

# Xinjun Wang

Email: XIW119@pitt.edu

Immigration status: U.S. Citizen

## EDUCATION

---

- University of Pittsburgh**, Pittsburgh, PA Anticipated 2022  
Ph.D. in Biostatistics  
Dissertation advisors: Ying Ding and Wei Chen
- University of Washington**, Seattle, WA Aug 2017  
M.S. in Biostatistics  
Thesis advisor: Xiao-Hua (Andrew) Zhou
- University of Iowa**, Iowa City, IA May 2015  
B.S. in Mathematics  
B.A. in Biochemistry

## WORK EXPERIENCE

---

- UPMC Children's Hospital of Pittsburgh**, Pittsburgh, PA Aug 2019 – present  
*Graduate Student Researcher*  
Supervisor: Wei Chen, Ph.D.
  - Develop novel statistical methods for single-cell multi-omics data
  - Analyze scRNA-seq, single-cell multi-omics (e.g., CITE-seq, DOGMA-seq) and spatial transcriptomics data for novel biological discoveries
- University of Pittsburgh Department of Occupational Therapy**, Pittsburgh, PA, Sep 2017 – Aug 2019  
*Graduate Student Researcher*  
Supervisor: Lauren Terhorst, Ph.D.
  - Analyzed post-acute-care data to support public policy making
- U.S. Food & Drug Administration (FDA)**, White Oak, MD Jun 2018 – Aug 2018  
*ORISE Fellow (Division of Biometrics VII)*  
Supervisor: Hana Lee, Ph.D.
  - Examined the utility of doubly-robust estimation in regulatory setting
- Peking University Clinical Research Institute (PUCRI)**, Beijing, China Aug 2016 - Sep 2016  
*Clinical Research Intern (Biostatistics)*  
Supervisor: Xiaoyan Yan, Ph.D.
  - Statistical analysis of clinical trials data
- University of Iowa Department of Obstetrics and Gynecology**, Iowa City, Iowa Mar 2011 - May 2015  
*Undergraduate Research Assistant*  
Supervisor: Xiangbing Meng, Ph.D.
  - Studied MTDH function in endometrial cancer and ovarian cancer

## RESEARCH FUNDING

---

- Funding Agency: UPMC Children's Hospital of Pittsburgh  
Category: Research Advisory Committee (RAC) Graduate Student Fellowship  
Grant Title: Machine Learning and Statistical Methods for Analyzing Single-cell Multi-omics Data  
Role on Grant: **Principal Investigator**

Years Inclusive: 7/1/2020 - 6/30/2022  
Total Direct Costs: \$80,000 (\$40,000 per year)

- Funding Agency: NIH/Clinical and Translational Science Institute, University of Pittsburgh  
Grant Number: UL1TR001857  
Grant Title: Joint Analysis of Single-cell Multi-omics Data  
Role on Grant: **co-Principal Investigator**  
Years Inclusive: 9/1/2019 - 8/31/2020  
Total Direct Costs: \$10,000

## **AWARDS**

---

Outstanding Graduate Student Researcher Award, Pitt Biostatistics, 2021  
Biostatistics Doctoral Award, Pitt GSPH Dean's Day, 2021  
ASA Student of the Year, Pittsburgh Chapter, 2021  
Outstanding Student Research Award, Pitt Biostatistics Research Day, 2021  
ICSA Student Paper Award, 2020

## **TEACHING EXPERIENCE**

---

*Teaching Assistant* (University of Pittsburgh, Fall 2019)

- BIOST2066: Applied Survival Analysis: Methods and Practice (independently gave a lecture on *competing risks in survival analysis*)

## **PUBLICATIONS**

---

+: co-first author

### *Methodology*

- **Wang X**, Xu Z, Zhou X, Zhang Y, Lafyatis R, Chen K, Huang H, Ding Y, Duerr R, Chen W. SECANT: a biology-guided semi-supervised method for clustering, classification, and annotation of single-cell multi-omics. Under review.
- Zhang Y<sup>+</sup>, **Wang X**<sup>+</sup>, Luo L, Chen W, Huang H. Fast Estimation of Concordance Matrix for Single-Cell Multi-Omics Analysis. Submitted.
- Wei Y, **Wang X**, Chew EY, Ding Y. Confident identification of subgroups from SNP testing in RCTs with binary outcomes. *Biometrical Journal*. 2021. DOI: 10.1002/bimj.202000170.
- Xin H, Lian Q, Jiang Y, Luo J, **Wang X**, Erb C, Xu Z, Zhang X, Heidrich-O'Hare E, Yan Q, Duerr RH, Chen K, Chen W. GMM-Demux: sample demultiplexing, multiplet detection, experiment planning, and novel cell-type verification in single cell sequencing. *Genome biology*. 2020 Dec;21(1):1-35.
- **Wang X**<sup>+</sup>, Sun Z<sup>+</sup>, Zhang Y, Xu Z, Xin H, Huang H, Duerr RH, Chen K, Ding Y, Chen W. BREM-SC: a bayesian random effects mixture model for joint clustering single cell multi-omics data. *Nucleic acids research*. 2020 Jun 19;48(11):5814-24.

### *Book Chapters*

- Ding Y, Wei Y, **Wang X**, Hsu JC. Testing SNPs in Targeted Drug Development. Book Chapter In: Handbook of Multiple Comparisons. *Chapman & Hall/CRC*, 2021+. In Press.

- Ding Y, Wei Y, **Wang X**. Logical Inference on Treatment Efficacy When Subgroups Exist. Book Chapter In: Design and Analysis of Subgroups with Biopharmaceutical Applications 2020 (pp. 209-228). Springer, Cham.

#### *Collaborative Papers*

- Mirizio E, Liu C, Yan Q, Waltermire J, Mandel R, Schollaert K, Konnikova L, **Wang X**, Chen W, Torok K. Genetic Signatures from RNA sequencing of Pediatric Localized Scleroderma (LS) Skin. *Frontiers in Pediatrics*. 2021;9. DOI: 10.3389/fped.2021.669116.
- Schutt C, Mirizio E, Salgado C, Reyes-Mugica M, **Wang X**, Chen W, Grunwaldt L, Schollaert KL, Torok KS. Transcriptomic evaluation of pediatric localized scleroderma skin with histological and clinical correlation. *Arthritis & Rheumatology (Hoboken, NJ)*. 2021 Apr 12. DOI: 10.1002/art.41758.
- Mirizio E, Tabib T, **Wang X**, Chen W, Liu C, Lafyatis R, Jacobe H, Torok KS. Single-cell transcriptome conservation in a comparative analysis of fresh and cryopreserved human skin tissue: pilot in localized scleroderma. *Arthritis research & therapy*. 2020 Dec;22(1):1-0.

#### *Biochemistry and Cell Biology*

- Meng X, Yang S, Zhang Y, **Wang X**, Goodfellow RX, Jia Y, Thiel KW, Reyes HD, Yang B, Leslie KK. Genetic deficiency of Mtdh gene in mice causes male infertility via impaired spermatogenesis and alterations in the expression of small non-coding RNAs. *Journal of Biological Chemistry*. 2015 May 8;290(19):11853-64.
- Yang S, Jia Y, Liu X, Winters C, **Wang X**, Zhang Y, Devor EJ, Hovey AM, Reyes HD, Xiao X, Xu Y. Systematic dissection of the mechanisms underlying progesterone receptor downregulation in endometrial cancer. *Oncotarget*. 2014 Oct;5(20):9783.
- Yang S, Xiao X, Jia Y, Liu X, Zhang Y, **Wang X**, J Winters C, J Devor E, Meng X, W Thiel K, K Leslie K. Epigenetic modification restores functional PR expression in endometrial cancer cells. *Current pharmaceutical design*. 2014 Apr 1;20(11):1874-80.
- Meng X, Dizon DS, Yang S, **Wang X**, Zhu D, Thiel KW, Leslie KK. Strategies for molecularly enhanced chemotherapy to achieve synthetic lethality in endometrial tumors with mutant p53. *Obstetrics and gynecology international*. 2013 Dec 7;2013.
- Meng X, Laidler LL, Kosmacek EA, Yang S, Xiong Z, Zhu D, **Wang X**, Dai D, Zhang Y, Wang X, Brachova P. Induction of mitotic cell death by overriding G2/M checkpoint in endometrial cancer cells with non-functional p53. *Gynecologic oncology*. 2013 Mar 1;128(3):461-9.
- Meng X, Zhu D, Yang S, **Wang X**, Xiong Z, Zhang Y, Brachova P, Leslie KK. Cytoplasmic Metadherin (MTDH) provides survival advantage under conditions of stress by acting as RNA-binding protein. *Journal of Biological Chemistry*. 2012 Feb 1;287(7):4485-91.
- Milhem MM, Knutson T, Yang S, Zhu D, **Wang X**, Leslie KK, Meng X. Correlation of MTDH/AEG-1 and HOTAIR expression with metastasis and response to treatment in sarcoma patients. *Journal of cancer science & therapy*. 2011 Dec 29(4).

#### **PRESENTATIONS**

---

- (Virtual Poster) “SECANT: a biology-guided semi-supervised method for clustering, classification, and annotation of single-cell multi-omics”, *Advancing Research through Computing (ARC) Symposium*, Apr 2021.
- (Virtual Poster) “SECANT: a biology-guided semi-supervised method for clustering, classification, and annotation of single-cell multi-omics”, *GSPH Dean’s Day*, Apr 2021.

- (Virtual Talk) “SECANT: a biology-guided semi-supervised method for clustering, classification, and annotation of single-cell multi-omics”, *ENAR*, March 2021.
- (Virtual Talk) “SECANT: a biology-guided semi-supervised method for clustering, classification, and annotation of single-cell multi-omics”, *Pitt Biostatistics Student Research Day*, March 2021.
- (Invited Virtual Talk) “SECANT: a biology-guided semi-supervised method for clustering, classification, and annotation of single-cell multi-omics”, *The Pulmonary Medicine Research Conference*, University of Pittsburgh, December 2020.
- (Virtual Talk) “BREM-SC: A Bayesian Random Effects Mixture Model for Clustering Single Cell Multi-Omics Data”, *ICSA*, December 2020.
- (Virtual Talk) “BREM-SC: A Bayesian Random Effects Mixture Model for Clustering Single Cell Multi-Omics Data”, *ENAR*, March 2020.
- (Poster) “BREM-SC: A Bayesian Random Effects Mixture Model for Clustering Single Cell Multi-Omics Data”, *Pitt Biostatistics Student Research Day*, Pittsburgh, PA, February 2020.
- (Invited Talk) “A joint clustering method for single cell multi-omics data”, *The Pulmonary Medicine Research Conference*, University of Pittsburgh, Pittsburgh, PA, October 2019.
- (Invited Talk) “A joint clustering method for single cell multi-omics data”, *Center for Systems Immunology Research Forum*, University of Pittsburgh, Pittsburgh, PA, September 2019.
- (Poster) “A Comparison and Assessment of Tree-based Methods for Subgroup Identification with Time-to-Event Data”, *GSPH Dean’s Day*, Pittsburgh, PA, April 2019.
- (Poster) “A Comparison and Assessment of Tree-based Methods for Subgroup Identification with Time-to-Event Data”, *ENAR*, Philadelphia, PA, March 2019.
- (Poster) “A Comparison and Assessment of Tree-based Methods for Subgroup Identification with Time-to-Event Data”, *Pitt Biostatistics Student Research Day*, Pittsburgh, PA, March 2019.
- (Talk) “Examining the Utility of Doubly-Robust Estimation in Regulatory Setting”, *ORISE Statistical Conference*, FDA, White Oak, MD, August 2018.
- (Poster) “Examining the Utility of Doubly-Robust Estimation in Regulatory Setting”, *Annual Summer Student Scientific Poster Day Event*, FDA, White Oak, MD, August 2018.

## **STATISTICAL PACKAGES**

---

*SECANT*: A Python package (with GPU acceleration) for a novel biology-guided semi-supervised method for clustering, classification, and annotation of single-cell multi-omics. Available on GitHub ([link](#)).

*BREM-SC*: An R package for joint clustering droplet-based CITE-seq data. Available on GitHub ([link](#)).

## **COMPUTATIONAL SKILLS**

---

R, Python (PyTorch, Scikit-learn, NumPy, Pandas, Matplotlib, etc.), UNIX shell scripting, Stata, SAS, PASS, cloud-based Jupyter notebook environment (e.g., Google Colab)

## **PROFESSIONAL AFFILIATIONS**

---

Member of American Statistical Association (ASA)

Member of Eastern North American Region International Biometric Society (ENAR)

Member of International Chinese Statistical Association (ICSA)

## **SERVICES**

---

- Student representative, Pitt Biostatistics Chair Search Committee, 2021-present
- Group meeting and journal club organizer, Dr. Wei Chen's Lab, 2019-present
- Executive Committee Member, National Institute of Statistical Sciences (NISS) Graduate Student Network, 2020-2021
- President, ASA Pittsburgh Student Chapter, 2020-2021
- Session chair, ENAR, 2021
- Student Host for Pitt BIOST accepted applicants, 2021
- Session chair, ENAR, 2020
- Student Ambassador for Pitt GSPH accepted applicants, 2020
- Student Host for Pitt BIOST accepted applicants, 2020
- Student volunteer, Lifetime Data Science (LiDS) Conference, 2019