

TRANG TU

Boston, Massachusetts | (781)-697-5909 | tu.tra@northeastern.edu | [LinkedIn](#) | [GitHub](#)

EDUCATION

Northeastern University, Boston, MA

Master of Science in Bioinformatics

Expected December 2025

GPA: 3.87

Relevant Coursework: Bioinformatics Programming (Python), Introduction to Computational Methods in Bioinformatics (Bash), Statistics for Bioinformatics (R), Database Design and Security (SQL), Transcriptomics in Bioinformatics (Python, R), Genomics in Bioinformatics (Python), Collect/Store/Retrieve Data (R)

Underwood International College, Yonsei University, Seoul, South Korea

August 2019

Bachelor of Science in Life Science and Biotechnology

TECHNICAL SKILLS

Programming Languages: Python, R, SQL, Bash

Tools: R Studio, Pycharm, VSCode, Microsoft SQL Server, MS PowerPoint, MS Excel, MS Word, BLAST, PyMOL, YASARA, POOL, Nextflow, Snakemake, Git, Bowtie2, samtools, Singularity

Operating Systems: Mac, Windows, Linux

Languages: Vietnamese (native), English (proficient), Korean (proficient)

WORK EXPERIENCE

The Ondrechen Research Group, Northeastern University

Boston, MA

Research Assistant

Jul 2024- Present

- Joining an NFS-supported project applying machine learning (Partial Order Optimum Likelihood, POOL) on 3D structures of 297 proteins (37 noncanonical and 260 canonical program to predict RNA binding sites using Bash and Linux-based tools. Visualizing protein-RNA interactions using YASARA and PyMOL. Identifying top 10% residues that interact with RNA within 3.2Å and 5Å to support structural binding analysis.
- Developing and optimizing Python and R scripts to process large-scale datasets of POOL scores, Solvent Accessible Surface Area (SASA), Consurf, Concavity score on HPC clusters to investigate binding mechanisms, protein conformations and RNA structure.
- Serving as lead analyst of the team. Automated data pipelines and web scraping workflows in Python, improving data processing efficiency by 60% and accelerating downstream analysis. Creating data visualizations (e.g., Python Matplotlib, Seaborn) to explore features predictive of RNA-binding potential and inform ML model training.

Northeastern University

Boston, MA

Teaching Assistant for MATH7340 Statistics for Bioinformatics

Sept 2024- Present

- Supporting 100+ Master's students in Bioinformatics through weekly office hours on concepts such as Central Limit Theorem, MLE, Confidence Intervals, Shapiro-Wilk Test, Fisher's Exact Test, ANOVA, Monte Carlo Simulation.
- Providing two 2-hour office hour sessions each week, offering additional support for code debugging and guidance on complex modules.

RESEARCH EXPERIENCE

Yonsei University

Seoul, Korea

Nucleus isolation and Hydrogels synthesis

Jul 2018 – Aug 2019

- Prepared cell cultures and conducted nucleus isolation and examined nucleus deformation using SAW device.
- Self-designed microscale collagen-based hydrogels and PEGDA hydrogels externally coated with actin for potential use in drug delivery.

Yonsei University

Seoul, Korea

Microneedles contamination and effectiveness record

Jun 2016 – Dec 2017

- Quantitatively tested out biomarker LPS and microneedles contamination before and after purification processes by electron beams and gamma rays using LPS ELISA.
- Recorded and analyzed the effectiveness of purified adenosine-loaded microneedle patches incorporated with skincare products through clinical efficacy tests on 22 women in terms of wrinkle improvement, dermal density, skin elasticity, and hydration.
- Second-authored "Adenosine-loaded dissolving microneedle patches to improve skin wrinkles, dermal density, elasticity, and hydration" paper in the International Journal of Cosmetic Science.

PROJECTS

Database Design and Data Management (SQL, MySQL Workbench, R)

- Designed and forward-engineered a relational database using MySQL Workbench, optimizing data storage and retrieval.
- Analyzed large datasets using SQL queries and statistical modeling, understanding data-driven decision-making. Visualized complex data relationships using R (ggplot2) to identify trends and optimize workflow efficiency.

Reproducibility and Validation Study (R, Bash)

- Reproduced the single-cell RNA-seq analysis pipeline from the published paper "*Integrative Analysis of Clinical and Bioinformatics Databases to Reveal the Role of Peripheral Innate Immunity in Kawasaki Disease*" using R on HPC clusters; successfully validated findings from Figures 1 and 2 with 98% concordance.
- Analyzed ~49,000 cells from control and patient groups (GEO dataset), reformatted data to match CellRanger output, performed PCA and UMAP using Seurat package, visualized biomarker clusters, annotated cells using SingleR, and conducted differential gene expression analysis using FindAllMarkers identifying 300+ significant genes, along with GO and KEGG pathway enrichment.

PUBLICATIONS

Kang, G., Tu, T.N.T., Kim, S., Yang, H., Jang, M., Jo, D., Ryu, J., Baek, J. and Jung, H. (2018), Adenosine loaded dissolving microneedle patches to improve skin wrinkles, dermal density, elasticity and hydration. *Int J Cosmetic Sci*, 40: 199-206. doi:10.1111/ics.1245