CURRICULUM VITAE of Wenyi Wang

CONTACT

E-mails:	wwang7@mdanderson.org
Mailing address	Department of Bioinformatics and Computational Biology 7007 Bertner Ave Houston, TX 77030
Website:	http://odin.mdacc.tmc.edu/∼wwang7

EDUCATION

2003-2007 JOHNS HOPKINS BLOOMBERG SCHOOL OF PUBLIC HEALTH Baltimore, MD
PhD, Biostatistics

2001-2003 COLUMBIA UNIVERSITY COLLEGE OF PHYSICIANS AND SURGEONS
MA, Human nutrition New York City, NY

1997-2001 FUDAN UNIVERSITY Shanghai, China
BS, Honor Science Program, Biology

PROFESSIONAL EXPERIENCE

9/2020 - present *Tenured Professor*
Department of Bioinformatics and Computational Biology
Department of Biostatistics (joint appointment)
University of Texas M. D. Anderson Cancer Center, Houston TX

7/2021 - present *Adjunct Professor*
Department of Statistics,
Rice University, Houston TX

6/2013 - present *Adjunct Professor*
Department of Statistics,
Texas A&M University, College Station TX

9/2015 - 8/2020 *Tenured Associate Professor*
Department of Bioinformatics and Computational Biology
University of Texas M. D. Anderson Cancer Center, Houston TX

9/2014 - 08/2018 *Program co-Director*
Quantitative Sciences (Biostatistics, Bioinformatics and Systems Biology),
The University of Texas Graduate School of Biomedical Sciences at Houston

5/2017 - 8/2017 *Visiting scientisit (Sabbatical)*
Cancer, aging and somatic mutation group (CASM)
The Wellcome Trust Sanger Institute, Hinxton UK

9/2010 - 8/2015 *Tenure-track Assistant Professor*
Department of Bioinformatics and Computational Biology,
University of Texas M. D. Anderson Cancer Center, Houston TX

8/2007 - 8/2010 *Postdoctoral Fellow*
Professor Terry Speed, Department of Statistics, University of California at Berkeley
Professor Ron Davis, Stanford Genome Technology Center, Stanford University

SELECTED PUBLICATIONS

1. Cao S, et al. **Wang W**✉ Tumor cell total mRNA expression shapes the molecular and clinical phenotype of cancer. *Nature Biotechnology*, 40, pages1624-1633 (2022). ePub June 13 2022. https://www.nature.com/articles/s41587-022-01342-x
   Wang lab identified the first pan-cancer biomarker, which is a feature of cancer cell plasticity, through an integrative deconvolution of matched DNA/RNAseq data.

2. Dentro S, et al. Portraits of genetic intra-tumour heterogeneity and subclonal selection across cancer types. *Cell*. 2021 Apr 15; 184(8): 2239-2254.e39. *Wang is a co-last author for leading the effort in consensus subclonal reconstruction of the PCAWG WGS data. Selected for the MD Anderson Wall of Science and Excellence in Science Webpage.*

3. Pan-cancer analysis of whole genomes. The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. *Nature*. 2020 Feb 5;578:82–93.
   This paper summarizes major efforts and findings of the PCAWG consorium over the analysis of 2,658 whole genomes of tumor samples during 2014-2020. We led efforts in two computational tasks: MuSE mutation calling, and consensus subclonal reconstruction of the PCAWG WGS data.

4. Wang Z, Morris JS, Cao S, Ahn J, Liu R, Tyekucheva S, Li B, Lu W, Tang X, Wistuba II, Bowden M, Mucci L, Loda M, Parmigiani G, Holmes CC, **Wang W**✉. Transcriptome deconvolution of heterogeneous tumor samples with immune infiltration. *iScience*. 2018 Nov 30;9:451-460. doi: 10.1016/j.isci.2018.10.028.
   This paper describes an advanced model, to deconvolve RNA sequencing data into tumor, stroma and immune-specific expression components. It uniquely dissect a whole tumor sample into three distinct compartments, estimating both mixing proportions and sample-specific expression levels. Bioconductor R package: DeMixT.

5. Nguyen NH, Dodd-Eaton EB,Corredor JL, Woodman-Ross J, Green S, Hernandez ND, Gutierrez Barrera AM, Arun BK, **W Wang**. Validating risk prediction models for multiple primaries and competing cancer outcomes in families with Li-Fraumeni syndrome using clinically ascertained data at a single institute. *JCO*, 2024 Apr 3;:JCO2301926. doi: 10.1200/JCO.23.01926. [Epub ahead of print]

6. Ji S, Zhu T, Sethia A, **Wang W**. Accelerated somatic mutation calling for whole-genome and whole-exome sequencing data from heterogenous tumor samples. *Genome Research*, 2024 May 3;. doi: 10.1101/gr.278456.123. [Epub ahead of print] *This paper provides an accelerated solution for mutation calling: MuSE 2, with a benchmarking study.*

7. Shin SJ, Ying Yuan✉, Strong LC, Bojadzieva J, **Wang W**✉. Bayesian semiparametric estimation of cancer-specific age-at-onset penetrance with application to Li-Fraumeni syndrome. *JASA*. 2018 doi: 10.1080/01621459.2018.1482749.
   This paper describes a novel statistical approach to estimate competing cancer outcomes over time for individuals with deleterious mutations in TP53, using family history data. Its application to data from families with Li-Fraumeni syndrome showed breast cancer occurrs earlier in life in TP53 mutation carriers as compared to BRCA1/2 carriers.

8. Fan Y, Xi L, Hughes DST, Zhang J, Zhang J, Futreal AP, Wheeler DA, **Wang W**✉. Accounting for tumor heterogeneity using a sample-specific error model improves sensitivity and specificity in mutation calling for sequencing data. *Genome Biology.* 2016 Aug 24;17(1):178. doi:

10.1186/s13059-016-1029-6.
This paper describes a Bayesian Markov model to call somatic mutations based on evolutionary principles: Markov Substitution model for Evolution, or MuSE. MuSE is particularly advantageous in calling subclonal somatic mutations. It was used to generate consensus mutation calls on ICGC ∼2,700 WGS data and TCGA ∼11,000 WES data.

9. Ahn J, Yuan Y, Parmigiani G, Suraokar MB, Diao L, Wistuba II, and **Wang W**✉. DeMix: deconvolution for mixed cancer transcriptomes. *Bioinformatics* 2013 doi: 10.1093/bioinformatics/btt301. *This paper describes a new statistical method, DeMix, to deconvolve RNA sequencing data into tumor and stroma-specific expression components. DeMix was used for the RNAseq deconvolution of 11,000 tumor samples across 30 cancer types at TCGA.*

PROFESSIONAL ACTIVITIES

Institutional activities: Member of the Clinical Research Advisory Committee, Member of the Rice-MDACC joint biostatistics program admissions committee, Member of the GSBS (graduate school) membership committee, Member of the GSBS academic standards committee, Member of the Task Force to develop the new core curriculum for GSBS, Member of the GSBS admissions committee, Faculty senate member (9.2021-9.2023), GSBS student fellowship review panel (2022), GSBS QS steering committee member (9.2021-9.2023), Faculty member of the TRIUMPH (Translational Research in Multidisciplinary Programs) training program (9.2021-), REACT (Research Enablement, Activities and Conduct) committee member (10.2021-). UTGSBS Dean search committee (03.2022 - 09.2022).

Biostatistics service on clinical trials:
Pembrolizumab in Combination with Dabrofenib and Trametinib as a Neoadjuvant Strategy Prior to surgery in BRAF-mutated Anaplastic Thyroid Cancer. Sponsor: Merck. 2020-present (Trial initiated)
Neoadjuvant/Adjuvant Treatment with LOXO-292 in RET-Altered Thyroid Cancers. 2020-present (Trial initiated)

Professional membership: American Statistics Association, ENAR/International Biometric Society, International Mathematical Society, International Society for Computational Biology, American Society of Human Genetics, International Chinese Statistical Association, International Society of Bayesian Analysis, American Association for Cancer Research ERCIM Working Group on Computational and Methodological Statistics, Program Chair 2018: ASA Section on Statistics in Genetics and Genomics.

Journal referee: Nature Biotechnology, Nature Genetics, Science, Nature Methods, Molecular Cell, JAMA Oncology, Developmental Cell, Genome Biology, Genome Medicine, Journal of Clinical Oncology, Journal of Clinical Oncology Precision Oncology Nature Communication, BMC Research Notes, Biometrics, Proceedings of the National Academy of Sciences, Genome Research, European Journal of Human Genetics, Nucleic Acids Research, Cancer Research, Statistical Applications in Genetics and Molecular Biology, Biostatistics, IEEE/ACM Transactions on Computational Biology and Bioinformatics, Journal of Biomedicine and Biotechnology, Molecular Carcinogenesis, Annals of Applied Statistics, Bayesian analysis, Clinical Genetics, British Journal of Dermatology, British Journal of Cancer, Biometrics Journal of Applied Statistics, Bioinformatics, PLOS Computation Biology, NeurIPS 2019 paper reviewer, Science Translational Medicine, JASA, Cancer Cell

Grant referee: The Netherlands Organization for Health Research and Development,
Cancer Research-UK,
National Sciences Centre Poland,
Univ. of Cambridge Medical Research Council-Biostatistics (Quinquennial review),
The American Cancer Society,
Swiss National Science Foundation,
Medical Research Council - UK,
National Sciences and Engineering Research Council of Canada.

Study section member:
NIH Cancer, Heart, and Sleep Epidemiology (CHSA) Study Section (2017-2021)
DoD Prostate Cancer Research Program Data Science Award review panel (2022)
NIH Special Emphasis panel (2024)
NHGRI R25 - Medical Student Curriculum and Short Courses Meeting (2024)

Conference program committee: American Association for Cancer Research (subcommittee on Bioinformatics and Computational Biology 2015), RECOMB2015 Satellite Workshop on Cancer Computational Biology and Sequencing, Bayesian Biostatistics and Bioinformatics Conference 2014, American Society of Human Genetics 2016, 2017 (reviewer for Cancer), ENAR 2018 educational program advisory committee, Program Chair 2018: ASA Section on Statistics in Genetics and Genomics, ICSA 2020 excecutive committee member/short course committee chair, ACM-BCB 2023 program co-chair, MD Anderson Annual Symposium in Leading Edge Cancer Research 2023 Organizing Committee member.

Other consortium membership: LiFE consortium for Li-Fraumeni studies; the Cancer Genome Atlas projects, PanCanAtlas working groups; ICGC PanCancer Analysis of Whole Genomes (PCAWG) evolution and heterogeneity working group; Human Cell Atlas Seed Network, Clinical Proteomic Tumor Analysis Consortium (NCI), Clinical Genetics Consortium Computational Subgroup (NCI), Variant Curation Expert Panel (NCI).

GRANT SUPPORT

FUNDED

Principal Investigator (20%)
7/8/2022-6/30/2027
Statistical methods for genomic analysis of heterogeneous tumors
NIH/NCI (R01, $2,000,000)

Principal Investigator (15%)
9/30/2022-8/31/2025
An integrated genomic definition and therapeutic strategy for androgen indifferent prostate cancers
DoD (Prostate Cancer Program Data Science Award, $1,619,425)

Co-Principal Investigator (8%)
8/31/2020-8/30/2024 (NCE)
Improving risk prediction for Li-Fraumeni Syndrome: A practical tool for clinical health care providers
($853,943, $284,481/year)
CPRIT
PI: Banu Arun/Wenyi Wang

Co-Investiator (10%)
1/23/2023-1/22/2025
Targeting B7-H3 in Aggressive Variant Prostate Cancer
Prostate Cancer Foundation Challenge Award ($499,845)
PI: Di Zhao

Co-Investigator (10%)
03/01/2023-02/28/2026
Monitoring of Field Effects for the Detection of Bladder Cancer
CPRIT ($912,376)
PI: Bogdan Czerniak

Co-Investiator (5%)
09/01/2022-8/31/2027
Novel Approaches Targeting B7-H3 in Castration-resistant Prostate Cancer
NIH/NCI (R01 $3,595,130)
PI: Di Zhao

Co-Investigator (2%)
2/1/2020-1/31/2025
Role of p53 Missense Mutations on Tumorigenesis in Vivo
NIH/NCI ($1,250,000, $250,000/year)
PI: Guillermina Lozano

COMPLETED (selected)

Principal Investigator (27%)
4/1/2019-3/31/2024 (NCE)
Statistical methods and tools for cancer risk prediction in families with germline mutations in *TP53*
1R01CA239342 ($941,486, $224,849/year)

Co-Investigator/Project co-leader
9/1/2022-8/31/2024
Deconvolution of young-onset CRC to characterize tumor-stroma-immune heterogeneity
($50k per year)
MDACC Colorectal cancer moonshot project
PI: Scott Kopetz

Sponsored Research Agreement (15%)
10/1/2021 - 9/30/2022
Curis Inc. ($78,034)
PI: Wenyi Wang

Development Research Project (15%)
6/1/2021 - 5/31/2022
Integrative genomic analysis to characterize tumor subclonal expansion and immune environment in primary and metastatic colorectal cancer ($50,000/year)
NCI SPORE in Gastrointestinal Cancer

Co-Principal Investigator
9/1/2020-8/31/2022

Intratumor Heterogeneity in Anaplastic Thyroid Carcinoma: Implications for Response to Neoadjuvant BRAF- and Immune- Directed Therapies
Mark Aspire Award ($500,000)
Mark Foundation
PI: Mark Zafereo/Jennifer Rui Wang/Wenyi Wang

Co-Principal Investigator (5%)
7/1/2019-6/30/2022
Chan-Zuckerberg Initiative Human Cell Atlas Network Fund ($92,217, $30,739/year)
PI: Rui Chen

Co-Investigator/Project leader
9/1/2021-8/31/2022
An integrated definition and therapeutic strategy for androgen indifferent prostate cancers.
($110k)
MDACC Prostate cancer moonshot project
PI: Christopher Logothetis

Co-Investigator
9/1/2019-8/31/2022
Deconvolution of CRC to characterize tumor-stroma-immune heterogeneity
($41.5k per year)
MDACC Colorectal cancer moonshot project
PI: Scott Kopetz

PI of the MDACC subcontract (17%)
8/1/2016-7/31/2021
Consistent variable selection in $p >> n$ settings.
2R01CA158113 (subcontract $320,000,$64,000/year)
NIH/NCI
PI: Valen Johnson

Principal Investigator (15%)
09/24/2014-8/31/2020
Statistical methods for genomic analysis of heterogeneous tumors
1R01CA183793 ($1,275,421, $255,084/year);
NIH/NCI

PI of the MDACC subcontract (11%)
02/01/2013-01/31/2018
Bioinformatics tools for genomic analysis of tumor and stromal pathways in cancer
1R01CA174206-01 ($1,248,881, subcontract $276,870, $55,374/year)
NIH/NCI
PI - Giovanni Parmigiani

Co-Principal Investigator
04/01/2015-03/30/2017
Cancer risk in Li Fraumeni syndrome (LFS) kindreds in regions of high founder mutation prevalence and regions of low prevalence in absence of founder as determined by LFSPRO
$100,000
MD Anderson Cancer Center Sister Institute of Network Fund

Co-PI: Louise Strong

Principal Investigator (40%)
06/01/2013-05/31/2016
Personalized risk assessment for families with Li-Fraumeni Syndrome
$464,000
Cancer Prevention Research Institute of Texas

PUBLICATIONS

underline: current trainee; *:contribute equally, not including Selected Publications.

Manuscripts under review

1. Corredor JL*, Dodd-Eaton EB*, Woodman-Ross J, Woodson A, Nguyen NH, Peng G, Green S, Gutierrez AM, Arun BK?, Wang W?. Performance of LFSPRO TP53 germline carrier risk predictions compared to standard genetic counseling practice on prospectively collected probands. under review, medRxiv:
https://www.medrxiv.org/content/10.1101/2024.07.09.24310095v1.

2. Guo S*, Liu X*, Cheng X*, Jiang Y, Ji S, Liang Q, Koval A, Li Y, Owen LA, Kim IK, Aparicio A, Weinstein JN, Kopetz S, DeAngelis MM, Chen R, **Wang W**. The DeMixSC deconvolution framework uses single-cell sequencing plus a small benchmark dataset for improved analysis of cell-type ratios in complex tissue samples. *Genome Research*, under revision. bioRxiv:
https://www.biorxiv.org/content/10.1101/2023.10.10.561733v1.

3. Duan, Y, Guo, S, Wang, W, Mueller, P. Immune Profiling among Colorectal Cancer Subtypes using Dependent Mixture Models. *JASA*, under 2nd review, bioRxiv:
https://www.biorxiv.org/content/10.1101/2023.07.24.550400v1.

4. Nguyen NH, Shin SJ, Dodd-Eaton EB, Ning J, **Wang W**✉. Personalized Risk Prediction for Cancer Survivors: A Bayesian Semi-parametric Recurrent Event Model with Competing Outcomes. under review. bioRxiv:
https://www.biorxiv.org/content/10.1101/2023.02.28.530537v2. ASA Section of Statistics Genomics and Genetics Student Paper Award 2022; ASA Section of Lifetime Data Science Student Paper Award 2023.

5. Chowdhury S, Ferri-Borgogno S, Yang P, **Wang W**, Peng J, Mok S, Wang P. Learning directed acyclic graphs for ligands and receptors based on spatially resolved transcriptomic analysis of ovarian cancer. submitted.
bioRxiv: https://www.biorxiv.org/content/10.1101/2021.08.03.454931v1.

6. Jiang Y*,Montierth MD* Yu K*, Ji S, Shin SJ, Shuai G, Cao S, Tang Y, Lesluyes T, Kopetz S, Msaouel P, Aparicio A, Shen JP, Sood AK, Wang JR, Kimmel M, Van Loo P, Zhu H, **Wang W**✉. Pan-cancer subclonal mutation analysis of 7,827 tumors predicts clinical outcome. under review
bioRxiv: https://www.biorxiv.org/content/10.1101/2024.07.03.601939v1
*authors contributed equally. ASA Section of Statistics Genomics and Genetics Student Paper Award 2021.

Published Articles

7. Yousef et al. Understanding Causes of Racial/Ethnic Survival Disparity in 47,178 Patients with Colorectal Cancer: Evaluation of Molecular, Socioeconomic, and Clinical Factors. *JAMA Oncology*, accepted

8. Aparicio AM, Tidwell RSS, Yadav SS, Chen JS, Zhang M, Liu J, Guo S, Pilie PG, Yu Y, Song X, Vundavilli H, Jindal S, Zhu K, Viscuse PV, Lebenthal JM, Hahn AW, Soundararajan R, Corn PG, Zurita AJ, Subudhi SK, Zhang J, **Wang W**, Huff C, Troncoso P, Allison JP, Sharma P, Logothetis CJ. A Modular Trial of Androgen Signaling Inhibitor Combinations Testing a Risk-Adapted Strategy in Patients with Metastatic Castration-Resistant Prostate Cancer. *Clin Cancer Res*. 2024 Apr 29;. doi: 10.1158/1078-0432.CCR-23-3740. [Epub ahead of print]

9. Bahrambeigi V, Lee JJ, Branchi V, Rajapakshe KI, Xu Z, Kui N, Henry JT, Wang K, Stephens BM, Dhebat S, Hurd MW, Sun R, Yang P, Ruppin E, **Wang W**, Kopetz S, Maitra A, Guerrero PA. Transcriptomic Profiling of Plasma Extracellular Vesicles Enables Reliable Annotation of the Cancer-specific Transcriptome and Molecular Subtype. Cancer Res. 2024 Mar 7;. doi: 10.1158/0008-5472.CAN-23-4070. [Epub ahead of print][Full text]

10. Nguyen NH, Dodd-Eaton EB, Peng G, Corredor JL, Jiao W, Woodman-Ross J, Arun BK, **Wang W**. LFSPROShiny: an interactive R/Shiny app for prediction and visualization of cancer risks in families with deleterious germline *TP53 mutations*. *JCO Clinical Cancer Informatics*. 2024 Feb;8:e2300167. doi: 10.1200/CCI.23.00167.

11. Wang JR, Zafereo ME, **Wang W**, Joshu C, Debashree R. Association of polygenic score with tumor molecular subtypes in papillary thyroid carcinoma. *Journal of Clinical Endocrinology & Metabolism*, 2023 Jul 15;dgad407. doi: 10.1210/clinem/dgad407.

12. Cermakova K, Tao L, Dejmek M, Sala M, Montierth MD, Chan YS, Patel I, Chambers C, Cabrera ML, Hoffman D, Parchem RJ, **Wang W**, Nencka R, Barbieri E, Hodges HC. Reactivation of the G1 enhancer landscape underlies core circuitry addiction to SWI/SNF. Nucleic Acids Research. gkad1081, https://doi.org/10.1093/nar/gkad1081 Published: 23 November 2023.

13. Zheng C et al. CRISPR-Cas9-based functional interrogation of unconventional translatome reveals human cancer dependency on cryptic non-canonical open reading frames. *Nature Structural & Molecular biology*, accepted

14. Rangel R et al. TP53 gain-of-function mutation modulates the immunosuppressive microenvironment in non-HPV associated oral squamous cell carcinoma. *Journal for ImmunoTherapy of Cancer*, 2023 Aug;11(8):e006666. doi: 10.1136/jitc-2023-006666.

15. Wang JR, Zafereo ME, **Wang W**, Joshu C, Debashree R. Association of polygenic score with tumor molecular subtypes in papillary thyroid carcinoma. *Journal of Clinical Endocrinology & Metabolism*, 2023 Jul 15;dgad407. doi: 10.1210/clinem/dgad407.

16. Zheng C et al. Multiomics analyses reveal DARS1-AS1/YBX1-controlled post-transcriptional circuits promoting glioblastoma tumorigenesis/radioresistance. *Science Advances*, Aug 4;9(31):eadf3984. doi: 10.1126/sciadv.adf3984. Epub 2023 Aug 4.

17. Chachad D, Patel LR, Recio CV, Pourebrahim R, Whitley EM, **Wang W**, Su X, Xu A, Lee DF, Lozano G. Unique transcriptional profiles underlie osteosarcomagenesis driven by different p53 mutants. Cancer Research, 2023 May 19:CAN-22-3464. doi: 10.1158/0008-5472.CAN-22-3464. Online ahead of print.

18. Wang JR, Montierth M, Li X, Goswami M, Zhao X, Cote G, **Wang W**, Iyer P, Dadu R, Busaidy NL, Lai SY, Grosss ND, Ferrarotto R, Lu C, Gunn GB, Williams MD, Routbort M, Zafereo ME, Cabanillas ME. Impact of Somatic Mutations on Survival Outcomes in Anaplastic Thyroid Carcinoma Patients. JCO Precision Oncology no. 6 (2022) e2100504. Published online August 17, 2022. DOI: 10.1200/PO.21.00504

19. Bondaruk J et al., **Wang W**, McConkey D, Wei P, Kimmel M, Czerniak B. The origin of bladder cancer from mucosal field effects. iScience 2022 Jun 7; 25(7):104551. doi: 10.1016/j.isci.2022.104551. eCollection 2022 Jul 15.

20. Cao S, et al., **Wang W**✉Estimation of tumor cell total mRNA expression in 15 cancer types predicts disease progression. Nature Biotechnology Published online June 13 2022. doi 10.1038/s41587-022-01342-x.

21. Wang Z, Kaseb AO, Amin HM, Hassan MM, **Wang W**, Morris JS. Bayesian edge regression in undirected graphical models to characterize interpatient heterogeneity in cancer. *JASA*. Published online: 05 Jan 2022. https://doi.org/10.1080/01621459.2021.2000866.

22. Cao L et al. Clinical Proteomic Tumor Analysis Consortium. Proteogenomic Characterization of Pancreatic Ductal Adenocarcinoma. *Cell.* 2021 Sep 16;184(19):5031-5052.e26. doi: 10.1016/j.cell.2021.08.023.

23. Taavitsainen S et al., **Wang W**, Nykter M, Urbanucci A. Single-cell ATAC and RNA sequencing reveal pre-existing and persistent subpopulations of cells associated with relapse of prostate cancer. *Nature Communications*, 2021 Sep 6;12(1):5307. doi: 10.1038/s41467-021-25624-1.

24. Dentro S, et al. Portraits of genetic intra-tumour heterogeneity and subclonal selection across cancer types. *Cell*. 2021 Apr 15; 184(8): 2239-2254.e39. *Wang is a co-last author for leading the effort in consensus subclonal reconstruction of the PCAWG WGS data.*

25. Morris JS, Hassan MM, Zohner YE, Wang Z, Xiao L, Rashid A, Abdel-Wahab R, Ballard KL, Wolff RA, George B, Li L, Allen G, Weylandt M, Li D, **Wang W**, Raghav K, Yao J, Amin HM, Kaseb AO. Deconvolution-based circulating biomarker aberration score enhances hepatocellular carcinoma staging. *Hepatology*. Published online 2020/9/15. https://doi.org/10.1002/hep.31555.

26. Haider H*, Tyekucheva S, Prandi D, Fox NS, Ahn J, Xu AW, Pantazi A, Park PJ, Laird PW, Sander C, **Wang W***, Demichelis F*, Loda M*, Boutros PC*, TCGA Research Network. Systematic assessment of tumour purity and its clinical implications. *JCO Precision Oncology*.No. 4 (2020) 995-1005. Published online September 4, 2020. https://ascopubs.org/doi/10.1200/PO.20.00016. *co-corresponding author

27. Gao F*, Pan X*, Dodd EB, Bojadzieva J, Mai PL, Vera Recio C, Zelley K, Johnson VE, Braun D, Kim NE, Garber JE, Savage SA, Strong LC, **Wang W**✉. A pedigree-based prediction model identifies carriers of deleterious de novo mutations in families with Li-Fraumeni syndrome. *Genome Research*. 2020 Aug 18. doi: 10.1101/gr.249599.119. Online ahead of print. *authors contributed equally

28. Nikooienejad A, **Wang W**, Johnson VE. Bayesian Variable Selection For Survival Data Using Inverse Moment Priors. *Annals of Applied Statistics.* Volume 14, Number 2 (2020), 809-828. arXiv version: https://arxiv.org/abs/1712.02964.

29. McCarthy DJ, Rostom R, Huang Y, Kunz DJ, Danecek P, Bonder MJ, Hagai T, HipSci Consortium, **Wang W**, Gaffney DJ, Simons BD, Stegle O, Teichmann SA. Cardelino: computational integration of somatic clonal substructure and single-cell transcriptomes. *Nature Methods*. Apr;17(4):414-421. doi: 10.1038/s41592-020-0766-3. Epub 2020 Mar 16.

30. Pan-cancer analysis of whole genomes. The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. *Nature*. 2020 Feb 5;578:82–93.

31. Wu CC, Beird H, Livingston JL, Advani S, Mitra A, Cao S, Reuben A, Ingram D, Wang WL, Ju Z, Leung CH, Lin H, Zheng Y, Roszik J, **Wang W**, Patel S, Benjamin R, Somaiah N, Conley A, Gordon M, Hwu P, Gorlick R, Lazar A, Daw N, Lewis V, and Futreal PA. Immuno-genomic landscape of

osteosarcoma. *Nature Communications*. 2020 Feb 21;11(1):1008. doi: 10.1038/s41467-020-14646-w.

32. Salcedo A, et al. A community effort to create standards for evaluating tumor subclonal reconstruction. *Nature Biotechnology*. 2020 Jan;38(1):97-107. doi: 10.1038/s41587-019-0364-z. Epub 2020 Jan 9.

33. Shin SJ, Dodd-Eaton EB, Peng G, Bojadzieva J,Chen J, Amos CI, Frone MN, Khincha PP, Mai PL, Savage SA, Ballinger ML, Thomas DM, Yuan Y, Strong LC, **Wang W**✉. Penetrance of different cancer types in families with Li-Fraumeni syndrome: a validation study using multicenter cohorts. *Cancer Research*. 2020 Jan 15;80(2):354-360. doi: 10.1158/0008-5472.CAN-19-0728. Epub 2019 Nov 12.

34. Shin SJ, Dodd EB, Gao F, Bojadzieva J, Chen J, Kong X, Amos CI, Ning J, Strong LC, **Wang W**✉. Penetrance estimates over time to first and second primary cancer diagnosis in families with Li-Fraumeni syndrome: a single institution perspective. *Cancer Research*. 2020 Jan 15;80(2):347-353. doi: 10.1158/0008-5472.CAN-19-0725. Epub 2019 Nov 12.

35. Gerstung M, et al., PCAWG Evolutionary and Heterogeneity Working Group-PCAWG network. The evolutionary history of 2,658 cancers. *Nature*. 2020 Feb 6;578:122–128.

36. Maura F, Agnelli L, Leongamornlert D, Bolli N, Chan J, Dodero A, Carniti C, Heavican T, Pellegrinelli A, Pruneri G, Butler A, Bhosle S, Chiappella A, Rocco AD, Zinzani PL, Zaja F, Piva R, Inghirami G, **Wang W**, Palomero T, Iqbal J, Neri A, Campbell PJ, Corradini P. Integration of transcriptional and mutational data improves the stratification of peripheral T-Cell lymphoma. *American Journal of Hematology*. March 4 2019 doi: 10.1002/ajh.25450.

37. Tarabichi M, Martincorena I, Gerstung M, Leroi AM, Markowetz F; PCAWG Evolution and Heterogeneity Working Group, Spellman PT, Morris QD, Lingjaerde OC, Wedge DC, Van Loo P. Neutral tumor evolution? *Nature Genetics*. 2018 Dec;50(12):1630-1633. doi: 10.1038/s41588-018-0258-x.

38. Wang Z, Morris JS, Cao S, Ahn J, Liu R, Tyekucheva S, Li B, Lu W, Tang X, Wistuba II, Bowden M, Mucci L, Loda M, Parmigiani G, Holmes CC, **Wang W**✉. Transcriptome deconvolution of heterogeneous tumor samples with immune infiltration. *iScience*. 2018 Nov 30;9:451-460. doi: 10.1016/j.isci.2018.10.028.

39. Shin SJ, Li J, Ning J, Bojadzieva J, Strong LC, **Wang W**✉. Bayesian estimation of a semiparametric recurrent event model with applications to the penetrance estimation of multiple primary cancers in Li-Fraumeni syndrome. *Biostatistics*. 2018 kxy066, https://doi.org/10.1093/biostatistics/kxy066.

40. Shin SJ, Ying Yuan✉, Strong LC, Bojadzieva J, **Wang W**✉. Bayesian semiparametric estimation of cancer-specific age-at-onset penetrance with application to Li-Fraumeni syndrome. *JASA*. 2018 doi: 10.1080/01621459.2018.1482749.

41. Li J, Fu C, Speed TP, **Wang W**✉, Symmans F✉. Accurate RNA sequencing from Formalin-fixed cancer tissue to represent high-quality transcriptome from frozen tissue. *Journal of Clinical Oncoloty Precision Oncology*. Published online January 26, 2018. doi: 10.1200/PO.17.00091. it Among the most accessed articles in 2018 at JCO PO.

42. Ahn J, Morita S, **Wang W**✉and Yuan Y✉. Bayesian shared-parameter models for longitudinal dyadic data with informative missing data. *Statistics Methods in Medical Research.* 2017 Jan 1:962280217715051. doi: 10.1177/0962280217715051.

43. Peng G, Bojadzieva J, Ballinger ML, Li J, Savage S, Mai P, Thomas DM, Strong LC, **Wang W**✉. Estimating *TP53* mutation carrier probability in families with Li-Fraumeni Syndrome using LFSPRO. *Cancer Epidemiology, Biomarker and Prevention.* Jan 2017. DOI: 10.1158/1055-9965.EPI-16-0695.

44. Holik AZ, Law CW, Liu R, Wang Z, **Wang W**, Ahn J, Asselin-Labat M, Smyth GK and Ritchie ME. RNA-seq mixology: designing realistic control experiments to compare protocols and analysis methods. *Nucleic Acids Research.* 2016 Nov 29. pii: gkw1063.

45. Fan Y, Xi L, Hughes DST, Zhang J, Zhang J, Futreal AP, Wheeler DA, **Wang W**✉. Accounting for tumor heterogeneity using a sample-specific error model improves sensitivity and specificity in mutation calling for sequencing data. *Genome Biology.* 2016 Aug 24;17(1):178. doi: 10.1186/s13059-016-1029-6. *MuSE was used to generate consensus mutation calls on ICGC ∼2,700 WGS data and TCGA ∼11,000 WES data.*

46. Nikooienejad A, **Wang W**✉, Johnson VE✉. Bayesian variable selection for binary outcomes in high dimensional genomic studies using non-local priors. *Bioinformatics.* 2016 May 1;32(9):1338-45. doi: 10.1093/bioinformatics/btv764. Epub 2016 Jan 6.

47. Palculict TB, Ruteshouser EC, Fan Y, **Wang W**, Strong L, Huff V. Identification of germline DICER1 mutations and loss of heterozygosity in familial Wilms tumor using whole genome sequencing. *Journal of Medical Genetics* Nov 13. pii: jmedgenet-2015-103311. doi: 10.1136/jmedgenet-2015-103311.

48. Lefterova MI*, Shen P*, Odegaard JI*, Fung E, Chiang T, Peng G, Davis RW, **Wang W**, Schrijver I, Scharfe C. Next-generation molecular testing of newborn dried blood spots for cystic fibrosis. *Journal of Molecular Diagnostics.* 2016 Mar;18(2):267-82. doi: 10.1016/j.jmoldx.2015.11.005. Epub 2016 Feb 1.

49. The Cancer Genome Atlas Research Network. The molecular taxonomy of primary prostate cancer. *Cell* 2015 Nov 5;163(4):1011-25. doi: 10.1016/j.cell.2015.10.025.

50. Fang LT, Afshar PT, Chhibber A,Mohiyuddin M, Fan Y, Mu J, Gibeling G, Barr S, Asadi NB, Gerstein M, Koboldt D, **Wang W**, Wong WH, Lam H. An ensemble approach to accurately detect somatic mutations using SomaticSeq. *Genome Biology*. 2015 Sep 17;16:197. doi: 10.1186/s13059-015-0758-2.

51. Ewing AD, Houlahan KE, Hu Y, Ellrott K, Caloian C, Yamaguchi TN, Bare JC, P'ng C, Waggott D, Sabelnykova VY; ICGC-TCGA DREAM Somatic Mutation Calling Challenge participants, Kellen MR, Norman TC, Haussler D, Friend SH, Stolovitzky G, Margolin AA, Stuart JM, Boutros PC. Combining tumor genome simulation with crowdsourcing to benchmark somatic single-nucleotide-variant detection. *Nat Methods.* 2015 Jul;12(7):623-30. doi: 10.1038/nmeth.3407. Epub 2015 May 18.

52. Peng G, Fan Y, **Wang W**✉. FamSeq: a variant calling program for family-based sequencing data using graphics processing units. *PLOS Computational Biology* 2014 Oct 30;10(10):e1003880. doi: 10.1371/journal.pcbi.1003880.

53. Davis CF, et al.; Cancer Genome Atlas Research Network, Chin L, Meyerson M, Kucherlapati R, Park WY, Robertson AG, Laird PW, Henske EP, Kwiatkowski DJ, Park PJ, Morgan M, Shuch B, Muzny D, Wheeler DA, Linehan WM, Gibbs RA, Rathmell WK, Creighton CJ. The somatic genomic landscape of chromophobe renal cell carcinoma. *Cancer Cell.* 2014 Sep 8;26(3):319-30. doi: 10.1016/j.ccr.2014.07.014. Epub 2014 Aug 21. PMID: 25155756

54. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* 507(7492):315-22, 3/2014. e-Pub 1/2014.

55. Ahn J, Liu S, **Wang W**✉, Yuan Y✉. Bayesian latent-class mixed-effect hybrid models for dyadic longitudinal data with non-ignorable dropouts. *Biometrics* 2013 Dec;69(4):914-24. doi: 10.1111/biom.12100. Epub 2013 Nov 6

56. The Cancer Genome Atlas Research Network. The Cancer Genome Atlas Pan-Cancer analysis project. *Nature Genetics* 2013 Oct;45(10):1113-20. doi: 10.1038/ng.2764.

57. Srivastava S, **Wang W**, Zinny PO, Colen RR, Baladandayuthapani V. Integrating multi-platform genomic data using hierarchical bayesian relevance vector machines. *EURASIP Journal on Bioinformatics and Systems Biology* 2013 Jun 28;2013(1):9. doi: 10.1186/1687-4153-2013-9.

58. Shen P*, **Wang W***, Chi AK, Fan Y, Davis RW, Scharfe C. Target capture using double-stranded DNA probes. *Genome Medicine* 2013, 5:50 doi:10.1186/gm454. *authors contributed equallly

59. Ahn J, Yuan Y, Parmigiani G, Suraokar MB, Diao L, Wistuba II, and **Wang W**✉. DeMix: deconvolution for mixed cancer transcriptomes. *Bioinformatics* 2013 doi: 10.1093/bioinformatics/btt301.

60. Peng G, Fan Y, Palculict TB, Shen P, Ruteshouser EC, Chi A, Davis RW✉, Huff V, Scharfe C, **Wang W**✉. Rare variant detection using family-based sequencing analysis. *Proceedings of the National Academy of Sciences.* ePub, February 20, 2013, doi: 10.1073/pnas.1222158110.

61. Zhang N, Xu Y, O'Hely M, Speed TP, Scharfe C, **Wang W**✉. SRMA: an R package for sequence based calling in candidate genes with custom resequencing microarrays. *Bioinformatics.* e-Pub 05/2012.

62. Hua Y, Gorshkov K, Yang Y, **Wang W**, Zhang N, Hughes DPM. Slow down to stay live: HER4 protects against cellular stress and confers chemoresistance in neuroblastoma. *Cancer.* e-Pub 03/2012.

63. Rubio JP, Wilkins EJ, Kostchet K, Cowie TC, O'Hely M, Burfoot R, **Wang W**, Speed TP, Stankovich J, Horne M. A DNA resequencing array for genes involved in Parkinson's Disease. *Parkinsonism & Related Disorders.* e-Pub 01/2012.

64. Shen P*, **Wang W***, Krishnakumar S, Palm C, Chi AK, Enns GM, Davis RW, Speed TP, Mindrinos MN, Scharfe C. High-quality DNA sequence capture of 524 disease candidate genes. *Proceedings of the National Academy of Sciences.* 2011, Apr 19;108(16):6549-54. Epub 2011 Apr 5.

    *authors contributed equally

65. Lin S*, **Wang W***, Palm C, Davis RW, Juneau K. A Molecular inversion probe assay for detecting alternative splicing. *BMC Genomics* Dec 2010, 11:712.

    *authors contributed equally

66. **Wang W**✉, Shen P, Thyagarajan S, Lin S, Palm C, Horvath R, Klopstock T, Cutler D, Pique L, Schrijver I, Davis RW, Mindrinos M, Speed TP, Scharfe C✉. Identification of rare DNA variants in mitochondrial disorders with improved array-based sequencing. *Nucleic Acids Research* 2011 Jan;39(1):44-58. Epub 2010 Sep 15.

67. **Wang W**, Niendorf KB, Patel D, Blackford A, Marroni F, Sober AJ, Parmigiani G and Tsao H. Estimating *CDKN2A* Carrier probability and personalizing cancer risk assessments in hereditary melanoma using MelaPRO. *Cancer Research.* Jan 2010 doi:10.1158/0008-5472.

68. **Wang W**, Carvalho B, Miller N, Pevsner J, Chakaravarti A and Irizarry RA. Estimating genome-wide copy number using allele specific mixture models. *Journal of Computational Biology.* 2008, 15:857-66.

69. **Wang W**, Chen S, Brune KA, Hruban RH, Parmigiani G and Klein AP. Development and validation of a risk assessment tool for individuals with a family history of pancreatic cancer: PancPRO. *Journal of Clinical Oncology*. 2007;25:1417-22.

70. Nicodemus KK, **Wang W** and Shugart YY. Stability of variable importance scores and rankings using statistical learning tools on single nucleotide polymorphisms (SNPs) and risk factors involved in gene-gene and gene-environment interactions. *BMC Proceedings*. 2007;1 Suppl 1:S58.

71. Chen S, **Wang W**, Lee S, Nafa K, Lee J, Romans K, Watson P, Gruber SB, Euhus D, Kinzler KW, Jass J, Gallinger S, Lindor N, Casey G, Ellis N, Giardiello FM, the Colon Cancer Family Registry, Offit K, Parmigiani G. Prediction of germline mutations and cancer risk in the Lynch syndrome. *Journal of the American Medical Association* 2006;296:1479-87.

72. Gonzalez JR, **Wang W**, Ballana E and Estivill X. A recessive Mendelian model to predict carrier probabilities of DFNB1 for non-syndromic deafness. *Human Mutation* 2006;27:1135-42.

73. Chen S, **Wang W**, Broman K and Parmigiani G. BayesMendel: An R environment for Mendelian risk prediction. *Statistical Application in Genetics and Molecular Biology* 2004;3: Article 21.

Book Chapters

74. Ji S, Montierth M, **Wang W**. MuSE: A Novel Approach to Mutation Calling with Sample-Specific Error Modeling. Methods Molecular Biology. 2022; 2493:21-27.

75. **Wang W**, Fan Y, Speed TP. DNA variant calling in targeted sequencing data. In *Advances in statistical bioinformatics: Models and Integrative Inference for High-Throughput Data* (Editors: Do KA, Qin ZS, Vannuci M). Cambridge University Press. June 2013

PRESENTATIONS

Invited Talks

1. Statistical methods development for cancer risk prediction. UCLA Computational Genetics Summer Institute, Los Angeles 07/24/2023.

2. Estimation of tumor cell total mRNA expression through deconvolution. BIRS Mathematical Methods in Cancer biology, Evolution and Therapy. Banff, Canada 05/15/2023

3. Deciphering cancer cell evolution and ecology and their associations with cancer prognosis. AACR. Orlando, FL 4/17/2023

4. Characterizing tumor microenvironment and clonal expansion at single cell resolution. ENAR. Nashville, TN 03/21/2023

5. Benchmarking-related model development for deconvoluting cancer genomes and heterogeneous tissue transcriptomes. NCI Spring School on Algorithms in Cancer Biology. Bethesda, MD 03/13/2023

6. An integrated genomic definition and therapeutic strategy for heterogeneous tumors. Center for Applied Bioinformatics. St Jude Children's Research Hospital. Memphis, TN 03/02/2023

7. Deciphering cancer cell evolution and ecology. NCI Biostatistics Branch, Division of Cancer Epidemiology and Genetics. (virtual) 02/23/2023

8. Estimation of tumor cell total mRNA expression in 15 cancer types predicts disease progression. The 4th Annual Meeting of China Chapter of the International MAQC Society (virtual). 12/17/2022.

9. Cancer Risk Prediction Modeling and Deciphering Cancer Evolution and Ecology. Keynote speaker. The UTGSBS Quantitative Sciences Program Student Symposium. Houston, TX 11/17/2022.

10. Cancer risk modeling for deleterious mutations in TP53 using a multi-center consortium. 6th International LFS Association Symposium. Bethesda, MD 10/15/2022.

11. Computational deconvolution of cancer genomes and transcriptomes. Computational Genomics Summer Institute UCLA. Los Angeles, CA 07/11/2022.

12. Deciphering cancer cell evolution and ecology. Keynote Speaker. RECOMB 2022. San Diego, CA 05/25/2022.

13. Risk Prediction Models of Li-Fraumeni Syndrome for Genetic Counseling. NCI Cancer Genetics Branch Seminar (virtual) 05/03/2022.

14. Cancer risk modeling for deleterious mutations in TP53 using a multi-center consortium. Statistics Colloqium Rice University. Houston TX 01/24/2022.

15. Tumor Cell Total mRNA Expression Shapes the Molecular and Clinical Phenotype of Cancer. Enjoy Science Webinar Series UT MD Anderson Cancer Center. 01/06/2022.

16. Tumor Cell Total mRNA Expression Shapes the Molecular and Clinical Phenotype of Cancer. NCI DCEG Biostatistics Branch. 12/08/2021.

17. Statistical methods for genomic analysis of heterogeneous tumor samples. ASA Section on Statistics in Genomics and Genetics Webinar. 09/20/2021.

18. Cancer risk modeling for deleterious mutations in *TP53* using a multi-center consortium. Genetics of Hereditary Cancer Symposium Honoring Louise C. Strong, M.D. Houston TX. 06/22/2021.

19. Tumor cell total mRNA expression shapes the molecular and clinical phenotype of cancer. UCSD Department of Bioengineering (virtual). San Diego CA. 05/07/2021.

20. Tumor cell total mRNA expression shapes the molecular and clinical phenotype of cancer. Endocrine Society ENDO 2021 (virtual). 03/22/2021.

21. Tumor cell total mRNA expression shapes the molecular and clinical phenotype of cancer. Translational Colorectal Cancer: From Genomics to Therapy Symposium (virtual). Houston TX. 03/04/2021.

22. Tumor cell total mRNA expression shapes the molecular and clinical phenotype of cancer. Special Topics in Bioinformatics Seminar (virtual). UT Southwestern Medical Center. Dallas TX. 02/08/2021.

23. Tumor cell total mRNA expression shapes the molecular and clinical phenotype of cancer. ICSA 2020 Applied Statistics Symposium (virtual). Houston TX, 12/13/2020.

24. Tumor cell total mRNA expression shapes the molecular and clinical phenotype of cancer. Victorian Cancer Bioinformatics Symposium (virtual). Melbourne, Australia 10/27/2020.

25. Tumor cell total mRNA expression shapes the molecular and clinical phenotype of cancer. MD Anderson Cancer Center Symposium on Cancer Research: Leading Edge in Cancer Research (virtual). Houston TX, 10/22/2020.

26. Clinical Impact of de novo mutations in TP53 as illustrated in families with Li-Fraumeni syndrome. 5th international LFS Association Symposium (virtual). Boston MA, 10/03/2020.

27. Understanding Tumor Transcriptional Activity, Heterogeneity and Evolution using Deconvolution Models. Department of Genetics and Genomics Series (virtual). Icahn School of Medicine at Mount Sinai. 03/16/2020.

28. Global tumor transcriptional activity reveals aggressiveness across multiple cancers. Department of Biostatistics, Fred Hutchinson Cancer Research Center, Seattle Washington, 12/12/2019.

29. Global tumor transcriptional activity reveals aggressiveness across multiple cancers. Department of Human Genetics UCLA, Los Angeles CA, 12/2/2019.

30. Global tumor transcriptional activity reveals aggressiveness across multiple cancers. Wellcome Centre for Human Genetics, Oxford UK, 11/20/2019

31. Global tumor transcriptional activity reveals aggressiveness across multiple cancers. Department of Biostatistics, Vanderbilt University. Nashville TN, 10/23/2019.

32. Robust subclonal architecture reconstruction from $\sim$2,700 cancer genomes. Research Frontiers in Biomathematics, UCLA, Los Angeles CA, 10/3/2019.

33. Transcriptomic deconvolution for tumor-stroma-immune interactions across cancer (sub)types. Cancer System Genetics Workshop. Berlin, Germany. 06/14/2019.

34. Mutation-based expression deconvolution identifies differential transcriptional activity across cancer (sub)types. Towards In silico-Guided Clinical Trial in Cancer Workshop. Oslo, Norway. 05/15/2019

35. Ares: allele read specific expression estimation for somatic mutations in a pancancer study. Single Cell and Massively Parallel Approaches Workshop. Bridgetown, Barbados. 01/26/2019

36. Statistical methods for the deconvolution of transcriptomes and genomes from heterogeneous tumor samples. UT Health School of Public Health Human Genetics Center. Houston, TX 01/16/2019

37. Robust subclonal architecture reconstruction from $\sim$ 2,700 cancer genomes. University of Houston Department of Mathematics. Houston, TX 11/30/2018

38. Deconvolution of multi-omics Data from Heterogeneous Tumor Samples. TAMU Bioinformatics & Cancer Symposium, College Station, TX 09/21/2018

39. A study on de novo mutations in TP53 using families with Li-Fraumeni syndrome. (Chalk talk) Systems Genetics Cancer Workshop. Portland, OR 06/15/2018

40. Cancer risk assessment and penetrance estimation for deleterious mutations in TP53 using a multi-centre consortium. The 4th International Li-Fraumeni Syndrome Conference. 04/26/2018

41. Robust subclonal architecture reconstruction from $\sim$ 2,700 cancer genomes. ENAR Spring Meeting, Atlanta, Georgia. 03/27/2018

42. Statistical methods for the deconvolution of high-throughput sequencing data from heterogeneous tumor samples. National Cancer Institute Division of Cancer Epidemiology and Genetics. Bethesda, MD 03/01/2018

43. Statistical methods for the deconvolution of high-throughput sequencing data from heterogeneous tumor samples. University of Warwick Department of Statistics. Warwick, UK 02/12/2018

44. Statistical methods for the deconvolution of high-throughput sequencing data from heterogeneous tumor samples. Barts Cancer Institute, London, UK 02/08/2018

45. Statistical methods for the deconvolution of mixed cancer transcriptomes. Baylor College of Medicine Breast Disease Working Group. Houston, TX 01/10/2018

46. Robust subclonal architecture reconstruction from ∼2,700 cancer genomes. Fudan University Shanghai Cancer Center Department of Pathology. Shanghai, China 12/22/2017

47. Statistical methods for the deconvolution of mixed cancer transcriptomes. Center for Functional Cancer Epigenetics at the Dana Farber Cancer Institute. Boston, MA 12/08/2017

48. Robust subclonal architecture reconstruction from ∼2,700 cancer genomes. CMO-BIRS: Challenges and Synergies in the Analysis of Large-Scale Population-Based Biomedical Data. Oaxaca, Mexico 11/28/2017

49. Robust subclonal architecture reconstruction from ∼2,700 cancer genomes. Department of Bioinformatics and Computational Biology. University of Southern California. Los Angelos, CA 11/09/2017

50. Statistical inference problems for the gene expression deconvolution of heterogeneous tumor samples. Systems Genetics of Cancer, London UK 09/18-09/20/2017.

51. Statistical methods for the gene expression deconvolution of heterogeneous tumor sample. The European Molecular Biology Laboratory (EMBL), Heidelberg Germany 07/27/2017.

52. Statistical methods for the gene expression deconvolution of heterogeneous tumor sample. Cancer Research UK Cambridge Institute, Cambridge UK 07/24/2017.

53. Statistical methods for the gene expression deconvolution of heterogeneous tumor sample. Oxford University Big Data Institute, Oxford UK 07/21/2017.

54. Statistical methods for the gene expression deconvolution of heterogeneous tumor sample. UNC Chapel Hill Department of Biostatistics, Chapel Hill, NC 04/27/2017.

55. Statistical methods for the gene expression deconvolution of heterogeneous tumor sample. Rice University Department of Statistics, Houston TX 04/17/2017.

56. Statistical methods for the gene expression deconvolution of heterogeneous tumor sample. BIRS workshop: Statistical and Computational Challenges in Large Scale Molecular Biology, Banff, Canada 03/28/2017.

57. Accounting for Tumor Heterogeneity Using a Sample-Specific Error Model Improves Sensitivity and Specificity in Mutation Calling for Sequencing Data, University of Iowa Department of Pulmonary Medicine, Cedar Rapids, IA 03/24/2017.

58. Statistical methods for the gene expression deconvolution of heterogeneous tumor sample. The University of Texas School of Public Health Department of Biostatistics, 02/14/2017.

59. Accounting for Tumor Heterogeneity Using a Sample-Specific Error Model Improves Sensitivity and Specificity in Mutation Calling for Sequencing Data, University of Evry, LaMME, Evry, France, 12/14/2016.

60. Accounting for Tumor Heterogeneity Using a Sample-Specific Error Model Improves Sensitivity and Specificity in Mutation Calling for Sequencing Data, UCLA, Program in Bioinformatics, 11/7/2016.

61. Accounting for Tumor Heterogeneity Using a Sample-Specific Error Model Improves Sensitivity and Specificity in Mutation Calling for Sequencing Data, Rice University, Keck Seminar, 10/14/2016.

62. Accounting for Tumor Heterogeneity Using a Sample-Specific Error Model Improves Sensitivity and Specificity in Mutation Calling for Sequencing Data, Ontario Institute of Cancer Research, 10/7/2016.

63. Accounting for Tumor Heterogeneity Using a Sample-Specific Error Model Improves Sensitivity and Specificity in Mutation Calling for Sequencing Data, Memorial Sloan Kettering Cancer Center, 10/5/2016.

64. Cancer-specific characterization of Li-Fraumeni Syndrome. The 2016 International LFS conference & the 3rd annual LiFE consortium and LFS association conference. Columbus, Ohio, 06/03/2016.

65. Accounting for tumor heterogeneity using a sample-specific error model improves sensitivity and specificity in mutation calling for sequencing data. KU Leuven Center for Human Genetics, Leuven, Belgium, 04/14/2016

66. Gene expression deconvolution of heterogeneous tumor samples: DeMix-Bayes. The Francis Crick Institute. London, UK, 12/15/2015

67. Bayesian variable selection for binary outcomes in high dimensional settings. 8th International Conference of the ERCIM WG on Computational and Methodological Statistics. London, UK, 12/12/2015

68. Gene expression deconvolution in heterogeneous tumor samples using DeMix-Bayes. NCI DCCPS New Grantee Workshop, 11/5/2015

69. Bayesian variable selection for binary outcomes in high dimensional settings. Joint Statistics Meeting. Seattle, WA, 8/9/2015

70. Cancer-specific characterization of the Li-Fraumeni syndrome. A.C. Carmargo Cancer Center, Sau Paulo, Brazil, 7/28/2015

71. Statistical methods for analysis of genomic data from heterogeneous cancer samples. BioC 2015, Seattle, WA, 7/21/2015

72. Statistical methods for analysis of genomic data from heterogeneous cancer samples. International Bioinformatics Workshop. Harbin, China, 7/10/2015

73. Statistical methods for analysis of genomic data from heterogeneous cancer samples. University of Evry, Essonne, France, 5/19/2015

74. Statistical methods for analysis of genomic data from heterogeneous cancer samples, Johns Hopkins University, Department of Biostatistics, Baltimore, MD, 4/6/2015

75. LFSpro: a risk assessment tool to estimate TP53 mutation status in families with Li-Fraumeni Syndrome, The University of Texas School of Public Health, Human Genetics Center, Houston, TX, 2/9/2015

76. DeMix-Bayes: a Bayesian model for the deconvolution of mixed cancer transcriptomes in microarray and RNA sequencing data, Baylor College of Medicine, Computational and Integrative Biomedical Research Center, Houston, TX, 1/14/2015

77. Cancer-specific characterization of the Li-Fraumeni Syndrome, Worcester Polytechnic Institute, Biomedical Engineering, Worcester, MA, 11/6/2014

78. DeMix-Bayes: a Bayesian model for the deconvolution of mixed cancer transcriptomes in microarray and RNA sequencing data, *TCGA Face-to-face network meeting*, Bethesda, MD, 2014.10.28

79. MuSE: somatic evolution estimation for mutation calling in sequencing data of matched tumor-normal samples, The Joint Statistical Meetings, *The Joint Statistical Meetings*, Boston, MA, 2014.8.4

80. Bayesian variable selection for binary outcomes in high dimensional settings, *The International Society of Bayesian Analysis World Meeting*, Cancun, Mexico, 2014.7.17

81. Somatic Mutation Calling and Gene Expression Deconvolution in Heterogeneous Tumor Samples, University of Oxford, Department of Oncology, Oxford, United Kingdom, 2014.6.12

82. Discussant for Keynote Speaker: Two Aspects in Tumor Heterogeneity: Subclonal Mutations and Stromal Expression, *The Southern Regional Council on Statistics Summer Research Conference*, Galveston, TX, 2014.6.3

83. Gene expression deconvolution in heterogeneous tumor samples, Texas A&M Unversity, Department of Statistics, College Station, TX, 2014.2.21

84. Gene expression deconvolution in heterogeneous tumor samples, *7th Annual Bayesian Biostatistics and Bioinformatics Conference*, Houston, TX, 2014.2.13

85. Gene expression deconvolution in heterogeneous tumor samples. *The 9th ICSA International Conference*, HongKong, China, 2013.12.23

86. Personalized risk assessment for families with Li-Fraumeni Syndromes. *The LiFE Consortium Meeting*. Boston, MA, 2013.10.26

87. Gene expression deconvolution in heterogeneous tumor samples. *The Joint Statistics Meetings*. Montreal, Canada, 2013.8.3

88. Rare variant detection using family-based sequencing analysis. Washington University Department of Genetics. St. Louis, MO, 2012.11

89. Determining probability of rare variants: implications for designs of family-based sequencing studies. *The Joint Statistical Meetings*. San Diego, CA, 2012.08

90. Determining probability of rare variants in sequencing studies for familial cancer syndromes. *The International Workshop on Cancer Systems Biology*. Changchun, China, 2012.07

91. Determining probability of rare variants: implications for designs of family-based sequencing studies. *The Southern Regional Council on Statistics: Summer Research Conference*. Jekyll Island, GA, 2012.06

92. Determining probability of rare variants: implications for designs of family-based sequencing studies. *International Conference on Risk Assessment and Evaluation of Predictions*. Silver Spring, MD, 2011.10

93. Determining probability of rare variants: implications for designs of family-based sequencing studies. UT School of Public Health Human Genetics Center, 2011.9

94. Determining probability of germline mutations in family-based sequencing studies. The First Wuxi International Statistics Forum, Wuxi, China, 2011.7

95. Validating risk prediction models using family registries, *Fourth Annual Bayesian Biostatistics Conference*, Houston, TX, 2011.1

96. Statistical methods for DNA resequencing analysis in disease-gene studies, Rice University, Department of Statistics, Houston, TX, 2011.1

SOFTWARE

1. `BayesMendel` (co-author): a comprehensive environment for prediction of inherited cancer susceptibility

    - R package, http://bcb.dfci.harvard.edu/BayesMendel

2. `SRMA`: Sequence robust multi-array analysis for resequencing arrays

- R package, http://odin.mdacc.tmc.edu/ wwang7/SRMAIndex.html

3. `FamSeq`: Analysis of family-based sequencing data
   - C++ based, GPU based http://bioinformatics.mdanderson.org/main/FamSeq
4. `LFSPRO`: Personalized risk assessment for families with Li-Fraumeni syndromes
   - R package, http://bioinformatics.mdanderson.org/main/LFSPRO
   - R function in BayesMendel, http://bcb.dfci.harvard.edu/BayesMendel
   - Update in 2019: cancer specific risk prediction, multiple primary cancer specific risk prediction for families with Li-Fraumeni Syndrome.
5. `DeMixT`: Deconvolution of mixed transcriptomes from tumor samples
   - R package, http://github.com/wwylab/DeMixT.
   - Biocondcutor, https://bioconductor.org/packages/release/bioc/html/DeMixT.html.
6. `MuSE`: Mutation somatic evolution estimation for sequencing data from matched tumor-normal pairs.
   - C++, http://bioinformatics.mdanderson.org/main/MuSE
   - Docker container, available from ICGC and TCGA PanCanAtlas
   - MuSE2.0 is now avialable and 50x faster than 1.0
7. `Famdenovo`: calculates the probability of a germline mutation to be de novo based on family history data.
   - R package, http://github.com/wwylab/Famdenovo.
8. `CliPP`: Clonal identification of subclonal structure through pairwise penalization using sequencing data from tumor samples. (C code)
   - Code, http://github.com/wwylab/CliPP.
9. `CSR`: Consensus Clustering for Subclonal structure Reconstruction.
   - Code, http://github.com/wwylab/CSR.
10. `DeMixSC`: Single-cell based deconvolution of bulk RNA sequencing data from heterogeneous tissue samples.
    - Code, http://github.com/wwylab/DeMixSC.

TEACHING EXPERIENCE

The UNIVERSITY OF TEXAS MD ANDERSON CANCER CENTER
POSTDOC

2019.10-present Shuangxi Ji (PhD in Biological Sciences, University of Birmingham, UK)
2022.9-2024.6 Xiaoqian Liu(PhD in Statistics, North Carolina State University),
Current position: Tenure-track assistant professor at UC Riverside Statistcs
2021.6-2022.9 Vankata Vundavilli (PhD in Electrical Engineering, Texas A&M)
New position:Assistant Professor, Indian Institute of Technology.
2016.9-2022.1 Shaolong Cao (PhD in Biomedical Engineering, University of Tulane)

New position: Senior data scientist, Biogen.
2016.11-2018.5 Kaixian Yu (joint with Hongtu Zhu, PhD in Statistics, Florida State University)
Current position: Senior data scientist, Didi Corp.
2015.8-2017.7 Xuedong Pan (joint with Val Johnson, PhD in Genetics, Fudan University)
Current position: Senior Bioinformatician.
2012.4-2016.8 Yu Fan (PhD in Evolution and Ecology, University of Connecticut)
Current position: Software engineer, ExxonMobil
2013.7-2015.2 Seung Jun Shin (joint with Ying Yuan, PhD in Statistics,
North Carolina State University)
Current position: Tenured Associate Professor in Statistics,
Korea University, Seoul, South Korea
2011.9-2013.8 Jaeil Ahn (joint with Ying Yuan, PhD in Biostatistics, University of Michigan)
Current position: Tenured Associate Professor in Biostatistics and Bioinformatics,
Georgetown University, Washington DC

PHD STUDENT
2023.9-present Hao Yan (UTSPH Biostatistics)
2022.9-present Haoming Shi (Rice Statistics)
2021.3-present Yaoyi Dai (BCM Quantitative and Computional biology)
2020.1-present Shuai Guo (UTGSBS Quantitative Sciences)
2020.1-present Matthew Montierth (BCM Quantitative and Computional biology)
2020.6-2024.4 Hoai Nam Nguyen (Rice Statistics)
currently Data Scientist at Trusting Social)
2020.1-2024.4 Yujie Jiang (Rice Statistics),
currently Senior Statistician at AbbVie)
2017.9-2021.5 Carlos Vera Recio (GSBS MD-PhD Biostatistics, NLM fellow, graduated with an MS)
2013.6-2017.4 Zeya Wang (Rice Statistics, currently Tenure-track Assistant Professor at
University of Kentucky Statistics)
2013.5-2017.4 Jialu Li (GSBS Biostatistics, currently Biostatistician at Novatis)
2011.1-2015.8 Gang Peng (GSBS Biostatistics, AAAS/Science Program for Excellence in Science,
currently Tenure-track Assistant Professor
of Medical and Molecular Genetics, Indiana University)

VISITING/GUEST CLINICAL FELLOW
2021.9-2022.9 Alexander Martin (Hematology fellow, currently Assistant Professor, Tufts Medical Center)
2019.5-present Jennifer Rui Wang (MDACC Head&Neck Surgery, Assistant Professor, PhD in Epidemiology, Johns Hopkins University)

External thesis examiner of
2019 Stefano Mangliola (University of Melbourne)
2019 Moustafa Abdalla (University of Oxford)

Computational Genetics Summer Institute at UCLA
2024.7 Instructor (300 students)
2023.7 Instructor (100 students)
2022.7 Instructor (300 students)

Otto Warburg International Summer School and Research Symposium
2019.8 Instructor (among 10 speakers), Cell-Type Heterogeneity and Single-Cell Analysis, Shanghai

China

TEXAS A&M UNIVERSITY DEPARTMENT OF STATISTICS
2017.1-5 Instructor, Statistical Bioinformatics (STAT646): 51 graduate students

THE GRADUATE SCHOOL OF BIOLOGICAL SCIENCES AT HOUSTON
2018.1 - 2020 Instructor, Computational Approaches for Single-Cell Data Analysis GS01 1041
2013 - 2021 Lecturer, Introduction to Bioinformatics GS0011062
2021 - present Faculty moderator for Quantitative Sciences, Scientific writing course GS211152-523

RICE UNIVERSITY, DEPARTMENT OF STATISTICS
2024.1-5 Co-instructor, STAT623/423, Probability in bioinformatics and genetics
2023.9 Lecturer, STAT600 Rice Reseaerch Presentation
2022.9 Lecturer, STAT600 Rice Reseaerch Presentation
2019.7 Lecturer, Quantitative Biology Summer School
2013.1 Lecturer, Biostatistics/Bioinformatics STAT453/553

FUDAN UNIVERSITY, SCHOOL OF LIFE SCIENCES, SHANGHAI, CHINA
2015.6.24-6.28 Statistical Genomics Workshop
Lecturer: Drs.Terry Speed, Wenyi Wang, Andrew Teschendorff
Level: Graduate students
Enrollment: 200

TONGJI UNIVERSITY, DEPARTMENT OF BIOINFORMATICS, SHANGHAI, CHINA
2012.7 Lecturer, Invited short course: Assessing risk in familial cancer syndromes
Enrollment: 30

SHANGHAI INSTITUTE OF BIOLOGICAL SCIENCES, CHINESE ACADEMY OF SCIENCES
2011.7 Invited short course: Biostatistics/Bioinformatics
Lecturer: Drs. Jun Liu, Wenyi Wang
Level: first-year graduate students
Enrollment: 400

VISITING TRAVEL SCHOLARSHIP
12.2013 Fudan University Key Laboratory Senior Visiting Scholarship, Fudan University, School of Computer Science, Shanghai, China

HONORS AND AWARDS

2022	Keynote speaker, The 26th Annual International Conference on Research in Computational Molecular Biology. La Jolla, USA, May 22-25, 2022.
2022	American Statistical Association Statistics in genomics & genetics Section Student paper award (PhD student Hoai Nam Nguyen)
2021	American Statistical Association Statistics in genomics & genetics Section Student paper award (PhD student Yujie Jiang)
2014	Outstanding service to graduate education The University of Texas Graduate School of Biomedical Sciences at Houston
2011	The Stellar Abstract Award, the 5th Annual Program in Quantitative Genomics,

Harvard School of Public Health

2008 Phi Beta Kappa, Delta Omega Alpha

2008 The Jane and Steve Dykacz Award for best student paper in medical statistics, Johns Hopkins Biostatistics

2007 Travel Award, the 11th International Conference on Research in Computational and Molecular Biology

2006 Travel Award, the International Genetic Epidemiology Society 15th Annual Meeting

2005 The June B. Culley Award for best performance in the qualifying oral exam, Johns Hopkins Biostatistics