Quang-Huy Nguyen

hqn0001@auburn.edu Google Scholar huynguyen250896.github.io

RESEARCH INTERESTS

Computational Biology: My work spans multi-omics studies integrating genomics, transcriptomics, metabolomics, and proteomics. This includes bulk, single-cell, and spatially resolved analyses, pathway analysis, variant calling (both germline and somatic), genome-wide association studies, and polygenic risk scores.
Computational Healthcare Informatics: I develop and apply data-driven methods for learning from heterogeneous clinical and biological data to support decision-making in medicine and public health.
I also address core AI challenges raised by these problems, including multi-modal modeling, agentic reasoning, uncertainty quantification, federated and collaborative learning, etc.

EDUCATION

Auburn University **Auburn, AL**
Ph.D. in Computer Science & Software Engr - ***cGPA: 3.73/4.0*** *08/2023 – Present*
Advisor: Professor Wei-Shinn Ku.

- Developed *FedeKD*, a federated knowledge distillation framework for robust knowledge transfer under heterogeneous settings. Achieved a significant reduction in negative transfer while improving predictive performance across 6 vision and medical imaging datasets.
- Developed *scBatchProx*, a federated learning-inspired framework for post-hoc single-cell batch integration via latent refinement, enabling continual and cumulative updates under dataset growth. Achieved Macro-F1 up to 0.76 with robustness to extreme batch imbalance.
- Developed *FedWQ-CP*, a federated conformal prediction framework for uncertainty quantification under joint data and model heterogeneity. Achieved robust client-level and global coverage across 7 vision and medical imaging datasets.

Dai Nam University **Hanoi, Vietnam**
Bachelor of Pharmacy - ***cGPA: 8.09/10.00*** *10/2015 – 08/2020*

RESEARCH EXPERIENCE

Department of Computer Science, HUST **Hanoi, Vietnam**
Visiting Graduate Student *12/2024 – 01/2025*
Advisor: Associate Professor Duc-Hau Le.

- Developed *DeepSSC*, a semi-supervised multi-omics classifier for cancer subtype prediction. Achieved 96.0% accuracy (99.2% AUC) across 3 TCGA cohorts, with up to 19% improvement over graph-based baselines; mentored undergraduate students on model development.

Department of Computer Science and Software Engineering, Auburn University **Auburn, AL**
Graduate Research Assistant *08/2023 – 08/2024*
Advisor: Research Assistant Professor Phi Hung Bya and Associate Professor Tin Nguyen.

- Conducted a literature review on functional analysis methodologies in LC-MS–based targeted and untargeted metabolomics. Dockerized 13 surveyed software tools as services, simplifying installation and execution to enhance accessibility for readers with limited IT expertise.
- Conducted a literature review on large language models and their applications in biomedical domain.
- Led the *Candidozyma auris* project; performed BLAST-based homology mapping and KEGG/GO enrichment analyses to characterize pathways involved in stress adaptation and antifungal resistance; contributed to manuscript editing and proofreading.
- Led the type 2 diabetes (T2D) polygenic score project; developed a genotype-to-pathway analysis pipeline that derives gene-level reference-deviation signals directly from VCF genotypes without matched transcriptomic data, identified T2D-associated pathways, and demonstrated strong predictive performance of pathway-specific polygenic risk scores; co-drafted the manuscript.

School of Computer Science and Engineering, Thuyloi University **Hanoi, Vietnam**
Visiting Scientist *10/2021 – 07/2023*
Advisor: Associate Professor Duc-Hau Le.

- Developed *DrGA*, an R toolkit for automated analysis of cancer driver genes, unifying gene set enrichment, gene–clinical association, module-level analysis, and patient stratification workflows.

School of Computer Science and Engineering, Thuyloi University **Hanoi, Vietnam**
Visiting Student *12/2017 – 07/2019*
Advisor: Associate Professor Duc-Hau Le.

- Read and replicated research papers, designed experiments and conducted algorithm on high-dimensional omics data in R.

INDUSTRY EXPERIENCE

Center for Biomedical Informatics, Vingroup Big Data Institute **Hanoi, Vietnam**
Research Scientist *10/2020 – 09/2021*
Managers: Associate Professor Duc-Hau Le and Dr. Nam Sy Vo.

- Developed *oCEM*, an R package for detecting overlapping co-expressed gene modules using decomposition methods (ICA/IPCA), with a permutation-based approach for selecting the optimal number of components.
- Designed and implemented a multi-omics analysis pipeline in R that detected novel prognostic subgroups of breast cancer, achieving a Cox log-rank p-value of 0.002 compared to the PAM50 classification scheme's p-value of 0.598. Led manuscript preparation and writing.
- Co-mentored an undergraduate student in clustering uveal melanoma patients in R.
- Contributed to the design and planning of the VN1K variant detection pipeline for large-scale genomic analysis of the Vietnamese population. Contributed to manuscript proofreading.

Research Intern *03/2020 – 10/2020*
Internship mentor: Associate Professor Duc-Hau Le.

- Developed and implemented a pipeline in R to systematically identify and analyze cancer driver genes using multi-omics data, ensuring a consistent approach to uncover key genetic factors contributing to cancer progression.

RESEARCH PROJECTS AND COMPETITIONS

Cell Maps for AI (CM4AI) CodeFest, UAB **Birmingham, AL**
NIH Bridge2AI Consortium *08/2025*

- Collaborated in a team of 5 to create separate embeddings of yeast fitness data, proteomic mass spectrometry, and cellular imaging, then fuse them into a co-embedding to construct a with-yeast multiscale subcellular hierarchy that uncovered hidden biological processes in triple-negative breast cancer.
- Implemented the analysis pipeline based on the agreed workflow in Python.

The 1000 Vietnamese Genomes Project (VN1K) **Hanoi, Vietnam**
Vingroup Joint Stock Company *2019 – 2024*
PI: Associate Professor Duc-Hau Le and Dr. Nam Sy Vo.

GenomeAsia100K **Hanoi, Vietnam**
NTU and others *2017 – 2022*
PI: Dr. Stephan Schuster (Singapore side) and Associate Professor Duc-Hau Le (Vietnam side).

TEACHING EXPERIENCE

Department of Computer Science and Software Engineering, Auburn University **Auburn, AL**
Graduate Teaching Assistant for COMP-1220 Python & COMP-1230 Matlab *08/2024 – 05/2026*

- Teaching Python and Matlab fundamentals to ≈ 60 undergraduates and guiding them through problem-solving exercises in lab sections.

PUBLICATIONS

†co-first author, *co-corresponding author

Preprints & Preparations

4. **Quang-Huy Nguyen**, Jiaqi Wang*, Wei-Shinn Ku*. *FedeKD: Energy-Based Gating for Robust Federated Knowledge Distillation under Heterogeneous Settings.* **arXiv**, 2605.05553. (Under review). 2026

3. **Quang-Huy Nguyen**, Zongliang Yue, Hao Chen, Wei-Shinn Ku*, Jiaqi Wang*. *Federated-inspired single-cell batch integration in latent space.* **arXiv**, 2602.00423. (Under review). 2026

2. **Quang-Huy Nguyen**, Jiaqi Wang*, Wei-Shinn Ku*. *Conformalized Neural Networks for Federated Uncertainty Quantification under Dual Heterogeneity.* **arXiv**, 2602.23296. (Under review). 2026

1. Hoang Le†, Van-Minh Nguyen†, Duc-Hai Tran, Van-Anh Tran, Van-Nui Nguyen, **Quang-Huy Nguyen**†*, Duc-Hau Le*. *A Semi-Supervised Deep Learning Framework for Multi-Omics Cancer Subtype Classification and Biomarker Discovery.* **bioRxiv**, 2022.01.13.476268. (Under review). 2026

Peer-reviewed publications

10. Trang T. H. Tran, ..., **Quang-Huy Nguyen**, ..., Quan Nguyen*, Van H. Vu*, Nam S. Vo*. *VN1K: a genome graph-based and function-driven multi-omics and phenomics resource for the Vietnamese population.* **bioRxiv**, 2025.04.15.648991. **Nature Communications** (Acceptance in Principle). 2026

9. Ching-Lan Chang†, Michael A. Moshi†, **Quang-Huy Nguyen†**, ..., David Hess*, Edwin C. Oh*. *Wastewater intelligence predicts the emergence of clinically-relevant and drug-resistant Candidozyma auris at healthcare facilities.* **Nature Communications**. 2026

8. **Quang-Huy Nguyen**, Ha Nguyen, Edwin C. Oh, Tin Nguyen. *Current approaches and outstanding challenges of functional annotation of metabolites: a comprehensive review.* **Briefings in Bioinformatics**, 25(6):bbae498. 2024

7. **Quang-Huy Nguyen**, Tin Nguyen, Duc-Hau Le. *DrGA: cancer driver gene analysis in a simpler manner.* **BMC Bioinformatics**, 23:86. 2022

6. **Quang-Huy Nguyen**, Tin Nguyen, Duc-Hau Le. *Identification and validation of a novel three hub long noncoding RNAs with m6A modification signature in low-grade gliomas.* **Frontiers in Molecular Biosciences**, 9:801931. 2022

5. **Quang-Huy Nguyen**, Duc-Hau Le. *oCEM: Automatic detection and analysis of overlapping co-expressed gene modules.* **BMC Genomics**, 23:39. 2022

4. Thi Hai Yen Nguyen†, Tin Nguyen, **Quang-Huy Nguyen†***, Duc-Hau Le*. *Re-identification of patient subgroups in Uveal Melanoma.* **Frontiers in Oncology**, 11:731548. 2021

3. **Quang-Huy Nguyen**, Duc-Hau Le. *Similarity calculation, enrichment analysis, and ontology visualization of biomedical ontologies using UFO.* **Current Protocols in Bioinformatics**, 1:e115. 2021

2. **Quang-Huy Nguyen**, Duc-Hau Le. *Improving existing analysis pipeline to identify and analyze cancer driver genes using multi-omics data.* **Scientific Reports**, 10:20521. 2020

1. **Quang-Huy Nguyen**, Hung Nguyen, Tin Nguyen, Duc-Hau Le. *Multi-omics analysis detects novel prognostic subgroups of breast cancer.* **Frontiers in Genetics**, 11:574661. 2020

PEER-REVIEWED CONFERENCES

1. Hung Nguyen, Bang Tran, Duc Tran, **Quang-Huy Nguyen**, Duc-Hau Le, Tin Nguyen. *Disease subtyping using community detection from consensus networks.* **2020 12th International Conference on Knowledge and Systems Engineering (KSE)**, Can Tho, Vietnam, IEEE, 2020; 318–323.
Oral Presentation.

PRESENTATIONS AND TALKS

1. 2024 MW CTR-IN Annual Meeting at The Westin Las Vegas Hotel & Spa, Las Vegas, NV. *Topic: A Systems-level Approach for Predicting Patient Survival using Multi-omics Data.* 2024

HONORS AND AWARDS

Gold Reviewer Award — Auburn, AL
Issued by ICML 2026 — *05/2026*

- Gold Reviewer Award (top 25% reviewers) at ICML 2026.

Google PhD Fellowship Nominee — Auburn, AL
Associated with Auburn University — *04/2026*

- One of three graduate students nominated university-wide by AU Graduate School for 2026 Google PhD Fellowship.

Certificate of merit for learning and class management - Class of 2015 — Hanoi, Vietnam
Associated with Dai Nam University — *08/2020*

- Top 3% of the Department's 2015 cohort (nearly 110 students).
- Class president in 5 consecutive years (2015-2020).

27th Mathematics Olympic for Students — Nha Trang, Vietnam
Issued by Vietnam Mathematical Association — *04/2019*

- 🏆 Received an honorable mention in Linear Algebra among 640 undergraduates representing 80 universities, academies, colleges across Vietnam.
- Taught myself linear algebra from scratch in preparation for the competition.
- Proudly attended the competition as the first and only representative of my university and became the first Vietnamese non-engineering student in the history of the competition.

TECHNICAL SKILLS

- **Programming:** Python, R, Matlab, JavaScript
- **Single-Cell & Omics:** Scanpy, AnnData, MuData, Seurat
- **Web Development:** HTML, CSS, RESTful APIs
- **Deep Learning & ML:** PyTorch, TensorFlow 2, scikit-learn
- **Database:** MySQL, SQLite, Oracle, MongoDB
- **Clouds & Tools:** BigQuery, AWS, Git, GitHub, Docker