

YUSI FANG

914-314-9339 | yuf31@pitt.edu | 230 N Craig St. APT707-Pittsburgh-PA-15213 | [Personal Website](#)

SUMMARY OF QUALIFICATION

- **Programming:** Proficient in R programming language and familiar with the Python and LaTeX programming language and statistical packages and software including Rcpp, RShiny, SAS, STATA, and UNIX system
- **Statistical Methodologies:** Excellent statistical background with more than 5 years of methodological research experience in meta-analysis, large-scale inference, and statistical learning, resulting in 4 published and 3 submitted methodological papers
- **Statistical Computing Skills:** Hands-on experience in analyzing genetic, genomic, and clinical data by various statistical methods and machine learning methods using R
- **Communication and Writing Skills:** Strong communication and writing skills from long-term collaborations with non-statistical researchers, resulting in 8 published collaborative papers

EDUCATION

School of Public Health, University of Pittsburgh

Pittsburgh, PA

Ph.D. in Biostatistics

Aug. 2017 – Apr. 2023

- Advisors: Prof. George C. Tseng, Sc.D. and Prof. Zhao Ren, Ph.D.
- Research Interest: Large-Scale Inference, Meta-Analysis of Heterogeneous Signals, Statistical Genomics and Genetics, Machine Learning Methods
- GPA: 4.00/4.00
- ASA Pittsburgh Chapter Student of the Year, 2023

School of Mathematical Sciences, Xiamen University

Xiamen, China

B.S. in Mathematics and Applied Mathematics

Sep. 2013 – Jul. 2017

- Relevant Coursework: Mathematical Statistics, Applied Linear Model, Statistical Calculation, Operation Research and Optimization, Probability Theory, Mathematical Analysis I, II, and III, Partial Differential Equations, Advanced Algebra.
- GPA: 3.86 / 4.00
- Outstanding Undergraduate Thesis

PROFESSIONAL EXPERIENCE

Graduate Student Researcher

Sep. 2017 – Apr. 2023

Department of Biostatistics, School of Public health, University of Pittsburgh

Pittsburgh, PA

Part Time: 20 Hours per Week

- Provided statistical consulting for multiple investigators from Magee-Womens Research Institute, University of Pittsburgh Medical Center
- Analyzed high-throughput multi-omics data, including ribonucleic acid sequencing (RNA-seq) and proteomics data, from raw data preprocessing using command-line tools to downstream analysis such as differentially expressed (DE) analysis, pathway enrichment analysis, WGCNA, and transcriptomic meta-analysis, and machine learning methods such as clustering and dimension reduction
- Analyzed cancer registry data using machine learning methods, survival analysis methods, and propensity score matching
- Published and co-authored eight collaboration papers in cancer and other diseases research

Intern: Biostatistics

Jun. 2021 – Aug. 2021

Biogen, Inc

Cambridge, MA

Mentors: George Kong, Ph.D. and Shelby Huang, Ph.D.

Full Time: 40 Hours per Week

- Designed and implemented simulations to justify the type I error control procedures of the 2-in-1 adaptive phase 2/3 clinical trial design and its variants

- Designed and implemented simulations to mimic drug development for lupus and evaluated the speed, efficiency, and risk of the above innovative clinical trial designs
- Attended and presented at the immune biostatistics group weekly meetings and presented the summer project at the biostatistics departmental meeting

SELECTED RESEARCH PROJECTS

P-Value Combination via Ensemble

Aug. 2021 – Mar. 2022

Department of Biostatistics, University of Pittsburgh

- Comprehensively evaluated multiple p-value combination methods in terms of asymptotic efficiencies (Bahadur slope) and finite-sample performance
- Developed an ensemble method to integrate the strengths of multiple p-value combination methods with a fast-computing algorithm and a theoretical guarantee on asymptotic efficiency
- Performed extensive simulations and a transcriptomics meta-analysis on the Atlas of Gene Expression in Mouse Aging Project (AGEMAP) dataset to demonstrate the universally competitive performance of the proposed methods and identify genes that related to aging or early development
- This work is published in *Statistica Sinica*

Association Study between Gene and Multiple Phenotypes of Complex Disease

Jul. 2018 – Apr. 2021

Department of Biostatistics, University of Pittsburgh

- Developed a new method from the perspective of p-value combination to efficiently aggregate heterogeneous phenotype-gene effects and adaptively select the truly contributed phenotypes; The proposed method also allows for phenotypes of different data types
- Proposed a bootstrapping-based procedure to derive the variability indices for evaluating the reliability of the selection of contributed phenotypes
- Applied the proposed method to a complex lung disease dataset and a breast cancer dataset separately to identify intriguing biological findings
- This work is published in *Genes*

Outcome-Guided Disease Subtyping using High-Dimensional Omics Data

Jun. 2019 – Jul. 2020

Department of Biostatistics, University of Pittsburgh

- Developed a unified latent generative model with feature selection to perform continuous outcome-guided clustering using omics data
- Derived the implementation algorithm of our methods based on EM and coordinate descent algorithms
- Extended the model to survival outcome by incorporating an accelerated failure time model
- Conducted expensive simulations and applied the method to a lung disease dataset with omics data to identify the disease subtypes of complex lung diseases

Robust Combination of Dependent P-Values with Unknown Dependent Structure

Oct. 2019 – Sep. 2020

Department of Biostatistics, University of Pittsburgh

- Developed a family of hypothesis tests that are robust for combining dependent p-values (in terms of type I error control) and optimal for detecting weak and sparse signals (in terms of detection boundary)
- Conducted comprehensive simulations to compare and evaluate the family of tests and provided a practical protocol for using the proposed tests
- Applied the tests to a neuroticism genome-wide association studies (GWAS) consortium data to illustrate the theoretical findings and the improvements of our proposed methods over the existing Cauchy method and identify neuroticism-related genes
- This work is published in *Statistica Sinica*

SELECTED HONORS AND AWARDS

- ASA Pittsburgh Chapter Student of the Year, 2023
- Travel award to Symposium on Data Science and Statistics (SDSS), 2022
- Travel award to Institute for Pure & Applied Mathematics (IPAM) workshop in “Emerging Opportunities for Mathematics in the Microbiome”, 2020
- Outstanding Undergraduate Thesis, 2017, Xiamen University

SELECTED PUBLICATIONS [[GOOGLE SCHOLAR](#)]

Statistical Methodology

1. **Yusi Fang**, Chung Chang, Yongseok Part, George Tseng. (2023+) “Heavy-tailed distribution for combining dependent p-values with asymptotic robustness.” *Statistica Sinica*, to appear. [[link](#)]
2. **Yusi Fang** Chung Chang, George Tseng. (2023+). “On p-value combination of independent and non-sparse signals: asymptotic efficiency and Fisher ensemble.” *Statistica Sinica*, to appear. [[link](#)]
3. Yujia Li*, **Yusi Fang***, Hung-Ching Chang, Michael Gorczyca, Peng Liu, George Tseng. (2023+). “Adaptively integrative association between multivariate phenotypes and transcriptomic data for complex diseases.” *Genes*, to appear (* co-first author)
4. **Yusi Fang**, Zhao Ren, George Tseng. (2023+). “Adaptive Fisher’s Method using weakly geometric grid for combining p-values with application to COVID-19 surveillance.” *submitted*.
5. Wenjia Wang, **Yusi Fang**, Chung Chang, George Tseng (2023+). “Accurate and ultra-efficient p-value calculation for higher criticism tests.” *Under revision in JCGS*.
6. Peng Liu, **Yusi Fang**, Zhao Ren, Lu Tang, George Tseng. (2023+). “Outcome-guided disease subtyping for high-dimensional omics data.” *submitted*. [[arxiv](#)]

Application

1. Azadeh Nasrazadani, Yujia Li, **Yusi Fang**, Osama Shah, Jennifer M. Atkinson, Joanna S. Lee, Priscilla F. McAuliffe et al. “Mixed invasive ductal lobular carcinoma is clinically and pathologically more similar to invasive lobular than ductal carcinoma.” (2023). *British Journal of Cancer*, 1-10.
2. Adria S. Mora, Mary Strange, **Yusi Fang**, Ibrahim Uygun, Lixin Zhang, George C. Tseng, Pawel Kalinski, Robert P. Edwards, and Anda M. Vlad. (2022). “Longitudinal Modulation of Loco-Regional Immunity in Ovarian Cancer Patients Receiving Intraperitoneal Chemotherapy.” *Cancers*, 14(22), 5647.
3. Brian Orr, Haider Mahdi, **Yusi Fang**, Mary Strange, Ibrahim Uygun, Mainpal Rana, Lixin Zhang, Adria Mora, Alexandra Pusateri, Esther Elishaev, Chaeryon Kang, George Tseng, William Gooding, Robert Edwards, Pawel Kalinski, Anda Vlad. (2022). “Phase I trial combining chemokine-targeting with loco-regional chemoimmunotherapy for recurrent, platinum-sensitive ovarian cancer shows induction of CXCR3 ligands and markers of type 1 immunity.” *Clinical Cancer Research*, 28(10), 2038-2049.
4. Neil Carleton, Jian Zou, **Yusi Fang**, Stephen E. Koscomb, Osama Shiraz Shah, Fangyuan Chen, Sushil Beriwal, Emilia Diego, Adam Brufsky, Steffi Oesterreich, Steven Shapiro, Robert Ferris, Leisha Emens, George Tseng, Oscar Marroquin, Adrian Lee, Priscilla McAuliffe. (2021). “Outcomes After Sentinel Lymph Node Biopsy and Radiotherapy in Older Women With Early-Stage, Estrogen Receptor-Positive Breast Cancer.” *JAMA Network Open*, 4(4), e216322-e216322.
5. Peng Liu, Silvia Liu, **Yusi Fang**, Xiangning Xue, Jian Zou, George Tseng, and Liza Konnikova. (2020). “Recent advances in computer-assisted algorithms for cell subtype identification of cytometry data.” *Frontiers in Cell and Developmental Biology*, 8, 234.

SELECTED TALKS

- (Contributed Talk) “Heavy-tailed distribution for combining dependent p-values with asymptotic robustness with application to SNP-set test.”, Joint Statistical Meetings (JSM), Washington, D.C., August 2022
- (Invited Talk) “On p-value combination of independent and frequent signals”, International Chinese Statistical Association (ICSA) Applied Statistics Symposium, Gainesville, FL, June 2022
- (Contributed Talk) “Robust aggregation of p-values with unknown dependency structure for SNP-set test”, Eastern North American Region Spring meeting (ENAR), Houston, TX, March 2022
- (Seminar Talk) “Fisher ensemble method for combining p-values with independent and non-sparse signals”, Division of Biostatistics, Medical College of Wisconsin, Milwaukee, WI, February 2023