

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 – Jun 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
Computational Biologist intern at 23andMe	June 2023 – Aug 2023
Bioinformatician 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: (*: published work)

- Bioinformatic analysis for single-cell datasets at 23andMe** (2023.6-2023.8)
 - Leveraged Cell Ranger and Seurat (R) to develop internal pipelines for multiome single-cell analysis.
- Bioinformatic analysis for clinical virology at Gilead Sciences** (2022.5-2022.8)
 - Developed novel k-mer-based algorithms for HBV genotyping of clinical samples.
- Computational methods for metagenomic analysis** (2020-now, advised by Dr. David Koslicki):
 - Proved, implemented and benchmarked a truncation-based containment MinHash algorithm **CMash***, which can simultaneously estimate multi-resolution Jaccard and containment indices, in metagenomic analysis.
 - Developed a **FracMinHash-based pipeline*** to perform alignment-free metagenomic functional profile.
 - Demonstrated the similarities between **synmers** and the **FracMinHash algorithm*** and show their efficacy and advantage in metagenomic applications.
 - Designing and implementing **multi-resolutions synmers*** that can fit a range of k values.
- Metagenomic data mining with knowledge graphs** (2022.10-2024.3, advised by Dr. David Koslicki):
 - constructed a **metagenomic-specific knowledge graph*** (MKG) by integrating public resources, addressing conflicts and fitting a Biolink data model.
 - Leveraged MKG for metagenomic applications, including graph embedding for metagenomic samples, visualization of biological connections, and pathogen predictions (a graph completion problem).
- The National Center for Advancing Translational Sciences (NCATS)** (2021.1-2024.3, supervised by Dr. David Koslicki):
 - worked in the **ARAX*** team and in the **Biomedical Data Translator Consortium*** aiming at providing a graph-based reasoning tool to explore biomedical questions;
 - worked in the Translator API team to normalize query communications between the user and the knowledge database.
- Methodology improvement for ATAC-seq analysis** (2018-2019, advised by 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;

- c) explored rodent-specific TE rewired gene regulatory network from ENCODE ATAC-seq data;
- 7. **TaRGET II epigenome data analysis** (2018-2019, advised by Dr. Bo Zhang and Dr. Ting Wang):
 - a) worked in Data Coordination Center (DCC) of TaRGET 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;
 - b) implemented pipelines into Docker and Singularity;
 - c) explored the empirical distribution of key quality factors and provided schemes for data quality estimate.
- 8. **Various bioinformatic collaborations** (2016-2019, supervised by Dr. Bo Zhang):
 - a) performed single-cell RNA analysis by Seurat and scdiff / SC3 for cell trajectory prediction;
 - b) profiled epigenetic signature changes under different conditions based on ATAC-seq and RNA-seq data;
 - c) analyzed ChIP-seq data of NKX2, Twist1 transcription factors in human and mouse samples;
- 9. **Synthetic biology: biobricks for tandem promoter** (2013-2014, IGEN project):
 - a) created biobricks with tandem promoters combined with specific guide-RNA targets
 - b) utilized GFP to validate the multi-stage promoter function triggered by dCas9+Transcription factor protein complex

Publications (*: first or co-first authors)

- **Shaopeng Liu***, Judith S. Rodriguez*, Viorel Munteanu*, ... , Mihai Pop, David Koslicki, Serghei Mangul. "Primer: Analysis of human metagenomic data." (Under review)
- Chunyu Ma*, **Shaopeng Liu***, David Koslicki. "MetagenomicKG: a knowledge graph for metagenomic applications." (Under revision)
- Mahmudur Rahman Hera*, **Shaopeng Liu***, Judith S. Rodriguez, Wei Wei Chunyu Ma, and David Koslicki. "Metagenomic functional profiling: to sketch or not to sketch?" (*European Conference on Computational Biology* 2024, in press)
- **Shaopeng Liu**, David Koslicki. "Connecting Syncmers to FracMinHash: similarities and advantages" (Under revision)
- Karamarie Fecho, Anne E Thessen, ... , **Biomedical Data Translator Consortium**. "Progress toward a universal biomedical data translator." *Clinical and Translational Science*, 2022, 15(8): 1838-1847.
- Deepak R. Unni, Sierra A. T. Moxon, ... , **Biomedical Data Translator Consortium**. "Biolink Model: A universal schema for knowledge graphs in clinical, biomedical, and translational science." *Clinical and Translational Science*, 2022.
- 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." *Bioinformatics* (2023).
- **Shaopeng Liu**, David Koslicki. "CMash: fast, multi-resolution estimation of k-mer-based Jaccard and containment indices." (ISMB 2022) *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

Shaopeng Liu

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

- 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
- Catherine E. Lipovsky, Jesus Jimenez, ... , Shaopeng Liu, Bo Zhang, Stacey L. Rentschler. "Chamber-specific transcriptional responses in atrial fibrillation." *JCI insight* 5.18 (2020).
- 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. Huettner , 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: Python, R, Linux (Bash), SQL
- Domain of knowledge: Bioinformatics, Statistics, Algorithm, Knowledge Graph, Data Mining, Machine Learning
- Project management: Git, Docker, Singularity, Snakemake

Service

- TA in class "Statistical Computing with SAS", 2017
- Penn State University: student committee to organize annual Bioinformatics and Genomics Retreat in Huck Institute in 2021; host code reproducibility bootcamp in 2023.
- Sub reviewer in ISMB, BIBM and RECOMB

Awards

- | | |
|--|-----------|
| The J. Lloyd Huck Graduate Fellowship; Pennsylvania State University | 2019-2020 |
| 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.** (Relationship: advisor)
Associate Professor of Computer Science and Engineering
The Pennsylvania State University
W205C Westgate, State College, PA 16802
Phone: 814-865-1611
Email: dm333@psu.edu

- **Bo Zhang, Ph.D.** (Relationship: supervisor)
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

- **Erika Ganda, Ph.D.** (Relationship: advisor)
Assistant Professor of Food Animal Microbiomes
The Pennsylvania State University
308 Animal, Veterinary and Biomedical Sciences Building, State College, PA 16802
Phone: 814-865-4084
Email: ganda@psu.edu

- **Qunhua Li, Ph.D.** (Relationship: advisor)
Associate Professor of Statistics
The Pennsylvania State University
514A Wartik Lab, State College, PA 16802
Phone: 814-865-9468
Email: qul12@psu.edu

- **Paul Medvedev, Ph.D.** (Relationship: advisor)
Professor of Computer Science and Engineering, and Biochemistry and Molecular Biology
506B Wartik Lab, State College, PA 16802
Phone: 814-865-2733
Email: pzm11@psu.edu