CURRICULUM VITAE of Wenyi Wang

CONTACT

| E-mails: | wwang7@mdanderson.org |
|-----------------|--|
| Mailing address | Department of Bioinformatics and Computational Biology |
| | 1400 Pressler St |
| | Houston, TX 77030 |
| Website: | ${\tt http://odin.mdacc.tmc.edu/}{\sim}{\tt wwang7}$ |

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

Shanghai, China

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

PROFESSIONAL EXPERIENCE

| 9/2015 - present | Tenured Associate Professor Department of Bioinformatics and Computational Biology Department of Biostatistics (joint appointment) University of Texas M. D. Anderson Cancer Center, Houston TX |
|------------------|--|
| 6/2013 - present | <i>Adjunct faculty member</i> Department of Statistics, Texas A&M University, College Station 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 | <i>Visiting scientisit (Sabbatical)</i> Cancer, aging and somatic mutation group (CASM) The Wellcome Trust Sanger Institute, Hinxton UK |
| 9/2010 - 8/2015 | <i>Tenure-track Assistant Professor</i> 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
- 10/2008 12/2008 *Statistical consultant* Counsyl, on software for newborn genetic counseling, Palo Alto CA

PROFESSIONAL ACTIVITIES

Institutional activities: Member of the Clinical Research Advisory Committee (approve internal funding for clinical research including clinical trials), Member of the Rice-MDACC joint biostatistics program admission 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 graduate school admission committee for GSBS.

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 for: 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, Journal of Applied Statistics, Bioinformatics, PLOS Computation Biology

Grant referee for: 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

Study section member: NIH Cancer, Heart, and Sleep Epidemiology (CHSA) Study Section (2017-2021)

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

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.

GRANT SUPPORT

FUNDED

PI of the MDACC subcontract (21%) 4/1/2016-3/31/2021 Consistent variable selection in p >> n settings. 2R01CA158113 (\$654,219, subcontract \$158,524, \$39,631/year) NIH/NCI PI: Valen Johnson

Principal Investigator (15%) 09/24/2014-8/31/2019 Statistical methods for genomic analysis of heterogeneous tumors 1R01CA183793 (\$1,275,421, \$255,084/year); NIH/NCI

Co-Investigator (10%) 12/1/2016-11/30/2019 Identifying vulnerabilities in mutant p53 driven tumorigenesis RP170231 (\$825,739, \$281,237/year) Cancer Prevention Research Institute of Texas PI - Guillermina Lozano

Co-Investigator (4%) 9/1/2014-8/31/2019 Developing New Rational, Personalized Medicine for Lung Cancer Based on Understanding of Lung Cancer Molecular and Cellular Biology \$770,090 NIH/NCI (Subcontract from University of Texas Southwestern Medical Center) PI - John Minna

COMPLETED (selected)

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

Articles

- 1. 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. *under review*
- 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. *under review*
- Wu CC, Beird H, Livingston JL, Advani S, <u>Cao S</u>, Reuben A, Ingram D, Ju Z, Wang WL, Leung CH, Zheng Y, Mitra A, Lin H, RoszikJ, **Wang W**, Patel S, Benjamin R, Gordon M, Hwu P, Gorlick R, Lazar A, Daw N, Lewis V, and Futreal PA. Comprehensive immuno-genomic characterization of relapsed and metastatic osteosarcoma reveals multiple mechanisms of immune suppression. *under review*
- 4. 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. *under review*
- 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: Integrating whole exomes and single-cell transcriptomes to reveal phenotypic impact of somatic variants. *Nature Genetics. under review.* bioRxiv version: doi: https://doi.org/10.1101/413047.

- Dentro S, et al., PCAWG Evolutionary and Heterogeneity Working Group-PCAWG network, Wang W*, Morris QA*, Wedge D*, Van Loo P*. Portraits of genetic intra-tumour heterogeneity and subclonal selection across cancer types. *Nature. under review* *authors contributed equally. bioRxiv version: doi: https://doi.org/10.1101/312041.
- Nikooienejad A, Wang W, Johnson VE. Bayesian variable selection in high dimensional survival time cancer genomic datasets using nonlocal priors. *Annals of Applied Statistics. under review* arXiv version: https://arxiv.org/abs/1712.02964.
- Gerstung M, et al.*, PCAWG Evolutionary and Heterogeneity Working Group-PCAWG network. The evolutionary history of 2,658 cancers. *under 3rd revision at Nature*. bioRxiv version: doi: https://doi.org/10.1101/161562. *Wang leads the effort in consensus subclonal reconstruction of the PCAWG WGS data.
- 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.
- Wang Z, Morris JS, <u>Cao S</u>, <u>Ahn J</u>, <u>Liu R</u>, 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*. Nov 30;9:451-460. doi: 10.1016/j.isci.2018.10.028.
- Li J, Shin SJ, 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.
- 12. <u>Shin SJ</u>, 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.
- 13. <u>Li J</u>, 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.
- 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.
- Peng G, Bojadzieva J, Ballinger ML, Li J, Savage S, Mai P, Thomas DM, Strong LC, Wang WM. 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.
- 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.

- 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.*
- Nikooienejad A, Wang W A, 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.
- Palculict TB, Ruteshouser EC, <u>Fan Y</u>, **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.
- 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.
- 21. 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.
- Fang LT, Afshar PT, Chhibber A, Mohiyuddin M, <u>Fan Y</u>, 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.
- 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-nucleotidevariant detection. *Nat Methods.* 2015 Jul;12(7):623-30. doi: 10.1038/nmeth.3407. Epub 2015 May 18.
- 24. 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.
- 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
- 26. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* 507(7492):315-22, 3/2014. e-Pub 1/2014.

- <u>Ahn J</u>, 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
- 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.
- 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.
- 30. 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 equally
- <u>Ahn J</u>, 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.
- 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.
- 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.
- 35. 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.
- 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

37. 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

 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.

- 39. 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.
- 40. Wang W, Carvalho B, Miller N, Pevsner J, Chakaravarti A and Irizarry RA. Estimating genomewide copy number using allele specific mixture models. *Journal of Computational Biology.* 2008, 15:857-66.
- 41. 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.
- 42. 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.
- 43. 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.
- 44. 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.
- 45. 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

 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. 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
- 2. 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
- 3. Robust subclonal architecture reconstruction from \sim 2,700 cancer genomes. University of Houston Department of Mathematics. Houston, TX 11/30/2018
- Deconvolution of multi-omics Data from Heterogeneous Tumor Samples. TAMU Bioinformatics & Cancer Symposium, College Station, TX 09/21/2018

- 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
- Cancer risk assessment and penetrance estimation for deleterious mutations in TP53 using a multicentre consortium. The 4th International Li-Fraumeni Syndrome Conference. 04/26/2018
- Robust subclonal architecture reconstruction from ~ 2,700 cancer genomes. ENAR Spring Meeting, Atlanta, Georgia. 03/27/2018
- 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
- 9. 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
- 10. Statistical methods for the deconvolution of high-throughput sequencing data from heterogeneous tumor samples. Barts Cancer Institute, London, UK 02/08/2018
- 11. Statistical methods for the deconvolution of mixed cancer transcriptomes. Baylor College of Medicine Breast Disease Working Group. Houston, TX 01/10/2018
- 12. Robust subclonal architecture reconstruction from \sim 2,700 cancer genomes. Fudan University Shanghai Cancer Center Department of Pathology. Shanghai, China 12/22/2017
- 13. 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
- 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
- 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
- 16. Statistical inference problems for the gene expression deconvolution of heterogeneous tumor samples. Systems Genetics of Cancer, London UK 09/18-09/20/2017.
- 17. Statistical methods for the gene expression deconvolution of heterogeneous tumor sample. The European Molecular Biology Laboratory (EMBL), Heidelberg Germany 07/27/2017.
- Statistical methods for the gene expression deconvolution of heterogeneous tumor sample. Cancer Research UK Cambridge Institute, Cambridge UK 07/24/2017.
- 19. Statistical methods for the gene expression deconvolution of heterogeneous tumor sample. Oxford University Big Data Institute, Oxford UK 07/21/2017.
- 20. Statistical methods for the gene expression deconvolution of heterogeneous tumor sample. UNC Chapel Hill Department of Biostatistics, Chapel Hill, NC 04/27/2017.
- 21. Statistical methods for the gene expression deconvolution of heterogeneous tumor sample. Rice University Department of Statistics, Houston TX 04/17/2017.

- 22. 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.
- 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.
- 24. 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.
- 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.
- 26. 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.
- 27. 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.
- 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.
- 29. 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.
- 30. 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.
- 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
- 32. Gene expression deconvolution of heterogeneous tumor samples: DeMix-Bayes. The Francis Crick Institute. London, UK, 12/15/2015
- 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
- Gene expression deconvolution in heterogeneous tumor samples using DeMix-Bayes. NCI DCCPS New Grantee Workshop, 11/5/2015
- 35. Bayesian variable selection for binary outcomes in high dimensional settings. Joint Statistics Meeting. Seattle, WA, 8/9/2015
- 36. Cancer-specific characterization of the Li-Fraumeni syndrome. A.C. Carmargo Cancer Center, Sau Paulo, Brazil, 7/28/2015

- Statistical methods for analysis of genomic data from heterogeneous cancer samples. BioC 2015, Seattle, WA, 7/21/2015
- Statistical methods for analysis of genomic data from heterogeneous cancer samples. International Bioinformatics Workshop. Harbin, China, 7/10/2015
- Statistical methods for analysis of genomic data from heterogeneous cancer samples. University of Evry, Essonne, France, 5/19/2015
- 40. Statistical methods for analysis of genomic data from heterogeneous cancer samples, Johns Hopkins University, Department of Biostatistics, Baltimore, MD, 4/6/2015
- 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
- 42. 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
- 43. Cancer-specific characterization of the Li-Fraumeni Syndrome, Worcester Polytechnic Institute, Biomedical Engineering, Worcester, MA, 11/6/2014
- 44. 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
- 45. MuSE: somatic evolution estimation for mutation calling in sequencing data of matched tumornormal samples, The Joint Statistical Meetings, *The Joint Statistical Meetings*, Boston, MA, 2014.8.4
- 46. Bayesian variable selection for binary outcomes in high dimensional settings, *The International Society of Bayesian Analysis World Meeting*, Cancun, Mexico, 2014.7.17
- 47. Somatic Mutation Calling and Gene Expression Deconvolution in Heterogeneous Tumor Samples, University of Oxford, Department of Oncology, Oxford, United Kingdom, 2014.6.12
- 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
- 49. Gene expression deconvolution in heterogeneous tumor samples, Texas A&M Unversity, Department of Statistics, College Station, TX, 2014.2.21
- 50. Gene expression deconvolution in heterogeneous tumor samples, 7th Annual Bayesian Biostatistics and Bioinformatics Conference, Houston, TX, 2014.2.13
- 51. Gene expression deconvolution in heterogeneous tumor samples. *The 9th ICSA International Conference*, HongKong, China, 2013.12.23
- 52. Personalized risk assessment for families with Li-Fraumeni Syndromes. *The LiFE Consortium Meeting*. Boston, MA, 2013.10.26

- 53. Gene expression deconvolution in heterogeneous tumor samples. *The Joint Statistics Meetings*. Montreal, Canada, 2013.8.3
- 54. Rare variant detection using family-based sequencing analysis. Washington University Department of Genetics. St. Louis, MO, 2012.11
- 55. Determining probability of rare variants: implications for designs of family-based sequencing studies. *The Joint Statistical Meetings*. San Diego, CA, 2012.08
- 56. Determining probability of rare variants in sequencing studies for familial cancer syndromes. *The International Workshop on Cancer Systems Biology*. Changchun, China, 2012.07
- 57. 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
- 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
- 59. Determining probability of rare variants: implications for designs of family-based sequencing studies. UT School of Public Health Human Genetics Center, 2011.9
- 60. Determining probability of germline mutations in family-based sequencing studies. The First Wuxi International Statistics Forum, Wuxi, China, 2011.7
- 61. Validating risk prediction models using family registries, *Fourth Annual Bayesian Biostatistics Conference*, Houston, TX, 2011.1
- 62. Statistical methods for DNA resequencing analysis in disease-gene studies, Rice University, Department of Statistics, Houston, TX, 2011.1

SOFTWARE

- 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 cancer genomes
 - R package, http://github.com/wwylab/DeMixT.
- 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
- 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. CliP: Clonal identification of subclonal structure through pairwise penalization using sequencing data from tumor samples. (C code)
 - Code, http://github.com/wwylab/CliP.
- 9. CSR: Consensus Clustering for Subclonal structure Reconstruction.
 - Code, http://github.com/wwylab/CSR.

TEACHING EXPERIENCE

The UNIVERSITY OF TEXAS MD ANDERSON CANCER CENTER

| PO9 | STE | 00 | С |
|-----|-----|----|---|
| | | | - |

- 2016.9-present Shaolong Cao (PhD in Biomedical Engineering, University of Tulane)
- 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: Senior Research Scientist, Phillips
- 2013.7-2015.2 Seung Jun Shin (joint with Ying Yuan, PhD in Statistics, North Carolina State University) Current position: Tenure-track Assistant 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: Tenure-track Assistant Professor in Biostatistics and Bioinformatics, Georgetown University, Washington DC

PHD STUDENT

| 2019.1-present | Feng Zhang (UTSPH Biostatistics) |
|-------------------|---|
| 2018.9-present | Jingxiao Chen (UTSPH Biostatistics) |
| 2017.9-present | Fan Gao (Rice Statistics) |
| 2017.9-present | Carlos Vera Recio (GSBS MD-PhD Biostatistics, NLM fellow) |
| 2013.6-2017.4 | Zeya Wang (Rice Statistics, currently Data Scientist at Pentuum) |
| 2013.5-2017.4 | Jialu Li (GSBS Biostatistics, currently Senior Biostatistician at Genomic Health) |
| 2011.1-2015.8 | Gang Peng (GSBS Biostatistics, AAAS/Science Program for Excellence in Science, |
| currently Postdoc | at Yale Biostatistics) |
| RESEARCH ASS | ISTANT |
| 2014.8-present | Elissa Dodd |
| 2018.9-2019.1 | Huiming Lin (Rice Statistics) |
| 2013.9-2015.5 | Amir Nikooienejad (Texas A&M Statistics) |
| ROTATION STU | DENT |
| 2018.10-2018.12 | Wanli Wang (Baylor College of Medicine Quantitative Computational Biology) |
| 2018.9-2018.12 | Wenhao Li(GSBS Quantitative Sciences) |
| 2018.3-2018.5 | Yipeng Gao (Baylor College of Medicine Quantitative Computational Biology) |
| 2017.9-2017.12 | Ramiz Iqbal (GSBS Quantitative Sciences) |
| 2015.9-2015.12 | Yichi Zhang (GSBS Quantitative Sciences) |
| 2012.3-2012.5 | Yihua Liu (GSBS Quantitative Sciences) |
| INTERN | |
| 2019.5-2019.8 | Timothy Cheng (Texas A&M, Undergraduate in Statistics) |
| 2018.6-2018.8 | Yutong Yao (University of Notre Dame, Premed, MDACC Summer Experience Pro- |
| gram) | |
| 2018.5-2018.8 | Xin Su (Texas A&M, Undergraduate in Statistics) |
| 2013.6-2013.8 | Emily Johnson (Bellaire high school, Houston TX) |
| 2012.11-2013.7 | Sara Algeri (Texas A&M Statistics) |
| 2012.5-2012.8 | Ling Chen (Rice statistics) |
| 2012.5-2012.8 | Sanvesh Srivastava (Purdue statistics) |
| 2011.1-2011.9 | Yan Huang (Rice statistics) |
| 2011.1-2011.5 | Yulun Liu (UT SPH Biostat) |

External thesis examiner of

2019 Stefano Mangliola (University of Melbourne)

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- present Instructor, Computational Approaches for Single-Cell Data Analysis (GS01 1041) 2018.11 Lecturer, Introduction to Bioinformatics GS0011062 2017.11 Lecturer, Introduction to Bioinformatics GS0011062 2016.11 Lecturer, Introduction to Bioinformatics GS0011062 2015.12 Lecturer, Introduction to Bioinformatics GS0011062 2014.11 Lecturer, Introduction to Bioinformatics GS0011062 2013.10 Lecturer, Introduction to Bioinformatics GS0011062

RICE UNIVERSITY, DEPARTMENT OF STATISTICS 2013.1 Lecturer, Biostatistics/Bioinformatics STAT453/553

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

TONGJI UNIVERSITY, DEPARTMENT OF BIOINFORMATICS, SHANGHAI, CHINA 2012.7 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

Instructor: Drs. Jun Liu, Wenyi Wang Level: first-year graduate students Enrollment: 400

VISITING SCHOLARSHIP

01.2013-12.2014 Fudan University Key Laboratory Senior Visiting Scholarship, Fudan University, School of Computer Science, Shanghai, China

HONORS AND AWARDS

| 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