Sha Cao

Current Information

Assistant Professor
Department of Biostatistics, School of Medicine
Indiana University
410 West 10th Street, Suite 3073
Indianapolis, IN 46202
E-mail: shacao@iu.edu; Office: 317-274-2602

Education

2014.01-2017.05	Ph.D., Statistics, University of Georgia
2011.09-2014.01	Ph.D student in Bioinformatics, University of Georgia
2007.09-2011.07	B.S., Mathematics, Beijing Normal University

Professional Experiences

2019.12-now	(adjunct) Assistant Professor, Department of BioHealth Informatics, School of Informatics and Computing, IUPUI, Indianapolis, IN
2017.8-now	Assistant Professor, Department of Biostatistics, School of Medicine, Indiana University, Indianapolis, IN
2013.9-2017.7	Research Assistant, Computational Systems Biology Lab, University of Georgia, Athens, GA
2011.9-2013.9	Graduate Research Assistant, Institute of Bioinformatics, University of Georgia, Athens, GA
2009-2011	Undergraduate research assistant, Lab of Computational Molecular Biology, Beijing Normal University, Beijing, China

Publications

(Methodology)

1. Changlin Wan, Wennan Chang, Tong Zhao, **Sha Cao***, Chi Zhang*. Geometric all-way Boolean tensor decomposition. **Advances in Neural Information Processing Systems 33 (NeurIPS)** (2020). In press.
2. Wennan Chang, Changlin Wan, Yong Zang, Chi Zhang, and **Sha Cao***. Supervised clustering of high dimensional data using regularized mixture modeling. **Briefings in Bioinformatics**. (2020) In press.
3. Xiaoyu Lu, Szu-wei Tu, Wennan Chang, Changlin Wan, Yifan Sun, Baskar Ramdas, Xin Lu, Shannon Hawkins, Reuben Kapur, Xiongbin Lu*, **Sha Cao*,** Chi Zhang*. SSMD: A semi-supervised approach for a robust cell type identification and deconvolution of mouse transcriptomics data. **Briefings in Bioinformatics**. (2020) In press.

4. Zhang, Yifei; **Cao, Sha**; Zhang, Chi; Jin, Ick Hoon; Zang, Yong. A Bayesian Adaptive Phase I/II Clinical Trial Design with Late-onset Competing Risk Outcomes. **Biometrics**. 10.1111/biom.13347. In press.

5. Han Y, Liu H, **Cao S**, Zhang C, Zang Y. TSNP: A Two-Stage Nonparametric Phase I/II Clinical Trial Design for Immunotherapy. **Pharmaceutical Statistics**. 2020. In Press.

6. Wan C, Chang W, Zhao T, Zang Y, **Cao S**[*], Zhang C[*]. Denoising individual bias for a fairer binary submatrix detection. In Proceedings of the 29th ACM International Conference on Information and Knowledge Management (**CIKM**) Accepted.

7. Yu X, **Cao S**, Zhou Y, Yu Z, Xu Y. Co-expression based cancer staging and application. **Scientific Reports**. 2020 Jun 30;10(1):1-0.

8. Upadhyaya Y, Xie L, Salama P, **Cao S**, Nho K, Saykin AJ, Yan J, Alzheimer's Disease Neuroimaging Initiative. Differential co-expression analysis reveals early stage transcriptomic decoupling in alzheimer's disease. BMC medical genomics. 2020 Apr;13(5):1-0.

9. Huang Z, Johnson TS, Han Z, Helm B, **Cao S**, Zhang C, Salama P, Rizkalla M, Yu CY, Cheng J, Xiang S. Deep learning-based cancer survival prognosis from RNA-seq data: approaches and evaluations. BMC medical genomics. 2020 Apr;13:1-2.

10. Wan C, Chang W, Zhao T, Li M, **Cao S**[*], Zhang C[*]. MEBF: a fast and efficient Boolean matrix factorization method. **AAAI**. In press. (acceptance rate 20%)

11. Han Y, and Yuan Y, and **Cao S**, Li M, and Zang Y. On the Use of Marker Strategy Design to Detect Predictive Marker Effect in Cancer Immunotherapy and Targeted Therapy. **Statistics in Biosciences**. 10.1007/s12561-019-09255-1.

12. Xie J, Ma A, Zhang Y, Liu B, **Cao S**, Wang C, Xu J, Zhang C, Ma Q. QUBIC2: a novel and robust biclustering algorithm for analyses and interpretation of large-scale RNA-Seq data. **Bioinformatics**. 2019 Sep 10.

13. Wan C, Chang W, Zhang Y, Shah F, Lu X, Zang Y, Zhang A, **Cao S**, Fishel ML, Ma Q, Zhang C. LTMG: a novel statistical modeling of transcriptional expression states in single-cell RNA-Seq data. **Nucleic acids research**. 2019 Aug 2.

14. Zhi Huang, Travis Johnson, Zhi Han, Bryan Helm, **Sha Cao**, Chi Zhang, Paul Salama, Maher Rizkalla, Christina Yu, Jun Cheng, Shunian Xiang, Xiaohui Zhan, Jie Zhang and Kun Huang. Deep Learning-based Cancer Survival Prognosis from RNA-seq Data: Approaches and Evaluations. ICIBM 2019 (In press).

15. Zhang Y, Wan C, Wang P, Chang W, Huo Y, Chen J, Ma Q, **Cao S**, Zhang C. M3S: A comprehensive model selection for multi-modal single-cell RNA sequencing data. BMC bioinformatics. 2019 Dec;20(24):1-5.

16. Zang Y, Guo B, Han Y, **Cao S**, Zhang C. A Bayesian adaptive marker-stratified design for molecularly targeted agents with customized hierarchical modeling. **Statistics in medicine**. 2019 Jan 1.

17. Zang Y, Fung WK, **Cao S**, Ng HK, Zhang C. Robust tests for gene–environment interaction in case-control and case-only designs. **Computational Statistics & Data Analysis**. 2019 Jan 1;129:79-92.

18. Zhang Y, **Cao S**, Zhao J, Alsaihati B, Ma Q, Zhang C. MRHCA: a nonparametric statistics based method for hub and co-expression module identification in large gene co-expression network. **Quantitative Biology**. 2018 6(1).

(Collaboration)

1. Dong L, Watson J, **Cao S**, Arregui S, Saxena V, Ketz J, Awol AK, Cohen DM, Caterino JM, Hains DS, Schwaderer AL. Aptamer based proteomic pilot study reveals a urine signature indicative of pediatric urinary tract infections. **PloS one.** 2020 Jul 6;15(7):e0235328.

2. Yuanzhang Fang, Lifei Wang, Changlin Wan, Yifan Sun, Kevin Van der Jeught, Zhuolong Zhou, Tianhan Dong, Ka Man So, Tao Yu1, Yujing Li, Haniyeh Eyvani, Austyn B Colter, Edward Dong, **Sha Cao**, Jin Wang, Bryan P Schneider, George E. Sandusky, Yunlong Liu, Chi Zhang, Xiongbin Lu, Xinna Zhang. Mal2 drives immune evasion by reducing antigen presentation on tumor cells. **Journal of Clinical Investigation**. 2020.

3. Natascia Marino, Rana German1, Xi Rao, Ed Simpson, Sheng Liu, Jun Wan, Yunlong Liu, George Sandusky, Max Jacobsen, Miranda Stoval, **Sha Cao**, Anna Maria V. Storniolo. Upregulation of lipid metabolism genes in the breast prior to cancer diagnosis. **NPJ Breast Cancer**. 2020. In Press.

4. Spence JP, Lai D, Reiter JL, **Cao S**, Bell RL, Williams KE, Liang T. Epigenetic changes on rat chromosome 4 contribute to disparate alcohol drinking behavior in alcohol-preferring and–nonpreferring rats. **Alcohol**. 2020 Aug 14.

5. Parker JG, Diller EE, **Cao S**, Nelson JT, Yeom K, Ho C, Lober R. Statistical multiscale mapping of IDH1, MGMT, and microvascular proliferation in human brain tumors from multiparametric MR and spatially-registered core biopsy. **Scientific Reports**. 2019 10.1038/s41598-019-53256-5

6. Miller SA, Policastro RA, Savant SS, Sriramkumar S, Ding N, Lu X, Mohammad HP, **Cao S**, Kalin JH, Cole PA, Zentner GE. Lysine-specific demethylase 1 mediates AKT activity and promotes epithelial-mesenchymal transition in PIK3CA mutant colorectal cancer. **Molecular Cancer Research**. 2019 Nov.

7. Sun H, Chen L, **Cao S**, Liang Y, Xu Y. Warburg Effects in Cancer and Normal Proliferating Cells: Two Tales of the Same Name. **Genomics, proteomics & bioinformatics**. 2019 May 7.

8. Sun H, Zhang C, **Cao S**, Sheng T, Dong N, Xu Y. Fenton Reactions Drive Nucleotide and ATP Syntheses in Cancer. **Journal of molecular cell biology**. 2018 Jul 16.

9. Diller EE, **Cao S**, Ey B, Lober R, Parker JG. Predicted disease compositions of human gliomas estimated from multiparametric MRI can predict endothelial proliferation, tumor grade, and overall survival. arXiv preprint arXiv:1908.02334. 2019 Aug 6.

10. Xiyin Wang, Robert Emerson, Chi Zhang, **Sha Cao**, Xiaoyu Lu, Doug Rusch, Aaron Buechlein, Russell Broaddus, Francesco DeMayo, John Lydon, and Shannon Hawkins. High-risk histologic endometrial carcinoma in a preclinical mouse model with uterine-specific deletion of Pten and Dicer. (Submitted to Cancer Research)

(Prior to joining IU)

1. Tian Y, Du W, **Cao S**, Wu Y, Dong N, Wang Y, Xu Y. Systematic analyses of glutamine and glutamate metabolisms across different cancer types. **Chinese journal of cancer**. 2017 Dec;36(1):88. (Figure 5 selected as cover art.)

2. **Cao S**, Zhou Y, Wu Y, Song T, Alsaihati B, Xu Y. Transcription regulation by DNA methylation under stressful conditions in human cancer. **Quantitative Biology**. 2017 Dec 1;5(4):328-37. 1.

3. Song T, **Cao S**, Tao S, Liang S, Du W, Liang Y. A Novel Unsupervised Algorithm for Biological Process-based Analysis on Cancer. **Scientific reports**. 2017 Jul 5;7(1):4671.

4. **Cao S**+, Zhu X+, Zhang C, Qian H, Schuttler HB, Gong JP, Xu Y. Competition between DNA methylation, nucleotide synthesis and anti-oxidation in cancer versus normal tissues. **Cancer Res**. (2017) DOI: 10.1158/0008-5472.CAN-17-0262.

5. Song T, Wang Y, Du W, **Cao S**, Tian Y, Liang Y. The Method for Breast Cancer Grade Prediction and Pathway Analysis Based on Improved Multiple Kernel Learning. **J. Bioinform. Comput. Biol.** 15, 1650037 (2017).

6. Zhang C, Sheng T, **Cao S**, Xu Y. Computational analysis of the impact of Autophagy in different stages of cancer progression, a chapter in book "Autophagy and Cancer", Editor: Jin-Ming Yang, Springer, 2016 (in press).

7. Coothankandaswamy, V., **Cao, S**., Xu, Y., Prasad, P. D., Singh, P. K., Reynolds, C. P., Yang, S., Ogura, J., Ganapathy, V., and Bhutia, Y. D. (2016) Amino acid transporter SLC6A14 is a novel and effective drug target for pancreatic cancer. **British Journal of Pharmacology**. 173: 3292–3306. doi: 10.1111/bph.13616.

8. **Cao S**[+], Zhang C[+], and Xu Y. Somatic Mutations May Not Be the Primary Drivers of Cancer Formation. **International Journal of Cancer**. (2015) DOI: 10.1002/ijc.29639.

9. Zhang C[+], Chao L[+], **Cao S**[+] and Xu Y. Elucidation of Divers of High-Level Production of Lactates throughout a Cancer Development. **Journal of Molecular Cell Biology**. (2015) doi: 10.1093/jmcb/mjv031

10. Chou WC, Ma Q, Yang S, **Cao S**, Klingeman DM, Brown SD, Xu Y. (2015) Analysis of strand-specific RNA-seq data using machine learning reveals the structures of transcription units in Clostridium thermocellum. **Nucleic Acids Research**. doi: 10.1093/nar/gkv177

11. Liu C+, Zhang C+, Su J, Zhang DS, **Cao S**[*]. (2015) Stresses drive a cancer's initiation, progression and metastasis: Critical comments on the book "Cancer Bioinformatics". **Journal of Bioinformatics and Computational Biology**. DOI: 10.1142/S021972001571002X

12. Zhang C, **Cao S**, Toole B, and Xu Y. (2014) Cancer may be a pathway to cell survival under persistent hypoxia and elevated ROS: A model for solid-cancer initiation and early development. **International Journal of Cancer**: DOI: 10.1002/ijc.28975.

13. Zhang C, **Cao S** and Xu Y. (2014) Population Dynamics inside Cancer Biomass Driven by Repeated Hypoxia-Reoxygenation Cycles. **Quantitative Biology**. Doi: 10.1007/s40484-014-0032-8

14. Cui J, Yin Y, Ma Q, Wang G, Olman V, Zhang Y, Chou WC, Hong CS, Zhang C, **Cao S**, Mao X, Li Y, Qin S, Zhao S, Jiang J, Hastings P, Li F, Xu Y (2014) Comprehensive characterization of the genomic alterations in human gastric cancer. **International Journal of Cancer**. DOI: 10.1002/ijc.29352

Manuscripts Submitted

1. Wennan Chang, Changlin Wan,Chun Yu, Weixin Yao, Chi Zhang, and **Sha Cao***. RobMixReg: an R package for robust, flexible and high dimensional mixture regression. (Submitted to Bioinformatics)
2. Alex White, Yong Zang, Chi Zhang, **Sha Cao***. Submatrix Detection via Adaptive Thresholding SVD. (Ready for submission)
3. Junyi Zhou, Xiaoyu Lu, Wennan Chang, Changlin Wan, Chi Zhang, **Sha Cao***. PLUS: predicting pan-cancer metastasis potential based on positive and unlabeled learning. (Submitted to Genome Biology)
4. Chang W, Zhou X, Zang Y, Zhang C, **Cao S***. Component-wise Adaptive Trimming For Robust Mixture Regression. arXiv preprint arXiv:2005.11599. 2020 May 23.
5. Chang W, Wan C, Zhang Y, So K, Richardson B, Sun Y, Zhang X, Huang K, Zhang A, Lu X, **Cao S***, Zhang C*. ICTD: Inference of cell types and deconvolution--a next-generation deconvolution method for accurate assess cell population and activities in tumor microenvironment. bioRxiv. 2018 Jan 1:426593.
6. **Cao S**, Chang W, Wan C, Zang Y, Ma Q, Zhang C. Pipeline for Characterizing Alternative Mechanisms (PCAM) based on bi-clustering to study colorectal cancer heterogeneity. (In revision with NAR Cancer)
7. Wennan Chang, Changlin Wan, Xiaoyu Lu, Pengtao Dang,, Yue Fang, Yong Zang, Yunlong Liu*, **Sha Cao***, Chi Zhang*. Accurate identification of cell type and phenotypic marker genes in single cell transcriptomic data via a data augmentation approach.

+: co-first authors; *: corresponding author

Awards

1. Showalter Trust Young Investigator Award, Indiana CTSI, 04/2019
2. Grimes Family Distinguished Graduate Fellowship in Natural Sciences, The University of Georgia. 11/2015
3. Innovative and Interdisciplinary Research Award, The University of Georgia, 04/2014

Grant as Principle Investigators

1. CRII: III: Computational framework for disparate data integration and its application in studying cancer drug resistance.
   NSF/CISE, Primary Investigator
   $174,747 (5%), 09/2019-08/2021
2. Delineating the variational epigenetic regulation on gene expression conditional on cell fractions in cancer tissue and its implications on alternative drug resistance mechanisms
   Indiana CTSI/Showalter Trust, Primary Investigator

$60,000 (8%), 07/2019-06/2020
3. A computational framework to facilitate cancer epigenetics studies
    Indiana CTSI/BERD pilot grant, Primary Investigator
    $10,000 (8%), 05/2019-04/2020

Invited Talks

- 2020.08.03 The 2020 Joint Statistical Meetings, Virtual
- 2020.05.05 Melvin and Bren Simon Cancer Center, Indiana University School of Medicine, Indianapolis, IN
- 2019.01.25 Department of Electrical and Computer Engineering, IUPUI, Indianapolis, IN
- 2018.07.18 Sun Yat-sen University Cancer Center, Guangzhou, China
- 2018.07.02 The 8th International Workshop on Cancer Systems Biology, Changchun, China
- 2018.06.16 2018 ICSA Applied Statistics Symposium, New Brunswick, NJ
- 2018.04.07 WPI-Advancing Women's Impact in Mathematics Symposium, Worcester, MA
- 2018.04.03 Department of Mathematical Sciences, IUPUI, Indianapolis, IN
- 2017.02.20 Department of Mathematical Sciences, Michigan Technological University, Houghton, MI
- 2017.02.09 Department of Biostatistics, Indiana University School of Medicine, Indianapolis, IN
- 2017.02.03 Division of Biostatistics at the University of Minnesota. Minneapolis, MN
- 2016.10.25 Center for Simulational Physics, The University of Georgia, Athens, GA
- 2016.10.20 Department of Bioinformatics and Biostatistics, Emory University, Atlanta, GA
- 2015.07.08 The University of Texas MD Anderson Cancer Center, Houston, TX
- 2015.07.09 Professor Wei Li's lab, Baylor College of Medicine, Houston, TX
- 2015.07.13 Center for Systems Biology, University of Texas at Dallas, Dallas, TX
- 2015.07.14 Professor Yi Li's lab, Baylor College of Medicine, Houston, TX
- 2015.07.15 Department of Bioinformatics and Computational Biology, The University of Texas MD

Teaching Experiences

2019 Spring   B670, instructor, IUSM Department of Biostatistics.
2018 Fall, 2019 Fall G788, guest lecturer, IUSM Department of Medical and Molecule Genetics and Molecular Biology (Organizing instructor: Professor Yunlong Liu).
2018/2020 Spring   PBHL-B584, guest lecturer, IUSM Department of Biostatistics (Organizing instructor: Professor Barry Katz, Susan Perkins).

2017 Spring	BCMB8125, guest lecturer, UGA Department of Biochemistry and Molecular Biology (Instructor: Professor Ying Xu).
2016 Fall	BCMB8210, guest lecturer, UGA Department of Biochemistry and Molecular Biology (organizing instructor: Professor Ying Xu).
2016 Spring	BCMB/CBIO/GENE8112, guest lecture, UGA Department of Biochemistry and Molecular Biology (organizing instructor: Professor Kelley Moremen).

Mentoring Experiences

- Alex White (2019.08-now), PhD student in Dept of Biostatistics, IUPUI
- Xinyu Zhou (2018.09-now), visiting student from Jilin University
- Xiaoyu Lu (2018.04-now), PhD student from Department of BioHealth Informatics, IUPUI
- Yan Han (2017.09-2019.06), master student from Department of Biostatics, IUSM
- Xinhua Hu (2018.03-2018.06), master student from Department of Biostatics, IUSM
- Kayla Sepsick (2017.06-2017.07), undergraduate student from BCM, University of Georgia.
- Ashford-Carroll, Brianna Suree (2017.05-2017.07), undergraduate student from NSF-sponsored Research Experience for Undergraduates (REU) Summer Program, University of Georgia
- Jason Kwak (2015.03-2017.07), undergraduate student from BCMB, UGA
- Naqeeb Mohammed Faroqui (2014.07-2015.12), pre-med student from UGA
- Yen Xuan Le (2012.08-2013.04), undergraduate student from College of Pharmacy, UGA

Students that I'm serving on dissertation committees: Enze Liu, Junyi Zhou, Norah Alghamdi, Pengtao Dang
Students that I'm serving as Biostatistics PhD minor advisor: Ting Wang, Enze Liu, Duojiao Chen, Briana Paisley

Academic Services

2012-now	Reviewer, PLOS One
2014	Reviewer, Nucleic Acid Research
2015-now	Reviewer, IEEE TCBB
2015-now	Reviewer, Mathematical Biosciences
2015-now	Reviewer, Journal of Bioinformatics and Computational Biology
2016-now	Reviewer, BIBM
2017-now	Reviewer, Journal of Molecular Cell Biology
2017-now	PC member, BIBM
2017	PC member, BIOKDD'17
2018	Organizing committee member, ICSB 2018, Changchun
2018	Session chair, 2018 IBS-China, Guangzhou

2018-now	Academic editor, PLOS One
2019-now	Reviewer, Briefings in Bioinformatics
2020-now	Reviewer, PeerJ
2020-now	Reviewer, Genomics, Proteomics and Bioinformatics (GPB)