Oct.2024 – Present
Supervisor: Brian Kalish, MD
- Building reproducible pipelines for Xenium datasets (virgin, pregnancy, ±CORT models) to study how microglial signals (IGF2, CXCL12) regulate neural stem cell activity.

- Developed novel computational tools include *SpatialAssignR*, an autonomous anatomical segmentation and *ADAT* adaptive down-sampled cell annotation (see Projects & Tools section).
- Drafting first-author manuscript (Pregnancy-Driven Neural Stem-Cell Niche Remodeling) and co-authored publications in *Nature Neuroscience* and *Nature Communications*.

THE HOSPITAL FOR SICK CHILDREN(SICKKIDS)   Toronto, ON (Canada)
Undergraduate Researcher, Mental Health & Neuroscience   May.2022 – Aug. 2024
Supervisor: Brian Kalish, MD

- Investigated spatial omics of prenatal and perinatal brain development, focusing on maternal immune activation and injury models.
- Designed and applied computational frameworks for analyzing cell neighborhood composition, cross-modal gene imputation, and cell–cell distance shifts (see Projects & Tools section).
- Co-authored publication in *eNeuro* on perinatal brain injury, providing computational analysis of microglia–neuron spatial relationships.

DONNELLY CENTRE FOR CELLULAR & BIOMOLECULAR RESEARCH(CCBR) Toronto, ON (Canada)
Undergraduate Researcher, Molecular Genetics   May.2022 – July. 2024
Supervisor: Zhaolei Zhang, PhD

- Co-developed 3UTRBERT, a Transformer-based model for RNA regulation, improving AUROC (0.788 → 0.836) and F1 (0.56 → 0.75) across 22 CLIP-seq datasets.
- Designed preprocessing workflows for mRNA 3′UTR sequences, integrated prediction of RNA–protein interactions, m6A modification sites, and subcellular localization.
- Co-authored peer-reviewed publication in Advanced Science (2024) presenting 3UTRBERT as a state-of-the-art model for 3′UTR-mediated gene regulation.

THE HOSPITAL FOR SICK CHILDREN(SICKKIDS)   Toronto, ON (Canada)
Undergraduate Researcher, Molecular Medicine   Sep.2022 – Apr. 2023
Supervisor: Jason T. Maynes, MD/PhD

- Investigated theNaV1.5 R222Q cardiac mutation and its role in arrhythmia and inflammation.
- Performed cryo-sectioning, protein isolation, Western blotting, and quantitative imaging (FIJI) to assess inflammasome activation and immune cell infiltration.
- Discovered gender-specific effects of the mutation, including increased M2 macrophage activity in male mutant mice. Highlighted connections between inflammasome signaling (NINJ1, NLRP3, GasderminD) and cytokine-mediated cardiac inflammation.

PRESENTATIONS & POSTERS

[1] SickKids Summer Research Conference   Aug. 2024
- Topic : Spatial Transcriptomics in Down Syndrome: MERFISH-Based Analysis of Cellular Heterogeneity

[2] Canadian Association for Neuroscience (CAN) Annual Meeting, Vancouver, BC   May. 2024
- Topic: Microglia–NSC Interactions in the Maternal SVZ: Spatial Transcriptomics and Computational Modeling of Pregnancy and Postpartum Depression. [Link]

[3] SickKids Research Institute Scientific Retreat   Nov. 2023
- Topic: Spatiotemporal Characterization of Brain Disorganization in Down Syndrome: Insights from the Ts65Dn Mouse Model and Human Fetal Brain Tissue.

[4] University of Toronto Summer Research Conference   Jul. 2023
- Topic: Spatial Transcriptomics in Down Syndrome: MERFISH-Based Analysis of Cellular Heterogeneity.

[5] University of Toronto Summer Research Conference   Jul. 2022
- Topic: Predicting mRNA Subcellular Localization and Functional Regions by Natural Language Models.

PROJECTS & TOOLS

[1] SpatialAssignR | Independent Developer | Boston Children's Hospital   2025
-    R package automating anatomical region annotation in MERFISH datasets using clustering and KNN voting. Reduced boundary errors by ~90% and annotation time to under 1 minute; adopted as a lab standard pipeline and pending publication.

[2] Adaptive Downsampled Annotation Transfer | Independent Developer | Boston Children's Hospital   2025
-    Low-memory annotation workflow enabling 10M+ cell 5K Xenium datasets to be processed on 32GB laptops. Achieved >95% concordance with full-scale labels and detected rare T-cell populations.

[3] Cross-Modal Imputation Pipeline | Independent Developer | SickKids   2024
-    Workflow integrating scRNA-seq with low-plex spatial data to expand feature space from ~500 to >20,000 genes. Recovered ~70% of masked genes; enabled spatial mapping of autism-risk loci.

[4] Cell Micro-Environment Analysis Pipeline | Independent Developer | SickKids   2024
-    Statistical framework quantifying per-cell neighborhood composition across developmental stages. Identified niche rewiring in maternal brain models; cited in manuscript submitted to Nature Neuroscience.

[5] Cell Distance Analysis Framework | Independent Developer | SickKids   2024
-    KDTree-based approach to measure true nearest-neighbor shifts between cell types. Revealed microglial displacement in maternal immune activation models; results published in eNeuro.

[6] 3UTRBERT | Co-Developer | CCBR   2023
-    Transformer-based deep learning model for 3′UTR-mediated gene regulation. Improved AUROC to 0.836 and F1 to 0.75; interpretability confirmed through motif discovery. Published in Advanced Science (2024).

[7] CNN - Chest X-ray Classifier | Co-Developer | Harvard University   2025
-    EfficientNetB3 and DenseNet121 ensemble for multi-label thoracic disease detection on ChestX-ray14 dataset (20K images). Achieved AUROC 0.88 and macro-F1 0.63.

[8] BERT - Toxic Comments Classifier | Co-Developer | Harvard University   2025
-    DistilBERT fine-tuned on 1.8M Wiki comments with focal loss for imbalance. Achieved F1 score of 0.91.

PROFESSIONAL/NON-ACADEMIC EXPERIENCE

QUANTARIUM AI   Remote
Data Science Intern   May. 2025 – Aug. 2025
-    Worked with large-scale real estate data (TerraLake™ / TerraIndex™-style datasets) to explore features for residential property valuation and market-velocity modeling.
-    Prototyped machine learning models (gradient boosting, ensemble methods) to predict property-level price indices and risk scores, benchmarking against existing internal baselines.
-    Helped build Python pipelines for data cleaning, feature engineering (geospatial + image-derived features), and model evaluation, improving reproducibility and experiment tracking.

CHEERYOU INTERNATIONAL CONSULTING   New York City, NY
Application Counselor   May. 2024 – Present
-    Advised students on graduate applications (CVs, PS, RS), improving clarity and competitiveness.

TORONTO GENERAL HOSPITAL   Toronto, ON (Canada)
Patient Care Volunteer, Department of Surgery   May. 2023 – May. 2024
-    Supported surgical inpatients through communication, enhancing recovery and patient–staff interactions.

THE HOSPITAL FOR SICK CHILDREN   Toronto, ON (Canada)
Administrative Volunteer, Department of Suspected Child Abuse and Neglection   May. 2023 – May. 2024
-    Assisted with administrative workflows, increasing departmental efficiency in sensitive cases.

Curriculum Vitae – Wuxinhao (Tim) Cao

SKILLS

Programming & Software
- Python (PyTorch, TensorFlow, JAX, Hugging Face, LangChain, LlamaIndex, FastAPI, Scanpy, AnnData, Matplotlib, Seaborn)
- R (Seurat, ggplot2, tidyverse, lme4, MAST, Shiny, Bioconductor)
- SQL, Bash, Git, Docker, Conda, renv, SLURM, Snakemake, Nextflow, REST APIs
- Cloud & infrastructure: AWS, GCP, HPC environments
- Data formats & tools: Parquet, Arrow, HDF5, DVC, GIN

Machine Learning & Deep Learning
- Classical ML: logistic/linear regression, ridge/lasso, elastic net, SVM, decision trees, random forest, XGBoost, LightGBM, CatBoost, k-means, hierarchical clustering, Louvain/Leiden, spectral clustering
- Deep learning: CNNs, RNN/LSTM/GRU, autoencoders, Transformers (BERT, ViT, GPT-style), multimodal fusion models, GNNs
- Model evaluation & optimization: cross-validation, focal loss, class reweighting, SMOTE/oversampling, ensemble learning, calibration, AUROC/AUPRC/F1/MCC, Bayesian optimization, grid/random tuning
- Interpretability: attention visualization, SHAP, LIME, feature importance, embedding/latent space analysis (UMAP, PCA, t-SNE)

Agentic AI & AI Systems Engineering
- Building LLM-based agents with LangChain, LangGraph, CrewAI, OpenAI Assistant tools
- Designing tool-calling, memory graphs, retrieval, and multi-agent workflows
- Implementing RAG pipelines with vector search (FAISS/Qdrant) and structured knowledge stores
- Multi-agent backend design: task routing, reflection/critic loops, error recovery, action planning
- Agentic web app deployment using FastAPI, Next.js, serverless functions, and event-driven architecture

MLOps & Scalable AI Deployment
- End-to-end ML pipelines: data engineering → model training → CI/CD → deployment
- Containerization and orchestration: Docker, Kubernetes, serverless inference
- Model serving & monitoring: API endpoints, batch/online inference, drift monitoring
- LLMOps: prompt orchestration, scalable inference, model versioning, vector DB management
- Logging, experiment tracking, and reproducibility using MLflow, Weights & Biases, DVC

Statistics & Computational Biology
- GLMs (logistic, Poisson, NB), mixed-effects models (LMM/GLMM), longitudinal/time-series models, survival analysis
- Bayesian inference: Stan, PyMC; resampling methods: bootstrap, jackknife, permutation tests
- Experimental design, multiple testing correction (FDR, Bonferroni)

Single-Cell & Spatial Omics
- scRNA-seq, scATAC-seq, CITE-seq, MERFISH, Xenium, Visium HD, Slide-seq, STARmap
- Annotation & cell-state modeling: CellTypist, pseudotime, RNA velocity, SCENIC
- Differential expression: DESeq2, edgeR, MAST; functional enrichment: GSEA, GO/KEGG
- Spatial analysis: neighborhood statistics, Moran's I, spatial autocorrelation, ligand–receptor inference (CellChat, NicheNet), cell–cell distance modeling, microenvironment analysis
- Multi-omics integration: cross-modal imputation, scRNA→spatial projection, co-embedding, Harmony, canonical correlation, schema alignment