RUIBIN XI

School of Mathematical Sciences
Peking University
No.5 Yiheyuan Road, Haidian District
Beijing, P.R.China 100871
Email: ruibinxi@math.pku.edu.cn
Web Page: http://www.math.pku.edu.cn/teachers/xirb

EDUCATION

Washington University in St. Louis, Ph.D. in Mathematics, 2009
Dissertation Title: Statistical Aggregation: Theory and Applications
Supervisor: Nan Lin

Washington University in St. Louis, M.A. in Mathematics, 2007

Peking University, M.A. in Mathematics, 2005
Dissertation Title: Kupka-Smale Systems with Obstacle Sets
Supervisor: Lan Wen

Inner Mongolia University, B.S. in Mathematics, 2002

EMPLOYMENT

Tenured Associate Professor/Research Fellow, School of Mathematical Sciences and Center for Statistical Science, Peking University. Sep. 2019 - Now

Assistant Professor/Research Fellow, School of Mathematical Sciences and Center for Statistical Science, Peking University. Sep. 2012 - 2018

Research Associate, Center for Biomedical Informatics, Harvard Medical School. Aug. 2009 - Aug. 2012
Supervisor: Peter J. Park

RESEARCH INTERESTS

Statistical analysis of next-generation sequencing, bioinformatics, cancer genomics, precision oncology, single cell RNA sequencing, statistical computing, big data analysis, high-dimensinal data analysis, Bayesian statistics.

PUBLICATIONS

*: correspondence authors; #: co-first authors.

Methodology Work in Statistics and Bioinformatics

1. Jin, Z., Huang, W., Shen, N., Li, J., Wang, X., Dong, J., Park, P. and **Xi, R.*** (2022) Single cell gene fusion detection by scFusion, *Nature Communications*, 13, 1084.
2. Wu, L., Wang, H., Xia, Y. and **Xi, R.*** (2020) CNV-BAC: Copy number Variation Detection in Bacterial Circular Genome. *Bioinformatics*, 36(12), 3890-3891.
3. Chen, C., Wu, C., Wu, L. Wang, X., Deng, M. and **Xi, R.*** (2020) scRMD: Imputation for

single cell RNA-seq data via robust matrix decomposition. *Bioinformatics*, 36(10), 3156-3161.

4. Xia,Y.#, Liu, Y.#, Deng, M. and **Xi, R.*** (2019) Detecting virus integration sites based on multiple related sequencing data by VirTect, *BMC Medical Genomics*, 12(1), 19.
5. Chen, C., **Xi, R.** * and Lin, N. (2018) Community detection by L_0-penalized graph Laplacian. *Electronic Journal of Statistics*, 12(1), 1842-1866.
6. Xia, Y., Liu, Y., Deng, M. and **Xi, R.*** (2017) SVmine improves structural variation detection by integrative mining of predictions from multiple algorithms, *Bioinformatics*, 33(21):3348-3354.
7. Yuan, H., **Xi, R.***, Chen, C. and Deng, M. (2017) Differential network analysis via lasso penalized D-trace loss, *Biometrika*, 104 (4): 755-770.
8. Xia, Y., Liu Y., Deng, M. and **Xi, R.** * (2017) Pysim-sv: a package for simulating structural variation data with GC-biases. *BMC Bioinformatics*, 18(Suppl 3):53.
9. **Xi, R.***, Lee, S., Xia, Y., Kim, T. and Park, P. J.* (2016) Copy number analysis of whole-genome data using BIC-seq2 and its application to detection of cancer susceptibility variants, *Nucleic Acids Research*, 44(13):6274-6286.
10. **Xi, R.***, Li, Y. and Hu, Y. (2015) Bayesian quantile regression based on the empirical likelihood with spike and slab priors, *Bayesian Analysis*, 11(3): 821-855.
11. **Xi, R.** and Lin, N.* (2015) Direct regression modelling of high-order moments in big data, *Statistics and Its Interface*, 9(4):445-452.
12. Yang, L., Luquette, L.J., , Gehlenborg, N., **Xi, R.**, Haseley, P.S., Hsieh, C.H., Zhang, C., Ren, X., Protopopov, A., Chin, L., Kucherlapati, R., Lee, C. and Park, P.J.* (2013) Diverse mechanisms of somatic structural variations in human cancer genomes. *Cell*, 153(4):919-929.
13. **Xi, R.**, Lin, N., Chen, Y.* and Kim, Y. (2012) Compression and aggregation of Bayesian estimates for data intensive computing, *Knowledge and Information Systems*, 33(1):191-212.
14. **Xi, R.**, Lee, S. and Park, P.J.* (2012) A Survey of copy number variation detection tools based on high-throughput sequencing data, *Current Protocols in Human Genetics*, Chapter 7:Unit7.19, John Wiley & Sons, Inc.
15. **Xi, R.**, Hadjipanayis, A.G., Luquette, L.J., Kim, T.M., Lee, E., Zhang, J.H., Johnson, M.D., Muzny, D.M., Wheeler, D.A., Kucherlapati, R., and Park, P.J.* (2011) Copy number alteration detection in sequencing data using the Bayesian information criterion, *Proceedings of the National Academy of Sciences, USA*, 108(46):E1128-1136.
16. **Xi, R.**, Kim, T. M. and Park, P.J.* (2010) Detecting structural variation using next-generation sequencing, *Briefings in Functional Genomics*, 9(5-6):405-415.
17. Lin, N.* and **Xi, R.** (2010) Aggregated estimating equation estimation, *Statistics and Its Interface*, 4:73-84.
18. Kim, T.M., Luquette, L.J., **Xi, R.** and Park, P.J.* (2010) rSW-seq: algorithm for detection of copy number alterations in deep sequencing data, *BMC Bioinformatics*, 11:423.
19. Li, Q., **Xi, R.*** and Lin, N.* (2010) Bayesian regularized quantile regression, *Bayesian Analysis*, 5:533-556.
20. Lin, N.* and **Xi, R.** (2010) Fast surrogates of U-statistics, *Computational Statistics and Data Analysis*, 54:16-24.
21. **Xi, R.**, Lin, N., and Chen, Y.* (2009) Compression and aggregation for logistic regression analysis in data cubes, *IEEE Transactions on Knowledge and Data Engineering*, 21(4):479-492.

Biological Collaboration Work

1. Song, G.#, Shi, Y.#, Meng, L.#, Ma, J.#, Huang, S., Zhang, J., Wu, Y., Li, J., Yang, S., Rao, D., Cheng, Y., Lin, J., Ji, S., Liu, Y., Shan, J., Wang, X., Shu, Z., Ke, A., Wang, X., Cao, Y., Yuan, J., Jian, Z., Jia, F.*, Zhang, X.*, **Xi, R.***, Gao, Q.* (2022), Single-Cell Transcriptomic Analysis Suggests Two Molecularly Subtypes of Intrahepatic Cholangiocarcinoma, *Nature Communications*, 13, 1642.
2. Song, G.#; Shi, Y.#; Zhang,M.#; Goswami,Sh.#; Afridi, S.,· · ·; Zhang, X.*; **Xi, R.***; Gao, Q.* (2020). Global immune characterization of HBV/HCV-related hepatocellular carcinoma identifies macrophage and T-cell subsets associated with disease progression, *Cell Discovery*, 2020, 6(90).
3. Cui, Y.#, Chen, H.#, **Xi, R.#**, Cui, H.#, Zhao, Y.#, Xu, E.#, · · ·, Zhan, Q.*, Li, Y.* and Liu, Z.* (2020). Whole-genome sequencing of 508 patients identifies key molecular features associated with poor prognosis in esophageal squamous cell carcinoma. *Cell Research*. 30(10):902-913.
4. Yin, S. #, **Xi, R.#**, Wu, A.#, Wang, S.# ,Li, Y., Wang, C., Tang, L., ,Xia, Y., Yang, D., Li, J., Ye, B., Yu, Y., Wang, J., Zhang, H., Ren, F., Zhang, Y., Shen, D., Wang, L., Ying, X., Li, Z., Bu, Z. Ji, Xin, Gao, X. Jia, Y., Jia, Z., Li, N., Li, Z., Ji, J.* and Xi, J.Z.* (2020) Patient-derived tumor-like cell clusters for drug testing in cancer therapy. *Science Translational Medicine*, 12(549), eaaz1723.
5. Corts-Ciriano, I., Lee, J., **Xi, R.**, Jain, D., Jung, Y.L., Yang, L., Gordenin, D., Klimczak, L.J., Zhang, C.Z., Pellman, D.S. and Park, P.J.* (2020) Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. *Nature Genetics*, 52(3):331-341.
6. Dong, L.Q.#, Shi, Y.#, Ma, L.#, Gao, D., Yang, L., Wang, X., Zhang, S., Ding, Z., Zhang, Z., Wang, Z., Liu, L., Zheng, B., Yuan, K., Ke, A., Zhou, J., FAN, J., **Xi, R.*** and Gao, Q.* (2018) Spatial and temporal clonal evolution of intrahepatic cholangiocarcinoma, *Journal of Hepatology*, 69 (1), 89-98.
7. Duan, M., Hao, J., Cui, S., Worthley, D., Zhang, S., Wang, Z., Shi, J., Liu, L., Wang, X., Ke, A., Cao, Y., **Xi. R.**, Zhang, X., Zhou, J., Fan, J.*, Li, C.* and Gao, Q.* (2018) Diverse modes of clonal evolution in HBV-related hepatocellular carcinoma revealed by single-cell genome sequencing, *Cell Research*, 28 (3), 359-373.
8. Chang, J.#, Tan, W.#, Ling, Z.#, **Xi, R.#**, Shao, M., # · · · Mao, W.*, Wu, C.* and Lin, D. (2017) Comprehensive analysis of esophageal squamous-cell carcinoma identifies alcohol drinking-related mutation signature and novel genomic alterations. *Nature Communications*, 8:15290.
9. Gao, Q., Wang, Z., Duan, M., Lin, Y., Zhou, X., Worthley, D. L., · · ·, **Xi R.**, Shi, Y.* and Fan, J.* (2017) Cell culture system for analysis of genetic heterogeneity within hepatocellular carcinomas and response to pharmacologic agents. *Gastroenterology*, 152(1):232-242.
10. Ping, Z., Xia, Y., Shen, T., Parekh, V., Siegal, G.P., Eltoum, I.E., He, J., Chen, D., Deng, M., **Xi, R.**, and Shen, D. * (2016) A microscopic landscape of the invasive breast cancer genome, *Scientific Reports*, 6,27545.
11. The Cancer Genome Atlas Research Network (Including **Xi, R.**) (2015) Comprehensive molecular characterization of cutaneous melanoma, *Cell*, 161(7):1681-1696.
12. Nam, J.Y., Kim, N.K., Kim, S.C., Joung, J.G., **Xi, R.**, Lee, S. Park, P.J.* and Park, W.Y.* (2015) Evaluation of somatic copy number estimation tools for whole-exome sequencing data, *Briefings In Bioinformatics*, bbv055.
13. Shi, J.#, Xing, Q.#, Duan, M.#, Wang, Z., Yang, L., Zhao, Y., Wang, X., Liu, Y., Deng, M., Ding, Z., Ke, A., Zhou, J., Fan, J., Cao, Y., Wang, J.*, **Xi, R.**,*, and Gao, Q.* (2015) Inferring

the progression of multifocal liver cancer from spatial and temporal genomic heterogeneity, *Oncotarget*, 7(3):2867-2877.

14. De Los Angeles, A.#, Ferrari, F.#, **Xi, R.**, Fujiwara, Y., Benvenisty, N., Deng, H., Hochedlinger, K., Jaenisch, R., Lee, S., Leitch, H.G., Lensch, M. W., Lujan, E., Pei, D., Rossant, J., Wernig, M., Park, P.J. and Daley, G.Q.* (2015) Hallmarks of pluripotency, *Nature*, 525(7570):469-478.
15. Kim, T.M., **Xi, R.**, Luquette, L.J, Park, R.W., Johnson, M.D. and Park, P.J.* (2013) Functional genomic analysis of chromosomal aberrations in a compendium of 8000 cancer genomes, *Genome Research*, 23(2):217-227.
16. Cancer Genome Atlas Research Network (including **Xi. R**) (2012) Comprehensive molecular characterization of human colon and rectal Cancer, *Nature*, 487:330-337.
17. Tan, X. , Hu, L., Luquette, L.J., Gao, G., Liu, Y., Qu, H., **Xi, R.**, Lu, Z.J., Park, P.J.* and Elledge, S.J.* (2012) Systematic identification of synergistic drug pairs targeting HIV, *Nature Biotechnology*, 30:1125-1130
18. The Cancer Genome Atlas Research Network (Including **Xi, R.**) (2012) Comprehensive genomic characterization of squamous cell lung cancers, *Nature*, 489:519-525.
19. Kharchenko, P.V., Alekseyenko, A.A., Schwartz, Y.B., Minoda, A., Riddle, N.C., Ernst, J., Sabo, P.J., Larschan, E., Gorchakov, A.A., Gu, T., Linder-Basso, D., Plachetka, A., Shanower, G., Tolstorukov, M.Y., Luquette, L.J., **Xi, R.**, Jung, Y.L., Park, R.W., Bishop, E.P., Canfield, T.P., Sandstrom, R., Thurman, R.E., MacAlpine, D.M., Stamatoyannopoulos, J.A., Kellis, M., Elgin, S.C.R., Kuroda, M.I., Pirrotta, V., Karpen, G.H. and Park, P.J.*(2011) Comprehensive analysis of the chromatin landscape in Drosophila melanogaster, *Nature*, 471(7339):480-485.
20. Kharchenko, P.V., **Xi, R.**,and Park, P.J.* (2011) Evidence for dosage compensation between the X and autosomes in mammals, *Nature Genetics*, 43(12):1167-1169.

COMPUTATIONAL PACKAGES

scRMD: An R-package for single cell RNA-seq imputation.

VirTect: A pipeline for detecting virus integration sites.

BICseqX: A pipeline for germline and somatic copy number detection based on target sequencing.

Difdtl: An R package for differential network analysis based on the penalized D-trace loss.

SVmine: A pipeline for structural variation detection.

Pysim-sv: A pipeline for simulating high throughput sequencing data.

BICseq2: A pipeline for normalizing whole genome sequencing data and detecting germline and somatic copy number variation.

BEQR: An R-package for empirical likelihood-based bayesian quantile regression model selection.

BICseq: An R-package for somatic copy number variation detection for whole genome sequencing.

TEACHING

Graduate Courses

- Mathematical Statistics, 2012/2013 Fall Semester

- Quantile Regression, 2013 Spring Semester
- Biostatistics, 2014/2016 Spring Semester
- Data Mining and Statistical Learning (joint course), 2014 Fall Semester

Undergraduate Courses

- Undergraduate Thesis Seminar, 2013/2014 Spring Semester
- Applied Time Series Analysis, 2014/2015/2016/2017 Fall Semester
- Probability and Statistics B, 2017/2018 Spring Semester

STUDENT SUPERVISED

PhD Students

2011 Yuchao Xia (Joint supervised with Prof. Minghua Deng)

2012 Huili Yuan (Joint supervised with Prof. Minghua Deng)

2013 Yun Liu

2014 Linjie Wu

2015 Chong Chen, Han Wang, Yang Shi (Joint supervised with Prof. Jianzhong Xi)

2017 Zijie Jin, Yifei Xiao

Master Students

2012 Shan Qiao

2013 Haofei Peng, Lei Han

2014 Zhehao He, Xue Long, Songzhi Jiang, Zheng Yang

2015 Yawen Wu, Yuxin Yang, Liu Yang

2016 Chao Gao, Lina Xie, Boya Yin, Nan Xu

FUNDING

Statistical analysis of ambient air pollution, macro-economy and human health (2016), Responsible for RMB ¥430,000 (around US $70,000), National Key Basic Research Program of China, Co-Principal Investigator

Big data driven models and algorithms for management decision (2015), Responsible for RMB ¥320,000 (around US $50,000), National Natural Science Foundation of China, Co-Principal Investigator

Statistical analyses of genomic variations based on high-throughtput sequencing data (2015), RMB ¥600,000 (around US $100,000), National Natural Science Foundation of China, Principal Investigator

Statistical Learning of Unstructured Data: Mathematical Foundation and Algorithms (2015), Responsible for RMB ¥580,000 (around US $100,000), National Key Basic Research Program of China, Co-Principal Investigator

The Recruitment Program of Global Youth Experts of China (2013), RMB ¥2,000,000 (around US $330,000), the Organization Department of the Central Committee of the CPC grants, Principal Investigator

Nonlinear dimension reduction of high dimensional data (2013), RMB ¥30,000 (around US $5,000), Microsoft-PKU Joint Lab Grant, Principal Investigator

Statistical analysis of cancer genomics data (2013), RMB ¥300,000 (around US $50,000), 985 Project, Peking University, Principal Investigator

Agilent 2011 eMerging Insights Grants (2011), US $75,000, Co-Principal Investigator

CONFERENCES

Conference/Session Organization

Biostatistics and computational biology: recent breakthrough in high-throughput biology data analysis, Session organizer, *IMS-China International Conference on Statistics and Probability*, Nanning, China, 2017

Networks and Interactions, Session organizer, *Asian-Pacific Bioinformatics Conference*, Hongkong, China, 2017

Invited Presentations

Tumor Heterogeneity in Hepatocellular Carcinoma and Intrahepatic Cholangiocarcinoma, *The 27th International Korea Genome Organization Annual Conference*, Seoul, South Korea, 2018

Tumor Heterogeneity in Hepatocellular Carcinoma and Intrahepatic Cholangiocarcinoma, *International Conference on Intelligent Biology and Medicine*, Los Angeles, California, USA, 2018

Tumor Heterogeneity in Hepatocellular Carcinoma and Intrahepatic Cholangiocarcinoma , *2018 Joint Statistical Meetings*, Vancouver ,Canada, 2018

Differential network analysis via Lasso penalized D-trace loss, *2017 Annual Conference of Biostatistics of China*, Wuhan, China, 2017

Copy number analysis of whole-genome data using BIC-seq2 and its application to detection of cancer susceptibility variants, *Third ICSA-Canada Chapter Symposium*, Vancouver, Canada, 2017

Differential network analysis via Lasso penalized D-trace loss, *2017 ICSA Applied Statistics Symposium*, Chicago, USA, 2017

Pysim-sv: a package for simulating structural variation data with GC-biases, *Asian-Pacific Bioinformatics Conference*, Hongkong, China, 2017

Copy number analysis of whole-genome data using BIC-seq2 and its application to detection of cancer susceptibility variants, 2016 National Young PI Forum on Genetics and Epigenetics, Suzhou, China, 2016

BIC-seq2 and its application to detection of cancer susceptibility copy number variants, *The 3rd Taihu International Statistics Forum*, Shanghai, China 2016

Intratumor Heterogeneity in Hepatocellular Carcinoma, *The 4th young PI workshop on interdisciplinary research between mathematics, computer science and biology*, Beijing, China, 2016

Statistical Analysis of Big Data, *Annual Meeting of Society of Statistics Canada*, St. Catharines, Canada, 2016

Copy number analysis of whole-genome data using BIC-seq2 and its application to detection of cancer susceptibility variants, *CSHA/AACR Joint Meeting - Big Data, Computation and Systems Biology in Cancer conference*, Suzhou, China, 2015

Statistical Analysis of Big data, *The 12th National Conference of Mathematics*, Beijing, China, 2015

Stochastic Search Variable Selection in Quantile Regression Based on Empirical likelihood, *The 10th National Conference of Probability and Statistics*, Jinan, China, 2014

Copy Number Variation Detection Based on Exome Sequencing Data in Tumor Genomes, *The 3rd Young Bioinformatics PI workshop*, Guangzhou, China, 2014

Detecting Structural Variations and Copy Number Variations Using High-throughput Sequencing Data, *The 4th IMS-China International conference on Statistics and Probability*, Chengdu, China 2013

Empirical likelihood based Bayesian model selection in quantile regression, *Statistics Workshop of Chinese Academy of Sciences*, Beijing, China, 2013

Copy number variation detection with next-generation sequencing data using the Bayesian Information Criterion, *The 5th Chinese conference on Bioinformatics and Systems Biology*, Ha'erbin, China, 2012

Fast Computation of U-statistics: Application in Longitudinal Study, *The 17th ICSA Applied Statistics Symposium*, New York, USA, 2008.

Aggregated Quasi-likelihood Estimation, *Robert Bohrer Workshop for Student Papers in Statistics*, Urbana-Champaign, USA, 2007.

JOURNAL/CONFERENCE REFEREE

Nucleic Acid Research, *Journal of American Statistical Association*, *Bioinformatics*, *Biostatistics*, *PLOS Computational Biology*, *Briefings In Bioinformatics*, *Statistica Sinica*, *ISMB 2012*, *ISMB/ECCB 2013*, *Metrika*, *Statistics and Its Interface*, *Computational Statistics and Data Analysis*,*Intelligent Systems for Molecular Biology (2012)*, *BMC Genomics*, *PLOS One*, *Journal of the Korean Statistical Society*, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*

HONORS

Best Poster Award in Asian Pacific Bioinformatics Conference 2016

The Recruitment Program of Global Youth Experts of China, 2013

Robert Bohrer Award for Student Papers in Statistics, University of Illinois, 2007.

Judith Ross Scholarship in Mathematics & Physics, Washington University in St. Louis, 2007.

First Year Graduate Student Fellowship, Washington University in St. Louis, 2005.

Second Prize in China Undergraduate Mathematical Contest in Modelling, 2000 & 2001.