

## Education

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

## Research Experience

<b>Doctoral researcher mentored by <a href="#">Dr. David Koslicki</a> (PSU)</b>	May 2020 – now
<b>Computational Biology intern at 23andMe</b>	June 2023 – Aug 2023
<b>Bioinformatic intern at Gilead Sciences</b>	May 2022 – Aug 2022
<b>Bioinformatician supervised by <a href="#">Dr. Bo Zhang</a> (WUSTL)</b>	Jan 2018 – April 2019
<b>Research assistant supervised by <a href="#">Dr. Bo Zhang</a> (WUSTL)</b>	Aug 2016 – Dec 2017

### Projects: (\*: publication-related)

- Bioinformatic analysis for single-cell datasets at 23andMe** (2023.6-2023.8)
  - benchmarked public analysis tools and developed internal pipeline for multiome single-cell analysis
- Bioinformatic analysis for clinical virology at Gilead Sciences** (2022.5-2022.8)
  - developed 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 efficiently estimate multi-resolution Jaccard and containment indices, in metagenomic analysis.
  - utilized k-mer-sketching-based methods to develop **comprehensive pipelines\*** to explore the "microbial dark matter";
  - extended the **k-mer truncation idea\*** to flexible seedings in other k-mer sampling methods and read alignment tools.
  - constructed a **metagenomic-specific knowledge graph\*** by integrating public resources and adopted graph learning methods to for data mining purposes in metagenomics
- The National Center for Advancing Translational Sciences (NCATS)** (2021.1-now, 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, 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;
- TaRGET II epigenome data analysis** (2018-2019, with 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.
7. **Various bioinformatic collaborations** (2016-2019, with Dr. Bo Zhang):
- a) performed single-cell RNA analysis by Seurat and scdiff / SC3 for 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;
8. **Synthetic biology: biobricks for tandem promoter** (2013-2014, IGEM 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**

---

- Mahmudur Rahman Hera\*, **Shaopeng Liu\***, Judith S. Rodriguez, Wei Wei Chunyu Ma, and David Koslicki. "Fast, lightweight, and accurate metagenomic functional profiling using FracMinHash sketches" (*Submitted*)
- **Shaopeng Liu**, David Koslicki. "Connecting Syncmers to FracMinHash: similarities and advantages" (*Submitted*)
- Fecho K, Thessen A E, Baranzini S E, et al. Progress toward a universal biomedical data translator[J]. *Clinical and Translational Science*, 2022, 15(8): 1838-1847.
- Unni D R, Moxon S A T, Bada M, et al. Biolink Model: A universal schema for knowledge graphs in clinical, biomedical, and translational science[J]. *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
- 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- $\kappa$ 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

# Shaopeng Liu

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

- 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, SQL, Rust
- Domain of knowledge: Bioinformatics, Statistics, Algorithm, Knowledge Graph, 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 |
| • 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: dm333@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