



# Shaopeng Liu

Ph.D. candidate  
Bioinformatics



sml6467@psu.edu



+1 3145852220



State College, PA, 16803



PDF: Curriculum vitae

## Keywords

1. Bioinformatician
2. Genomic data scientist
3. Metagenomics
4. Biomedical knowledge graphs

## Skills

Python

Shell

Genomics & NGS

Machine Learning

Statistics

Algorithm

Docker & Git

SQL

## Summary

Passionate bioinformatician with 5+ years experience analyzing genomic datasets and computational tasks. Developed and implemented comprehensive pipelines and novel algorithms. Currently working on sketching-based algorithms and graph learning methods for metagenomic samples. Seeking to leverage efficient algorithms, multi-omic data, artificial intelligence, and biomedical knowledge graphs to address health-related issues.

## Education

2019-2024 Ph.D. in Bioinformatics & Genomics

Pennsylvania State University

2016-2017 M.Sc. in Biostatistics

Washington University in St. Louis

2010-2014 B.Sc. in Biology

Wuhan University

## Experience

### 1. Bioinformatic intern (2022.5-8, Gilead Sciences)

- developed k-mer-based algorithms for HBV genotyping of clinical samples.

### 2. Doctoral researcher in computational biology (2020-2024, PSU)

- proved, implemented, and benchmarked a truncation-based containment Min-Hash algorithm CMash for multi-resolution estimation of Jaccard and containment indices in metagenomic analysis;
- worked in the Biomedical Data Translator Consortium that aims at providing a graph-based reasoning tool for biomedical and translational studies;
- utilized k-mer-sketching-based methods to develop comprehensive pipelines to explore the "microbial dark matter";
- besides the sketching methods, adopted graph learning methods to analyze microbial samples for pathogen detection.

### 3. Bioinformatician (2017-2019, WUSTL)

- collaborated with various groups and performed bioinformatic analysis on genomic and epigenomic datasets, including ChIP-seq, RNA-seq, WGBS, ATAC-seq, and single-cell RNA;
- worked in the 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;
- developed AIAP, an advanced ATAC-seq analysis pipeline for comprehensive analysis of ATAC-seq dataset; and also implemented the pipeline into Docker and Singularity

## Publication

- 2022 Biolink Model: A universal schema for knowledge graphs in clinical, biomedical, and translational science[J]. *Clinical and Translational Science*
- 2022 Progress toward a universal biomedical data translator[J]. *Clinical and Translational Science*
- 2022 ARAX: a graph-based modular reasoning tool for translational biomedicine. *bioRxiv*
- 2022 \*CMash: fast, multi-resolution estimation of k-mer-based Jaccard and containment indices. *Bioinformatics*
- 2021 \*AIAP: A Quality Control and Integrative Analysis Package to Improve ATAC-seq Data Analysis. *Genomics, Proteomics & Bioinformatics*

\*: first author



# Shaopeng Liu

Ph.D. candidate  
Bioinformatics



sml6467@psu.edu



+1 3145852220



State College, PA, 16803

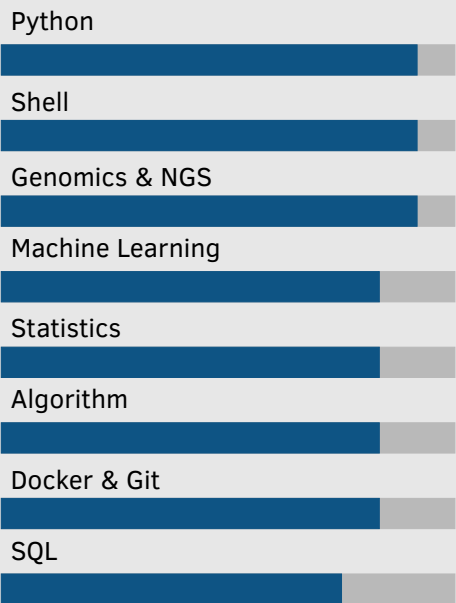


PDF: Curriculum vitae

## Keywords

- 1. Bioinformatician
- 2. Genomic data scientist
- 3. Metagenomics
- 4. Biomedical knowledge graphs

## Skills



2021 Deconstructing Stepwise Fate Conversion of Human Fibroblasts to Neurons by MicroRNAs. *CELL STEM CELL*

2020 Comparison of differential accessibility analysis strategies for ATAC-seq data. *Scientific reports*

2019 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*

2018 The role of Twist1 in mutant huntingtin-induced transcriptional alterations and neurotoxicity. *Journal of Biological Chemistry*

2017 Regulatory networks specifying cortical interneurons from human embryonic stem cells reveal roles for CHD2 in interneuron development. *Proceedings of the National Academy of Sciences*

2014 Multistage Regulator Based on Tandem Promoters and CRISPR/Cas. *ACS Synthetic Biology*

## Awards

2019-2021	The Graham Endowed Fellowship	Pennsylvania State University
2019	The J. Lloyd Huck Graduate Fellowship	Pennsylvania State University
2019	The Braddock Scholarship	Pennsylvania State University
2016	WU Scholarship and Grants	Washington University in St. Louis
2013	Golden medal	International Genetically Engineered Machine (IGEM)
2010-2012	College Scholarship	Wuhan University