

TAEHOON HA

📍 New York, NY · ✉ taehoonh91@gmail.com · 📞 (919) 308-3505 · 🏠 taehoonh.me · 📧 taehoonha

WORK EXPERIENCE

STONY BROOK MEDICINE (RF SUNY) | Biostatistician (Data Manager, Statistician) Stony Brook, NY
Nov 2024 – Present

- Provide statistical guidance on study design, data analysis, and result interpretation for clinical research.
- Develop and implement advanced analytics models, manage large datasets, and perform sophisticated statistical analyses using R and SAS.
- Support faculty, students, and researchers by conducting statistical workshops, teaching fundamental and advanced biostatistical concepts, and mentoring trainees.

COLD SPRING HARBOR LABORATORY | Biostatistician (Core Director, Biostatistics Core) Cold Spring Harbor, NY
Sep 2020 – Nov 2024

- Conducted weekly office hours, providing over 100 researchers with tailored statistical consultations.
- Collaborated with Cancer Center and Northwell Health investigators on pre-clinical and clinical studies, leveraging advanced statistical methods—including multivariate regression, random forest classifiers, and machine learning algorithms—to enhance data-driven decision-making, improve study design robustness, and increase analytical accuracy.
- Authored and reviewed more than 50 statistical analysis plans (SAPs) and methodological sections for research proposals, grant applications (including NIH and NCI), and manuscripts, contributing to a higher success rate in securing competitive research funding.
- Developed and optimized data pipelines in R and Python, reducing data preparation time by 30% and enabling advanced and predictive analytics in over 100 data analysis projects, including the design and validation of experiments.
- Led annual biostatistics workshops, training 100+ researchers in biostatistics, improving analytical capabilities and research output.

WEILL CORNELL MEDICINE | Research Assistant – Biostatistics & Data Science (Advisor: Xi Kathy Zhou, PhD) New York, NY
Aug 2019 – Sep 2020

Application: Collaboration with Andrew J. Dannenberg Group

- Collaborated extensively with investigators researching cancer, obesity, and metabolic diseases.
- Provided statistical consulting support to clinical (lab) and genomic data using R.
- Performed sample size and power calculations, designed and implemented database for clinical data collection.
- Interpreted statistical analysis reports for investigators and wrote statistical method sections for scientific publication.

Methodology: Application of Bayesian model averaging to better identify differentially expressed genes in high-dimensional setting

- Developed a new statistical method using Bayesian model averaging to identify DE genes associated with one or more patient characteristics (or phenotypes), as well as their interactions.
- Built and improved R package ‘BMAseq’ using Bayesian model averaging to analyze observational gene-expression data.
- Applied the Bayesian model averaging method to multiple types of datasets, from metabolomics data to NGS data, to check its performance.

JOHNS HOPKINS UNIVERSITY | Voluntary Researcher – Bioinformatics Analyst (Advisor: Bongsoo Park, PhD) Remote
Apr 2019 – Apr 2021

Transcriptome and Epigenome Atlas for Air Pollution PM2.5

- Generated a pipeline code to analyze liver single-cell RNA-seq data using R package Monocle to cluster the cells and predict cell types.
- Constructed analytical pipelines using R and Python to identify DE genes associated with the exposure to particulate matter.
- Checked sample quality by Spearman’s Rho correlation, hierarchical clustering, and perform Principal Component Analysis (PCA).
- Performed Differential Expression analysis and Gene Ontology (GO) Term analysis on six different brain sub-areas and liver cells.
- Conducted pathway analysis using Ingenuity Pathway Analysis (IPA) software.

TECHNICAL SKILLS

- **Programming:** R, Python, SAS
- **Database:** MySQL, PostgreSQL, SQL Server
- **Cloud/Distributed computing:** Amazon Web Services
- **Visualization:** Tableau, Power BI, Looker, Prism Graphpad
- **Deployment:** Shiny App
- **Version control:** Github

EDUCATION

CORNELL UNIVERSITY, WEILL CORNELL MEDICINE | Master of Science, Biostatistics and Data Science New York, NY
Dec 2019

- Academic Excellence (Over 4.0 cumulative GPA) Award
- Thesis: Application of a Bayesian Model Averaging Method to Observational Metabolomics Data Analysis

DUKE UNIVERSITY | Master of Science, Business Analytics Durham, NC
May 2018

- Capstone project: Duke University Hospital (Duke Health)

SUNGKYUNKWAN UNIVERSITY | Bachelor of Business Administration with an emphasis on Quantitative Methods Seoul, Korea
Jul 2017

- Dean’s list with distinction
- Study-abroad: School of Arts and Sciences at the University of Pennsylvania (2014) – Travel funding, Mar 2015
- Military Service: Republic of Korea Marine Corps (Rank: Sergeant, 2011 – 2013)

ADDITIONAL INFORMATION

PUBLICATIONS

Published

- O Klingbeil, D Skopelitis, C Tonelli, A Alpsy, F Minicozzi, D Aggarwal, **T Ha**, OE Demerdash, DL Spector, DA Tuveson, P Cifani, and CR Vakoc (2024). *MARK2/MARK3 kinases are catalytic co-dependencies of YAP/TAZ in human cancer*. Cancer Discovery.
- E Zhou, JI Yang, A Habowski, A Deschênes, P Belleau, **T Ha**, C Tzanavaris, J Boyd, C Hollweg, X Zhu, DA Tuveson, and DA King (2024). *GATA6 amplification is associated with improved survival of TP53-mutated pancreatic cancer*. Pancreas.
- S Henry, SM Lewis, SL Cyrill, MK Callaway, D Chatterjee, AVH Somasundara, G Jones, XY He, G Caligiuri, MF Ciccone, IA Diaz, A Biswas, E Hernandez, **T Ha**, JE Wilkinson, ME Egeblad, DA Tuveson, CO dos Santos (2024). *Host response during unresolved urinary tract infection alters mammary tissue homeostasis through collagen deposition and TIMP1*. Nature Communications.

- Y Gao, XY He, XS Wu, YH Huang, S Toneyan, JJ Ipsaro, **T Ha**, PK Koo, M Egeblad, L Joshua-Tor, and CR Vakoc (2023). *ETV6 Dependency in Ewing Sarcoma through Antagonism of EWS-FLI1-Mediated Enhancer Activation*. Nature Cell.
- S Bhatia, M Kramer, S Russo, P Naik, G Arun, K Brophy, P Andrews, C Fan, C Perou, J Preall, **T Ha**, D Plenker, D Tuveson, A Rishi, J Wilkinson, WR McCombie, K Kostroff, and D Spector (2022). *Patient-derived Triple Negative Breast Cancer Organoids Provide Robust Model Systems that Recapitulate Tumor Intrinsic Characteristics*. Cancer Research.
- CM Brennan, S Nadella, X Zhao, RJ Dima, N Jordan-Martin, BM Demestichas, SO Kleeman, M Ferrer, E Gablenz, N Mourikis, M Rubin, H Adnani, **T Ha**, S Prum, CB Schleicher, SS Fox, M Ryan, C Pili, J Poulard, G Goldberg, JM Crawford, S Goodwin, X Zhang, J Preall, S Costa, J Conigliaro, JR Masci, J Yang, DA Tuveson, KJ Tracey, T Janowitz (2022). *Oral Famotidine vs Placebo in Diverse Non-Hospitalized Patients with COVID-19: A Randomized Double-Blind, Data-Intense, Phase 2 Clinical Trial*. Gut.
- S Basu, C Liu, XK Zhou, N Ryohei, **T Ha**, J Chen, M Johncilla, RK Yantiss, DC Montrose, and AJ Dannenberg (2021). *GLUT5 is a Determinant of Dietary Fructose-mediated Exacerbation of Experimental Colitis*. AJP Gastrointestinal and Liver Physiology.
- JI Yang, **T Ha**, E Zhou, C Tzanavaris, CE Devoe, X Zhu, and J Boyd (2021). *Association of TP53 Mutation Status and GATA6 Amplification with Clinical Outcome of Pancreatic Cancer*. Journal of Clinical Oncology.
- DC Montrose, M Foronda, S Saha, EM McNally, XK Zhou, **T Ha**, J Krumsiek, A Verma, O Elemento, RK Yantiss, Q Chen, SS Gross, L Galluzzi, LE Dow and AJ Dannenberg (2021). *Exogenous and Endogenous Sources of Serine Contribute to Colon Cancer Metabolism and Growth*, Cancer Research.
- NM Iyengar, XK Zhou, H Mendieta, O El-Hely, DD Giri, L Winston, DJ Falcone, H Wang, L Meng, **T Ha**, M Pollak, CA Hudis, M Morrow, and AJ Dannenberg (2021). *Effects of Obesity on Breast Aromatase Expression and Systemic Metabo-Inflammation in Women with BRCA1 or BRCA2 Mutations*. npj Breast Cancer.
- R Nishiguchi, S Basu, HA Staab, N Ito, XK Zhou, H Wang, **T Ha**, M Johncilla, RK Yantiss, DC Montrose, and AJ Dannenberg (2021). *Dietary Interventions to Prevent High Fructose Diet-associated Worsening of Colitis and Colitis-associated Tumorigenesis in Mice*. Carcinogenesis.
- EH Williams, TR Flint, CM Connell, D Giglio, H Lee, **T Ha**, E Gablenz, N Bird, JMJ Weaver, H Potts, CT Whitley, MA Bookman, AG Lynch, HV Meyer, S Tavaré, and T Janowitz (2020). *CamGFR v2: A New Model for Estimating the Glomerular Filtration Rate from Standardized or Non-Standardized Creatinine in Patients with Cancer*. Clinical Cancer Research.

Accepted

TEACHING EXPERIENCE

- HBH550: Statistics in Life Sciences** | Teaching Assistant for Prof. Jie Yang, Stony Brook Medicine Jan 2025
- Biostatistics Course 2024** | Instructor, Cold Spring Harbor Laboratory Aug 2024
- Biostatistics Course 2023** | Instructor, Cold Spring Harbor Laboratory Jul 2023
- Big Data in Medicine: Biomedical Imaging** | Teaching Associate for Prof. Elizabeth Sweeney, Weill Cornell Medicine Spring 2020
- Big Data in Medicine: Genetics & Genomics** | Teaching Associate for Prof. Davide Risso, Weill Cornell Medicine Spring 2020
- Categorical and Censored Data Analysis** | Teaching Associate for Prof. Oleksandr Savenkov, Weill Cornell Medicine Fall 2019
- Led lab sessions for 32 Master's candidate students to teach biostatistical methods with R
- Reviewed and graded weekly homework and provide guidance on lab assignments
- Held regular office hours regarding questions on course materials, assignments, and academic concerns

PRESENTATION & PRESS INTERVIEW

- Virtual Core Knowledge: Biostatistics Workshop, Cold Spring Harbor Laboratory Jan 2021
- Interview Article: Analysis of 3,600 COVID-19 sequences on Nextstrain, Donga Science Apr 28th, 2020
- The Single-cell Pathology Landscape of Breast Cancer, Weill Cornell Medicine Mar 2020
- Genomic Signatures Predict the Immunogenicity of BRCA-Deficient Breast Cancer, Weill Cornell Medicine Dec 2019
- Profound Perturbation of the Metabolome in Obesity Is Associated with Health Risk, Weill Cornell Medicine Aug 2019

SERVICES

- Korean Data Science / Statistics Community Leader** Nov 2021 – Present
- Lead a community of over 500+ Korean data science and statistics students and professionals in the US, organizing meetups, guest speaker events, knowledge-sharing sessions, and career opportunities.
- Nextstrain** | Voluntary Technical Translator May 2020
- Translated technical document and weekly genomic analysis of COVID-19 situation reports into Korean