TAEHOON HA

New York, NY

(267) 225-2588 | tah4002@med.cornell.edu | taehoonh.me | taehoonh | taehoonha

Education

Weill Cornell Medicine New York, NY
M.S., Biostatistics and Data Science 2018 – 2019
· Advisor: Xi Kathy Zhou, PhD
· Thesis: *Application of a Bayesian model averaging method to observational metabolomics data analysis*

Duke University Durham, NC
M.S., Business Analytics 2017 – 2018
· Capstone project: Duke University Hospital (Duke Health)

Sungkyunkwan University Seoul, Korea
B.B.A. with an emphasis on Quantitative Methods (Statistics) 2009 – 2017
· Dean's list with distinction
· Study-abroad: School of Arts and Sciences at the University of Pennsylvania (2014)

Research Experience

Weill Cornell Medicine New York, NY
Research Assistant - Biostatistics (Advisor: Xi Kathy Zhou, PhD) 04/2019 – Present

Application: Collaboration with Andrew J. Dannenberg, MD group
· Collaborate with Andrew J. Dannenberg, MD group to identify the link among obesity, inflammation, and breast cancer
· Provide statistical consulting support to clinical (lab) and genomic data using R
· Design and implement database for clinical trial data collection and write statistical methods sections for scientific publications.
· Update and customize R package software '*BTKR*' which includes multiple functions that implement some commonly used biostatistics analysis methods for the summary of data

Methodology: Bayesian model averaging
· Assist advisor with developing a new statistical method using Bayesian model averaging to identify differentially expressed genes associated with one or more patient characteristics (or phenotypes), as well as their interactions
· Develop R package '*BMAseq*' using Bayesian model averaging to analyze observational gene-expression data
· Apply the Bayesian model averaging method to large scale public RNA-seq gene expression (NGS) data
· Utilize the Bayesian model averaging method to observational metabolomics data to improve differentially expressed (DE) metabolites identification in high dimensional setting

Johns Hopkins Bloomberg School of Public Health (Remote) Baltimore, MD
Voluntary Researcher – Bioinformatics Analyst (Advisor: Bongsoo Park, PhD) 04/2019 – Present

Transcriptome and epigenome atlas for air pollution $PM_{2.5}$
· Generate a pipeline code to analyze liver single-cell RNA sequencing data using R package Monocle to cluster the cells and predict their cell types.
· Align RNA sequences using STAR or TopHat2
· Construct analytical pipelines using EdgeR, DESeq2, or Limma/Voom to identify differentially expressed genes associated with the exposure to ultra-fine dust, $PM_{2.5}$
· Verify sample quality using Spearman's Rho correlation and hierarchical clustering methods
· Check the distribution by Relative Log Expression (RLE), and Principal Component Analysis (PCA)
· Normalize each sample and removed unwanted variances and filtered top N genes by p-values

- Perform DE (Differential Expression) analysis and Gene Ontology (GO) Term analysis on six different brain sub-areas and liver cells
- Conduct pathway analysis using Ingenuity Pathway Analysis (IPA) software

JB Lab & Clinic (Remote) Seoul, Korea
Research Scientist 08/2018 – 12/2019

- Identified the association of sodium intake and hypertension, metabolic syndrome, and ARB treatment effect from a 10K+ hypertension patients data acquired from a K-MetS study

Weill Cornell Medicine New York, NY
Independent Project 08/2018 – 08/2019

The Effect of Right Heart Catheterization (RHC) During the First 24 Hours on 30-day Mortality of Critically Ill Patients in Intensive Care Units 05/2019 – 08/2019

- Applied data adaptation method to estimate the propensity score
- Generated causal model estimands using multiple methods (random forest, GLM, elastic net, LASSO, and gradient boosting with SuperLearner)
- Conducted 5-fold cross-validation to evaluate each estimand's performance and coefficient in the ensemble estimator

Rate of General Anesthesia Use for Cesarean Delivery Among Anesthesiologists with and without Fellowship Training in Obstetric Anesthesia 01/2019 – 05/2019

- Fitted the logistic regression model to identify whether the fellowship-trained anesthesiologists are more or less likely to provide general anesthesia for non-routine C-section as compared to non-fellowship trained anesthesiologists
- Identified the significant difference between general and neuraxial anesthesia and the odds of using general anesthesia was 0.471 times lower in attending fellowship training anesthesiologists compared the non-fellowship training anesthesiologists

Cost-Effective Optimization of Model-Based Prediction of Cardiovascular Disease 03/2019 – 05/2019

- Aimed to increase the accuracy of cardiovascular disease diagnosis using a model-based approach
- Determined the demographic factors and medical tests that help predict the likelihood of heart disease using proportional odds model and logistics regression
- Identified a combination of necessary medical tests that help predict the probability of heart disease in a cost-effective manner and reduced the exam cost by $110.17

Neuroendocrine Prostate Cancer (NEPC) 11/2018 – 12/2018

- Identified that there is no difference in the proportions of 3-month progression status between patients in the NEPC and non-NEPC group treated with alisertib
- Determined that there is no association between 3-month progression-free survival for all clinical characteristics and gene abnormalities
- Identified an association between PSA level and prior systemic therapies between NEPC and non-NEPC groups treated with alisertib

Major Risk Factors of Low Birth Weight Babies 10/2018 – 12/2018

- Determined that low birth weight is associated with the mother's premature labor history, race, smoking status, weight of the last menstrual period, and history of hypertension
- Fitted multiple generalized linear models with the model selection based on deviance and p-values

Study design of Tommy John Surgery for MLB Pitchers by *Statcast* Measurements 09/2018 – 10/2018

- Identified the primary injury factors of Ulnar Collateral Ligament Reconstruction for Major League Baseball pitchers by Statcast measures
- Designed a study to identify major factors of Ulnar Collateral Ligament Reconstruction (a.k.a. Tommy John Surgery) for Major League Baseball Pitchers by Statcast measurements

Teaching

Big Data in Medicine: Biomedical Imaging Spring 2020
Teaching Associate for Prof. Elizabeth Sweeney, Weill Cornell Medicine

Big Data in Medicine: Genetics & Genomics Spring 2020
Teaching Associate for Prof. Davide Risso, Weill Cornell Medicine

Categorical and Censored Data Analysis Fall 2019
Teaching Associate for Prof. Oleksandr Savenkov, Weill Cornell Medicine

- Lead lab sessions for 32 Master's candidate students to teach biostatistics methods with R
- Review and grade weekly homework and provide guidance on lab assignments
- Hold regular office hours regarding questions on course materials, assignments, and academic concerns

Publications

Submitted

- Iyengar NM, Zhou XK, Mendieta H, El-Hely O, Giri DD, Winston L, Falcone DJ, Wang H, Meng L, **Ha T**, Pollak M, Hudis CA, Morrow M, Dannenberg AJ (2020), *Effects of Obesity on Breast Aromatase Expression and Systemic Metabo-Inflammation in Women with BRCA1 or BRCA2 Mutations*, Submitted to Nature Cancer.
- Cho BA, Zhou XK, Morrow M, Giri DD, Sharaiha RZ, Kumar R, Yaghoobzadeh H, **Ha T**, Verma A, Elemento O, Pollak M, Laurence J, Iyengar NM, and Dannenberg AJ (2020), *Overexpression of Complement-related Genes in Adipose Tissues of Obese Individuals: Implications for the Pathogenesis of COVID-19*, Submitted to JCI Insight.
- David C. Montrose, Miguel Foronda, Suchandrima Saha, Erin M. McNally, Xi Kathy Zhou, **Taehoon Ha**, Jan Krumsiek, Akanksha Verma, Olivier Elemento, Rhonda K. Yantiss, Qiuying Chen, Steven S. Gross, Lorenzo Galluzzi, Lukas E. Dow and Andrew J. Dannenberg (2020), *Exogenous and Endogenous Sources of Serine Contribute to Colon Cancer Metabolism and Growth*, Submitted to Cancer Research.

In Preparation

- Bongsoo Park, Jeffrey A. Deiuliis, Rengasamy Palanivel, **Taehoon Ha**, Ji Eun Park, Sanjay Rajagopalan, and Shyam Biswal, *Sex-difference in the metabolic effects of air pollution exposure*.
- **Taehoon Ha**, Ji Eun Park, Rengasamy Palanivel, Sanjay Rajagopalan, Shyam Biswal, and Bongsoo Park, *Sex-difference in liver transcriptome with environmental exposure*.
- B Park, S Kim, **Taehoon Ha**, JE Park, V Vinayachandran, KD Hansen, B Paul, S Rajagopalan, and S Biswal, *Brain transcriptome map of air pollution $PM_{2.5}$.*
- Hanhan Wang, Lingsong Meng, **Taehoon Ha**, Xi Kathy Zhou, *A Bayesian model averaging approach for RNA-seq counts data (BMA-seq) and its application*.

Awards

Academic Excellence (Over 4.0 cumulative GPA), *Weill Cornell Medicine* 07/2019

Exchange Student Scholarship at the University of Pennsylvania (Travel Funding), *Sungkyunkwan University* 03/2015

Academic Excellence Scholarship, *Sungkyunkwan University* 01/2015; 03/2014

Pursuit of Excellence Scholarship, *Sungkyunkwan University* 01/2014; 01/2010; 07/2009

Guest Lectures & Presentations & Interviews

How to organize, manage, share, and process bio/healthcare data, *Konkuk University Hospital* 12/2018

RNA-seq data analysis 101, *JB Lab & Clinic* 08/2018

How to collaborate with a statistician and data analyst/scientist, *Sungkyunkwan University* 06/2017

Interview Article: Analysis of 3,600 COVID-19 sequences on Nextstrain, *Donga Science* 04/28/2020

Service

Nextstrain (Open-source platform to harness the scientific and public health potential of pathogen genome data) 03/2020 – 05/2020

Voluntary Technical Translator

- Translated weekly genomic analysis of COVID-19 spread situation reports into Korean
- Translated Nextstrain's technical document into Korean: How to read phylogenetic tree

Prudential Foundation 2015

Project CD34: Stem Cell Donation Trend Visualization

- Integrated and constructed the stem cell donation trend databases in Korea
- Generated real-time dashboard to help the board's decision-making process

S-ONE, Sungkyunkwan University 2013 – 2014

Data Analysis & Visualization Education Session Chief

- Led lab sessions for 40 students to teach statistical methods and data visualization tools when analyzing social science data

Samsung Dream Class Program 2009 – 2010

High School Mathematics Tutor

- Taught high school and pre-college math to 20 low-income students

Technical Skills

- **Programming**: R, SAS, Python, bash (shell)
- **Database**: MySQL, SQL Server, Microsoft Access
- **Document**: LaTeX, Markdown
- **Cloud/Distributed computing**: Amazon Web Services, Google Cloud Platform, Microsoft Azure, Oracle Grid Engine (i.e., Sun Grid Engine)
- **Platform:** MacOS, Windows, Ubuntu, Linux, Unix
- **Visualization:** Tableau Software

Membership

- American Statistical Association (ASA)
- Korean Statistical Society (KSS)
- American Heart Association (AHA)
- American Association for the Advancement of Science (AAAS)