

Xiaoyu Lu

Department of BioHealth Informatics, School of Informatics and Computing,

Indiana University, Indianapolis, IN, USA

Tel.: 317-6037563 E-mail: lu14@iu.edu

RESEARCH INTERESTS

- Statistical and semi-supervised method development in tissue deconvolution on omics data.
- Machine learning method design to transfer binary data to continuous probability p and predict disease risk ratio using biomedical data
- Information mining in electronic medical records (EMRs) to supporting clinical decision-making

EDUCATION

Indiana University, PhD candidate in bioinformatics, Aug 2017—Now

Shandong University, China, B.S. in Statistics, July 2017.

Shandong University, China, B.S. minor in Finance, July 2017.

HONORS AND AWARDS

- | | |
|-----------|---|
| 05/2019 | IUPUI Graduate Student Travel Fellowship |
| 2015-2017 | Excellent Student Scholarship by Academic Performance, Shandong University |
| 2015-2016 | Silver Awards in Student Research Training Program(SRTP), Shandong University |
| 10/2015 | Golden Award in China Undergraduate Mathematical Contest in Modeling, Shandong Division |
| 2014-2015 | Outstanding Student Cadre, Shandong University |

STRENGTHS

- Well trained in statistics and mathematics. Good understanding of NGS and bioinformatical algorithm.
- Skilled in mathematical model, especially in statistical model of biology problem.
- Can work independently, teamwork player and coordinator
- Excellent in people skill and sports, championship of the university soccer league

PROGRAMMING AND SKILLS

- **Skills:** R language, Python, MATLAB, Microsoft Office
- **Courses:** Mathematical Analysis, Advanced Algebra, Probability Theory, Mathematical Modeling, Mathematical Statistics, Basic Stochastic Processes, Time Series Analysis

MANUSCRIPTS SUBMITTED

1. **Xiaoyu Lu**, Szu-Wei Tu, Wennan Chang, Changlin Wan, Yifan Sun, Baskar Ramdas, Shannon Hawkins, Xin Lu, Reuben Kapur, Xiongbin Lu, Sha Cao and Chi Zhang*. A semi-supervised approach for a robust cell type identification and deconvolution of mouse transcriptomics data. (ISMB 2020)
2. Jiannan Liu, Chuanpeng Dong, Guanglong Jiang, **Xiaoyu Lu**, Yunlong Liu, Huanmei Wu. Transcription factor expression as a predictor of colon cancer prognosis: A machine learning practice. BMC medical genomics
3. Szu-Wei Tu, **Xiaoyu Lu**, Wennan Chang, Sha Cao, Chi Zhang. IWOD: An integrative web-server for optimized tissue data deconvolution. *Nucleic Acids Research* Web Server Issue 2019
4. Wennan Chang, Changlin Wan, **Xiaoyu Lu**, Szu-Wei Tu, Yifan Sun, Xinna Zhang, Yong Zang, Anru Zhang, Kun Huang, Yunlong Liu, Xiongbin Lu, Sha Cao, Chi Zhang. ICTD: Inference of cell types and deconvolution -- a next-generation deconvolution method for accurate assess cell population and activities in tumor microenvironment. BioRxiv (2019) DOI: <https://doi.org/10.1101/426593>. (Submitted, under editorial decision)
5. Wennan Chang, Changlin Wan, **Xiaoyu Lu**, Pengtao Dang, Yue Fang, Yong Zang, Yunlong Liu, Sha Cao and Chi Zhang. Accurate identification of cell type and phenotypic marker genes in single cell transcriptomic data via a data augmentation approach. (RECOMB 2020)

PUBLICATIONS

1. Menghao Huang, Hyeong Geug Kim, Xiaolin Zhong, Chuanpeng Dong, Brian Zhang, Zhigang Fang, Yang Zhang, **Xiaoyu Lu**, Romil Saxena, Yunlong Liu, Chi Zhang, Suthat Liangpunsakul, X Charlie Dong, Sestrin 3 Protects Against Diet- Induced Nonalcoholic Steatohepatitis in Mice Through Suppression of Transforming Growth Factor beta Signal Transduction, *Hepatology*. 2019 Jun 19. <https://doi.org/10.1002/hep.30820>
2. Changlin Wan, Wennan Chang, Yu Zhang, Fenil Shah, **Xiaoyu Lu**, Yong Zang, Anru Zhang, Sha Cao, Melissa L Fishel, Qin Ma, Chi Zhang*, LTMG: a novel statistical modeling of transcriptional expression states in single-cell RNA-Seq data, *Nucleic Acids Research*, Volume 47, Issue 18, 10 October 2019, Page e111, <https://doi.org/10.1093/nar/gkz655>
3. Samuel A Miller, Robert A Policastro, Sudha S Savant, Shruthi Sriramkumar, Ning Ding, **Xiaoyu Lu**, Helai P Mohammad, Sha Cao, Jay H Kalin, Philip A Cole, Gabriel E Zentner, Heather M O'Hagan. Lysine-specific demethylase 1 mediates AKT activity and promotes epithelial-mesenchymal transition in PIK3CA mutant colorectal cancer. *Molecular Cancer Research*. [10.1158/1541-7786.MCR-19-0748](https://doi.org/10.1158/1541-7786.MCR-19-0748)

ABSTRACTS

1. **Xiaoyu Lu**, Szu-wei Tu, Wennan Chang, Yan Huo, Pengcheng Wang, Yu Zhang, Chi Zhang, Sha Cao. A new deconvolution algorithm for accurate assessing immune and stromal cell populations in mouse transcriptomic data. Cancer Research (2019)
2. Changlin Wan, Wennan Chang, **Xiaoyu Lu**, Yifan Sun, Kaman So, Sha Cao, Xiongbin Lu, Chi Zhang. Statistical modeling of transcriptional regulatory states in single-cell RNA-Seq data of tumor and infiltrated immune cells. Cancer Research (2019)

CONFERENCE POSTERS AND ABSTRACTS

1. **Xiaoyu Lu**, Szu-wei Tu, Wennan Chang, Yan Huo, Pengcheng Wang, Yu Zhang, Chi Zhang, Sha Cao. A new deconvolution algorithm for accurate assessing immune and stromal cell populations in mouse transcriptomic data. AACR Annual meeting 2019, Atlanta, USA.
2. Changlin Wan, Wennan Chang, **Xiaoyu Lu**, Yifan Sun, Kaman So, Sha Cao, Xiongbin Lu, Chi Zhang. Statistical modeling of transcriptional regulatory states in single-cell RNA-Seq data of tumor and infiltrated immune cells. AACR Annual meeting 2019, Atlanta, USA.

RESEARCH EXPERIENCES

09/2015 China Undergraduate Mathematical Contest in Modeling (CUMCM) Team Leader

Project: Determined the local date, time, latitude and longitude by analyzing the change of shadows

- Transformed 3D information into 2D information based on the screen coordinate perspective theory
- Built the dynamic model based on the sun's orbit, adjusted model parameter and optimized it

02/2016 Interdisciplinary Contest In Modeling (ICM) Team Leader

Project: Research of Water shortages with mathematical and computing methods and put forward the improvement measures

- Proposed the variable fuzzy evaluation model and used entropy weight method to determine the weight of different indexes according to the actual data. Then described the degree of water shortage in this region
- Applied ARMA model in time series to predict the future circumstances and explained how people's life would be influenced by water resource after analyzing each indexes
- Realized the model using computer software, calculated the index of different interventions and obtained the appropriate intervention

03/2016-09/2016 Consensus for Multi-agent Systems

Advisor: Dr. Fangcui Jiang, Shandong University, Weihai

Overview: Focuses on the consensus problem for high-order multi-agent systems with switching directed topologies

- Introduced an orthogonal linear transformation and Lyapunov-Krasovskii function approach
- Established some sufficient conditions for consensus convergence and the maximum allowable upper bounds of time-varying delays
- The design of protocol parameters was provided based on Lyapunov equation and algebraic inequality
- Guaranteed the solvability of the liner matrix inequalities in the sufficient conditions and the robust consensus with respect to asymmetric time-varying delays for multi-agent systems

08/2017-04/2018 Interpretation of the data in proteomics

Advisor: Dr. Xiaowen Liu, Indiana University

Overview: Design algorithm and develop software tools for the interpretation of top-down mass spectrometry data in proteomics

- Top-down spectral deconvolution: groups top-down spectral peaks into isotopomer envelopes and converts isotopomer envelopes to monoisotopic neutral masses
- Identifies and characterizes proteoforms at the whole proteome level by searching top-down tandem mass spectra against a protein sequence database

05/2018-10/2019 Semi-supervised method in deconvolution of mouse gene expression data

Advisor: Dr. Sha Cao & Dr. Chi Zhang, Indiana University

Overview: Design a semi-supervised method to explain the immune cell type and micro-environment inside of the cancer tissue

- Random walk based method for choosing cell type specific mark gene for each cell type
- Co-expression module to determine the marker based on their own structure
- A R-package has be released via GitHub: <https://github.com/zy26/SSMD>
- An user-friendly web server has be released: <https://ssmd.cccb.iupui.edu/>

11/2019-12/2019 IWOD: An integrative web-server for optimized tissue data deconvolution

Advisor: Dr. Sha Cao & Dr. Chi Zhang, Indiana University

Overview: IWOD is a webserver enables selection, comparison and optimization of methods and marker genes for a tissue omics deconvolution analysis.

- IWOD enables the utilization and comparison of a large set of state-of-the-arts deconvolution methods and formulations, with using a default setting or gene markers refined by their expression pattern in the tissue data or matched/independent single cell data sets
- Analysis of a broad range of omics data types, such as microarray, RNA-seq, ATAC-seq, DNA-methylation and protein assay data.
- Provides different sets of cell type specifically expressed genes to enable the analysis of the blood, central nervous systems, hematopoietic, normal organ, inflammatory, cancer tissues of human and mouse.
- IWOD can identify cell types and marker genes from matched/independent single cell data to assist the tissue data analysis.
- An user-friendly web server has be released: <https://shiny.ph.iu.edu/iwod/>

01/2020-now Redefining cancer metastasis potential using high-throughput sequence data

Advisor: Dr. Sha Cao, Indiana University

Overview: Design a machine learning method to defining the potential of cancer metastasis

- Assume gene expression of cancer tissue follow a mixture model which contains two distributions from metastasis and non-metastasis samples
- Design a machine learning approach to obtain the difference between metastasis and non-metastasis samples
- Define a probability to show the metastasis potential of non-metastasis patients