

Education

Ph.D. in Bioinformatics and Genomics (ongoing), GPA 3.9 / 4.0 The Pennsylvania State University (PSU), The Huck Institutes of Life Science	State College, PA Aug 2019 – May 2024
Master of Science in Biostatistics, GPA 3.7 / 4.0 Washington University in St. Louis (WUSTL), Division of Biostatistics	St. Louis, MO June 2016 – Dec 2017
Bachelor of Science: Biology, GPA 3.5 / 4.0 Wuhan University, College of Life Science	Wuhan, China Sep 2010 – June 2014

Research Experience

Doctoral researcher mentored by Dr. David Koslicki (PSU)	May 2020 – now
Bioinformatic intern at Gilead Sciences	May 2022 – Aug 2022
Bioinformatician supervised by Dr. Bo Zhang (WUSTL)	Jan 2018 – April 2019
Research assistant supervised by Dr. Bo Zhang (WUSTL)	Aug 2016 – Dec 2017

Projects: (*: publication)

- Bioinformatic analysis for clinical virology at Gilead Sciences** (2022.5-2022.8)
 - explored, implemented and benchmarked alignment-based and k-mer-based methods for HBV genotyping.
- Computational methods for metagenomic analysis** (2020-2021, advised by Dr. David Koslicki):
 - proved, implemented and benchmarked a truncation-based containment MinHash algorithm **CMash***, which can efficiently estimate multi-resolution Jaccard and containment indices, in metagenomic analysis.
 - developed, implemented and analyzed metagenomic to explore the “microbial dark matter”.
- The National Center for Advancing Translational Sciences (NCATS)** (2021.1-2021.7, supervised by Dr. David Koslicki):
 - worked in the **ARAX*** team that aims at providing a graph-based reasoning tool to explore biomedical questions;
 - worked in the Translator API team that aims at normalizing query communications between the user and the knowledge database.
- Methodology improvement for ATAC-seq analysis** (2018-2019, with Dr. Bo Zhang and Dr. Ting Wang):
 - developed **AIAP***, an advanced ATAC-seq analysis pipeline for comprehensive analysis of ATAC-seq dataset;
 - incorporated statistical ideas from DEG and genomic footprint into the pipeline to discover differentially opened DNA regions and insertion free DNA regions as downstream analysis;
 - explored rodent-specific TE rewired gene regulatory network from ENCODE ATAC-seq data;
- TaARGET II epigenome data analysis** (2018-2019, with Dr. Bo Zhang and Dr. Ting Wang):
 - worked in Data Coordination Center (DCC) of TaARGET II Consortium, streamlined and setup the routine pipelines for RNA-seq, ChIP-seq, WGBS, and ATAC-seq data; and processed part of data with preliminary quality analysis;
 - implemented pipelines into Docker and Singularity;
 - explored the empirical distribution of key quality factors and provided schemes for data quality estimate.
- Various bioinformatic collaborations** (2016-2019, with Dr. Bo Zhang):
 - performed single-cell RNA analysis by Seurat and scdiff / SC3 for trajectory prediction;
 - profiled epigenetic signature changes under different conditions based on ATAC-seq and RNA-seq data;
 - analyzed ChIP-seq data of NKX2, Twist1 transcription factors in human and mouse samples;

7. **Synthetic biology: biobricks for tandem promoter** (2013-2014, IGEM project):
- created biobricks with tandem promoters combined with specific guide-RNA targets
 - utilized GFP to validate the multi-stage promoter function triggered by dCas9+Transcription factor protein complex

Publications

- Amy K. Glen, Chunyu Ma, Luis Mendoza, Finn Womack, E. C. Wood, Meghamala Sinha, Liliana Acevedo, Lindsey G. Kvarfordt, Ross C. Peene, **Shaopeng Liu**, Andrew S. Hoffman, Jared C. Roach, Eric W. Deutsch, Stephen A. Ramsey, David Koslicki. "ARAX: a graph-based modular reasoning tool for translational biomedicine." *bioRxiv* (2022).
- Shaopeng Liu**, David Koslicki. "CMash: fast, multi-resolution estimation of k-mer-based Jaccard and containment indices." *Bioinformatics* 38.Supplement_1 (2022): i28-i35.
- Shaopeng Liu**, Daofeng Li, Cheny Lyu, Paul Gontarz, Ting Wang, Bo Zhang. "AIAP: A Quality Control and Integrative Analysis Package to Improve ATAC-seq Data Analysis." *Genomics, Proteomics & Bioinformatics*, 2021
- Matthew J McCoy, Kitra Cates, Yangjian Liu, Daniel G Abernathy, Bo Zhang, **Shaopeng Liu**, Paul Gontarz, Woo Kyung Kim, Shawei Chen, Wenjun Kong, Harrison W Gabel, Samantha A Morris, Andrew Yoo. "Deconstructing Stepwise Fate Conversion of Human Fibroblasts to Neurons by MicroRNAs." *CELL STEM CELL*, 2021
- Paul Gontarz, Shuhua Fu, Xiaoyun Xing, **Shaopeng Liu**, Benpeng Miao, Viktoriia Bazylianska, Akhil Sharma, Pamela Madden, Kitra Cates, Andrew Yoo, Anna Moszczynska, Ting Wang, Bo Zhang. "Comparison of differential accessibility analysis strategies for ATAC-seq data." *Scientific reports* 10.1 (2020): 1-13.
- Jennifer L David, Linda Cox, Christine Shao, Cheng Lyu, **Shaopeng Liu**, Rajeev Aurora, Deborah J Veis. "Conditional Activation of NF- κ B Inducing Kinase (NIK) in the Osteolineage Enhances Both Basal and Loading-Induced Bone Formation." *Journal of Bone and Mineral Research*, 2019
- Yanchun Pan, Ying Zhu, Wei Yang, Eric Tycksen, **Shaopeng Liu**, Linjian Zhu, Mukesh K Sharma, Albert H Kim, Bo Zhang, Hiroko Yano. "The role of Twist1 in mutant huntingtin-induced transcriptional alterations and neurotoxicity." *Journal of Biological Chemistry*, 2018
- Kesavan Meganathan, Emily MA Lewis, Paul Gontarz, **Shaopeng Liu**, Ed Stanley, Andrew G Elefanty, James E. Huettnier, Bo Zhang, Kristen L Kroll. "Regulatory networks specifying cortical interneurons from human embryonic stem cells reveal roles for CHD2 in interneuron development." *Proceedings of the National Academy of Sciences*, 2017
- Hangxing Jia, Tong Liang, Zhaoning Wang, Zhaoren He, Yang Liu, Lei Yang, Yan Zeng, **Shaopeng Liu**, Linyi Tang, Jianbo Wang, Yu Chen, and Zhixiong Xie. "Multistage Regulator Based on Tandem Promoters and CRISPR/Cas." *ACS Synthetic Biology*, 2014

Skills

- Programming: Shell, Python, R, Rust, Git, and SQL
- Domain of knowledge: Bioinformatics, Statistics, Algorithm, and Data mining

Service

- TA in class "Statistical Computing with SAS", 2017
- Student committee, annual Bioinformatics and Genomics Retreat in Huck Institute, 2021
- Subreviewer, ISMB 2022

Awards

- The J. Lloyd Huck Graduate Fellowship; Pennsylvania State University 2019-2020

Shaopeng Liu
(314) 585-2220 | sml6467@psu.edu
State College, PA

- | | |
|---|-----------|
| • The Graham Endowed Fellowship; Pennsylvania State University | 2019-2021 |
| • The Braddock Scholarship; Pennsylvania State University | 2019-2020 |
| • WU Scholarship and Grants; Washington University in St. Louis | 2016-2017 |
| • Golden medal; International Genetically Engineered Machine (IGEM) | 2013-2014 |
| • College Scholarship; Wuhan University | 2010-2012 |

Professional References

- **David Koslicki, Ph.D.**

Associate Professor of Computer Science and Engineering

The Pennsylvania State University
W205C Westgate, State College, PA 16802
Phone: 814-865-1611
Email: dmk333@psu.edu

- **Bo Zhang, Ph.D.**

Associate Professor of Developmental Biology

Washington University in St. Louis
Suite 2100, 4480 Clayton Avenue, St. Louis, MO 63110
Phone: 314-362-4757
Email: bzhang29@wustl.edu