

# WENJIA WANG

**Address:** 1800 5<sup>th</sup> Ave, Pittsburgh, PA 15219  
**Mobile Phone:** +1 (530) 5649145

**E-mail:** [wew89@pitt.edu](mailto:wew89@pitt.edu)  
**Website:** <https://wenjiaking.github.io/>

---

## EDUCATION

**University of Pittsburgh, School of Public Health** *Aug. 2020 - Apr. 2025*  
Ph.D. in Biostatistics, Advisor: [George C. Tseng](#)

**University of California, Davis** *Sep. 2018 - Dec. 2019*  
M.S. in Statistics

**East China Normal University** *Sep. 2013 - Jul. 2017*  
B.S. in Mathematics

## RESEARCH INTERESTS

**Statistics Methodology:** Statistical Computing; Bayesian Analysis; Data Integration; High-Dimensional Clustering; Statistical Learning and AI Machine Learning; Deep Learning

**Bioinformatics:** Multi-Omics Data Integration; Long-Read RNA Sequencing; Metabolome; Microbiome; CyToF; Statistical Genomics and Genetics.

**Disease Domain Experience:** Cancer; Sepsis; Atopic Dermatitis and Psoriasis; Alzheimer's Disease; Preterm Labor; Necrotizing Enterocolitis (NEC); Pediatric Inflammatory Bowel Disease (IBD).

## PUBLICATIONS

[Google Scholar](#); † indicates equal contribution

### PEER-REVIEWED ARTICLES (\*first/joint first-authored)

#### Methodological and Major Publications

- (\*P1) **Wenjia Wang**, Weihong Gu, Ron Schweitzer, Omry Koren, Soliman Khatib, George Tseng, Liza Konnikova. (2024) "In utero human intestine contains maternally derived bacterial metabolites". *Microbiome*. *Major revision*.
- (P2) Peng Liu, Yuchen Pan, Hung-Ching Chang, **Wenjia Wang**, Yusi Fang, Xiangning Xue, Jian Zou, Jessica M. Toothaker, Oluwabunmi Olaloye, Eduardo Gonzalez Santiago, Black McCourt, Vanessa Mitsialis, Pietro Presicce, Suhas G. Kallapur, Scott B. Snapper, Jia-Jun Liu, George C. Tseng, Liza Konnikova, Silvia Liu. (2024) "Comprehensive evaluation and practical guideline of gating methods for high-dimensional cytometry data: manual gating, unsupervised clustering, and auto-gating". *Briefings in Bioinformatics*. *Major revision*.
- (\*P3) **Wenjia Wang**, Yuzhen Li, Sungjin Ko, Ning Feng, Manling Zhang, Jia-Jun Liu, Songyang Zheng, Baoguo Ren, Yan P. Yu, Jian-Hua Luo, George C. Tseng, and Silvia Liu. (2024) "IFDlong: an isoform and fusion detector for accurate annotation and quantification of long-read RNA-seq data". *Genome Biology*. *Under review*.
- (\*P4) Yujia Li<sup>†</sup>, Peng Liu<sup>†</sup>, **Wenjia Wang**<sup>†</sup>, Wei Zong, Yusi Fang, Zhao Ren, Lu Tang, George C. Tseng. (2024) "Outcome-guided disease subtyping by generative model and weighted joint likelihood in transcriptomic applications". *Annals of Applied Statistics*. 18:1947-1964. (<sup>†</sup>*equal contribution*).
- (\*P5) **Wenjia Wang**, Yusi Fang, Chung Chang, George Tseng. (2024) "Accurate and ultra-efficient p-value calculation for higher criticism tests". *Journal of Computational and Graphical Statistics*. 33:2, 463-476. (*The paper received 2023 ASA Statistical Computing and Graphics Student Paper Award*)

#### Collaborative Publications

- (P6) Lixin Zhang, Yusi Fang, Ibrahim Uygun, Danyang Li, Mary Strange, Syed K Zaidi, **Wenjia Wang**, Julia Knight, Mackenzy Radolec, Esther Elishaev, Joan F Brozick, Allison Edwards, George Tseng, Sandra Cascio, Ronald

Buckanovich, Robert P Edwards, Anda M Vlad. (2024) “Novel murine cell lines with defined mutations model different histological subtypes of epithelial ovarian cancer”. *Disease Models and Mechanism. Major revision.*

- (P7) Silas A. Buck, Sophie A. Rubin, Tenzin Kunkhyen, Christoph D. Treiber, Xiangning Xue, Lief E. Fenno, Samuel J. Mabry, Varun R. Sundar, Zilu Yang, Divia Shah, Kyle D. Ketchesin, Darius D. Becker-Krail, Iaroslavna Vasylieva, Megan C. Smith, Florian J. Weisel, **Wenja Wang**, M. Quincy Erickson-Oberg, Emma I. O’Leary, Eshan Aravind, Charu Ramakrishnan, Yoon Seok Kim, Yanying Wu, Matthias Quick, Jonathan A. Coleman, William A. MacDonald, Rania Elbakri, Briana R. De Miranda, Michael J. Palladino, Brian D. McCabe, Kenneth N. Fish, Marianne L. Seney, Stephen Rayport, Susana Mingote, Karl Deisseroth, Thomas S. Hnasko, Rajeshwar Awatramani, Alan M. Watson, Scott Waddell, Claire E. J. Cheetham, Ryan W. Logan, Zachary Freyberg. (2024) “Sexually dimorphic mechanisms of VGLUT-mediated protection from dopaminergic neurodegeneration”. *Neuron. Under review.*
- (P8) Lauren Smith, Eduardo Gonzalez Santiago, Chino Eke, Weihong Gu, **Wenja Wang**, Dhana Llivichuzhca-Loja, Tessa Kehoe, Kerri St Denis, Madison Strine, Sarah Taylor, George Tseng, Liza Konnikova. (2024) “Human milk supports robust intestinal organoid growth, differentiation, and homeostatic cytokine production”. *Gastro Hep Advances. Accepted.*
- (P9) Yuan Gui, Yanbao Yu, **Wenja Wang**, Yuanyuan Wang, Hanyue Lu1, Sarah Mozdierz, Eskander, Kirolos, Yi-Han Lin, Hanwen Li, Xiaojun Tian, Silvia Liu, Dong Zhou. (2024) “Proteomes characterization of liver-kidney comorbidity after microbial sepsis”. *FASEB Journal. 38,7: e23597.*
- (P10) Silvia Liu, Caroline Obert, Yan-Ping Yu, Junhua Zhao, Baoguo Ren, Kelly Brease, Benjamin Krajacich, **Wenja Wang**, Kyle Metcalfe, Mat Smith, Tuval Ben-Yehzekel, Jianhua Luo. (2024) "Utility analyses of AVITI sequencing chemistry". *BMC Genomics. 25, 778.*
- (P11) Silvia Liu, Yan-Ping Yu, Bao-Guo Ren, Tuval Ben-Yehzekel, Caroline Obert, Mat Smith, **Wenja Wang**, Alina Ostrowska, Alejandro Soto-Gutierrez, Jianhua Luo. (2024) "Long-read single-cell sequencing reveals expressions of hypermutation clusters of isoforms in human liver cancer cells". *eLife. 12:RP87607.*

## ARTICLES UNDER PREPARATION

- (S1) **Wenja Wang**, Yusi Fang, Chung Chang, George C. Tseng. (2024) “Combining p-values: historical development, recent advances and future opportunities”. (*ready for submission*)
- (S2) Balamurugan Ramatchandirin, **Wenja Wang**, Marie Amalie Balamurugan, Yasemin Alnahaas, Suneetha Desiraju, Arjun Subramanya, Juanitaa George Raj, Zainab D Lawal, Megan Ferris, Liza Konnikova, George C. Tseng, Krishnan MohanKumar. (2024) “Immune landscape in liver of neonatal mice with phlebotomy-induced anemia”. (*ready for submission*)
- (S3) **Wenja Wang**, George C. Tseng. (2024) “A Bayesian outcome-guided clustering framework via consensus clustering for molecular disease subtyping”. (*under preparation*)
- (S4) **Wenja Wang**, George C. Tseng. (2024) “Bayesian consensus clustering for integration of metabolomics and microbiomics data”. (*under preparation*).
- (S5) Sarah Wang, **Wenja Wang**, George C. Tseng. (2024) “Efficient p-value calculation for adaptive rank truncated p-value combination tests”. (*under preparation*)
- (S6) Danyang Li, **Wenja Wang**, Wei Zong, George C. Tseng. (2024) “Multivariate guided disease subtyping for high-dimensional omics data”. (*under preparation*)
- (S7) Hung-Ching Chang, **Wenja Wang**, George C. Tseng. (2024) “The CCombP R package for p-value combination tests”. (*under preparation*)
- (S8) Hung-Ching Chang, **Wenja Wang**, Danyang Li, Sarah Wang, Yuxin Shi, Ria Sia, Jian Zou, George C. Tseng. (2024) “Evaluation of copy number variation inference methods for tumor subclone identification using single-cell RNA-sequencing data”. (*under preparation*)
- (S9) Jiajun Liu, **Wenja Wang**, Dong Zhou, Paul Monga, Silvia Liu (2024). “Meta-analytic framework for robust biomarker and cell subtype identification on single-cell and spatial transcriptomics data”. (*under preparation*)

## PROFESSIONAL EXPERIENCE

- **University of Pittsburgh**, Pittsburgh, PA, USA                      **Graduate Student Researcher**                      *Aug. 2020 - present*  
Provided statistical consulting and explored biological insights by high-throughput multi-omics data analysis mainly for
  - ✓ **Liza Konnikova Lab** (relocated to School of Medicine, Yale University in 2020): contributed to significant findings in microbial metabolites in fetal intestines by metabolomics data analysis (**\*P1**), fetal intestinal organoid development by cytometry data analysis (**P8**), fetal immune cells development by scRNA data analysis.
  - ✓ **Dr. Silvia Liu**, School of Medicine, University of Pittsburgh: lead the software development for long-read RNA-seq data analysis (**\*P3**), contributed to methods evaluation for cytometry data analysis (**P2**), findings about immune response in organoids by pathway enrichment analysis (**P9**), sequencing techniques evaluation (**P10-P11**).
- **Pfizer, Inc**, Cambridge, MA, USA                                      **Statistics Intern**                                      *May. 2022 – Aug. 2022*  
Developed a shiny app to compare gene signature improvement for Atopic Dermatitis and Psoriasis by Pfizer's therapies to competitor drugs. Predicted drug response using baseline transcriptomic data after data integration, comprehensive comparison of AI models such as convolutional neural network, variational auto-encoder, etc., toward precision medicine.
- **Eli Lilly and Company**, Indianapolis, IN, USA                      **Statistics Intern**                                      *May. 2024 – Aug. 2024*  
Learned the biological mechanism for the side effect of amyloid-targeting therapies in Alzheimer's disease by analyzing the longitudinal proteomics data combined with clinical data. Developed an efficient predictive model with delicate feature selection given patient's baselines to identify high-risk patients of occurring the adverse event.

## TEACHING EXPERIENCE

### Regular Course Teaching

**Spring 2023 & 2024    Teaching Fellow for BIOST 2094 Advanced R Computing course, University of Pittsburgh**

- ✓ An R programming course in advanced graduate level, taught by two PhD students under Dr. George Tseng's supervision
- ✓ Respectively 16 and 15 graduate students enrolled in Spring 2023 and Spring 2024 semester
- ✓ Lectured half of the classes, supervised and graded final projects

### Short Course Teaching

**April 2023                      Instructor for ASA-SSGG Short Course Series: Selective Introduction to Multi-Omics Analysis**

- ✓ One of the four instructors to teach the annual online short course hosted by ASA Section on Statistics in Genomics and Genetics (**125 paid students** and overall course rate of 4-5 on a scale 1-5 based on Google feedback survey)
- ✓ Delivered one of the four zoom sessions (45 minutes of lecture and 45 minutes of lab session): Overview of multi-omics data analysis and horizontal data integration. Answered questions from the audience in all zoom sessions
- ✓ Prepared for course materials covering statistical and bioinformatic methods for multi-omics data analysis with focus on horizontal meta-analysis, disease subtyping clustering, dimension reduction, causal mediation and single cell multi-omics
- ✓ Availability of all slides and data/code on [GitHub](#)

### Course Capable of Teaching

- **Advanced Statistics Courses**: Statistical Computing (optimization theory and R/Python implementation); Bayesian Analysis; Statistical Learning and Machine Learning; Deep Learning
- **Bioinformatics Courses**: Omics Data Analysis; Computational Genomics; Sequencing Technologies and Analysis
- Flexible and confident in teaching *standard statistics and biostatistics courses* when the department needs

## CONFERENCE TALKS

- 2024                      "IFDlong: an isoform fusion detector on long-read RNA-seq data", JSM, Portland, OR, USA (**\*P3**)
- 2024                      "Isoform and fusion detection on bulk and single-cell long-read RNA-seq data", Great Lakes Bioinformatics Conference, Pittsburgh, PA, USA (**\*P3**)

- 2023 **[invited]** “Accurate and Ultra-Efficient p-Value Calculation for Higher Criticism Tests”, JSM, Toronto, Ontario, Canada (ASA Statistical Computing and Graphics Student Paper Award) **(\*P5)**
- 2023 “Accurate and Ultra-Efficient p-Value Calculation for Higher Criticism Tests”, ENAR, Nashville, TN, USA **(\*P5)**

## OPEN-SOURCE SOFTWARE

- 2024 **ogClust** - [R package](#) (Developer); *Liu, Li, and Wang, et al., 2024 (\*P4)*  
An R package that implements two outcome-guided clustering methods for disease subtyping: the generative model (*ogClust<sub>GM</sub>*) and the weighted joint likelihood model (*ogClust<sub>WJL</sub>*)
- 2024 **IFDlong** - [Bash tool](#) (Developer); *Wang, et al., 2024+ (\*P3)*  
A unified software that performs long-read RNA-seq alignment, isoform/fusion annotation and quantification
- 2023 **HCp** - [R package](#) (Developer); *Wang, et al., 2024 (\*P5)*  
An R package that includes functions of different methods for p-value computation of Higher Criticism test

## SELECTED AWARDS & CERTIFICATES

### Major

- 2024 ASA Pittsburgh Chapter Student of the Year
- 2023 ASA Sections on Computing and Graphics Student Paper Awards
- 2012 Second Prize in National Olympic Biology Competition

### Others

- 2022 SAS Certified Specialist: Base Programming
- 2017 Excellent Graduate of East China Normal University
- 2017 American Society of Actuaries: Financial Mathematics Course
- 2017 American Society of Actuaries: Probability Course
- 2014-2016 First-class & Second-class National Scholarship: for top 3% and 10% students respectively

## TECHNICAL SKILLS

- **Programming Language**: Proficient in R, bash, Python, LaTeX, SQL, SAS, MATLAB and STATA
- **Statistical Methodologies**: Familiar with various statistical models, Bayesian inference, statistical computation, optimization, AI models including CNN, RNN, Hidden Markov Models, Bayesian Networks, Reinforcement Learning, Recommender Systems, and computer vision such as image processing, feature extraction, and segmentation
- **Bioinformatics and Genomic Data Analysis**: DNA-seq, bulk and single-cell RNA-seq, spatial transcriptomics, etc.

## PROFESSIONAL SERVICES

- **Leadership, Mentoring and Services**
  - ✓ Mentored three junior Ph.D. students in biostatistics (with papers **(S5-S7)**), University of Pittsburgh, 2023-2024
  - ✓ Mentored a summer undergraduate student in computer science, University of Rochester, 2022
  - ✓ Pittsburgh ASA Student Chapter Executive Committee, University of Pittsburgh, 2021-2023
  - ✓ Organized Internship Panel, Pittsburgh ASA Student Chapter, University of Pittsburgh, 2021 (invited UPitt biostatistics and statistics students to share their intern experiences; organized and hosted talks)
  - ✓ Computing server manager for Tseng’s lab: maintained lab computing servers and managed lab allocation on [Center For Research Computing](#), University of Pittsburgh, 2020-2024
- **Journal Peer Review**: The Journal of Emerging Investigators (2022-2024), and PLOS Digital Health (2022)
- **Memberships**: American Statistical Association (ASA) – Statistics in Genomics and Genetics, Eastern North American Region (ENAR)