Shaopeng Liu

Ph.D. candidate
Bioinformatics

sml6467@psu.edu

+1 3145852220

State College, PA, 16803

PDF: Curriculum vitae

Keywords

1. Bioinformatics
2. Metagenomics
3. K-mer-based algorithms
4. Genomic data science
5. Biomedical knowledge graphs

Skills

- Python
- Shell
- Genomics & NGS
- Algorithm
- Statistics
- Data mining
- Docker & Git
- SQL

Summary

Dynamic and innovative bioinformatician with a Ph.D. candidacy in Bioinformatics and 6+ years of experience in genomics and computational biology. Specializes in the development of advanced algorithms for genomic data analysis, focusing particularly on metagenomics and multi-omics integration.

Proven expertise in developing k-mer sketching techniques for precise metagenomic analysis, complemented by a strong proficiency in employing advanced bioinformatics tools. Demonstrated skill in integrating and utilizing biomedical knowledge graphs with AI techniques to significantly advance drug discovery and offer innovative hypotheses for complex biological interactions. Committed to contributing to groundbreaking research and development with a passion for advancing healthcare solutions through data-driven insights.

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

Computational Biology intern (2023.6-8, 23andMe)

- Created pipelines for the analysis and benchmarking of multiome single-cell datasets.

Bioinformatic intern (2022.5-8, Gilead Sciences)

- Developed innovative k-mer-based algorithms for HBV genotyping in clinical samples.

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

- Proved, implemented, and benchmarked a truncation-based containment MinHash algorithm **CMash** for multi-resolution estimation of Jaccard and containment indices in metagenomic analysis.
- Utilized k-mer-sketching-based methods to develop comprehensive pipelines to explore the "microbial dark matter".
- Actively participated in the Biomedical Data Translator Consortium, emphasizing collaborative efforts in developing a graph-based reasoning tool for biomedical and translational studies.
- Constructed a metagenomic-specific knowledge graph **MKG** by integrating public resources and adopted graph learning methods for data mining purposes (such as pathogen prediction) in metagenomics.

Bioinformatician (2017-2019, WUSTL)

- Engaged in collaborative efforts with diverse teams and conducted bioinformatic analyses on a spectrum of genomic and epigenomic datasets, including ChIP-seq, RNA-seq, WGBS, ATAC-seq, and single-cell RNA data.
- Worked in the Data Coordination Center (DCC) of TaRGET II Consortium, streamlined and set 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.

Shaopeng Liu

Ph.D. candidate
Bioinformatics

sml6467@psu.edu

+1 3145852220

State College, PA, 16803

PDF: Curriculum vitae

Keywords

1. Bioinformatics
2. Metagenomics
3. K-mer-based algorithms
4. Genomic data science
5. Biomedical knowledge graphs

Skills

- Python
- Shell
- Genomics & NGS
- Algorithm
- Statistics
- Data mining
- Docker & Git
- SQL

Leadership

- **Team Collaboration:** Demonstrated excellence in working collaboratively within multidisciplinary teams, contributing to a positive and productive work environment in both academic and industry settings.
- **Mentorship:** Actively involved in mentoring undergraduate students and junior colleagues, providing guidance and support in their academic and professional development.
- **Communication:** Expertly coordinated and spearheaded a series of department-level workshops and conferences; regularly presented at academic conferences, effectively communicating complex research findings to a broad scientific audience.
- **Interdisciplinary Project Leadership:** Led and participated in cross-functional projects, effectively facilitating communication between different research teams and integrating diverse viewpoints and skill sets to achieve project goals.
- **Problem-Solving:** Demonstrated a consistent ability to identify and resolve complex challenges in bioinformatics and computational biology. Excelled in devising innovative solutions to technical and research-related problems.

Publications

*: first or co-first author

2024 *Primer: Analysis of human metagenomic data *(In-progress)*

2024 *MKG: a microbial knowledge graph for metagenomic data mining *(In-progress)*

2023 *Fast, lightweight, and accurate metagenomic functional profiling using FracMinHash sketches. *bioRxiv*

2023 *Connecting Syncmers to FracMinHash: similarities and advantages *bioRxiv*

2023 ARAX: a graph-based modular reasoning tool for translational biomedicine. *Bioinformatics*

2022 Biolink Model: A universal schema for knowledge graphs in clinical, biomedical, and translational science. *Clinical and Translational Science*

2022 Progress toward a universal biomedical data translator. *Clinical and Translational Science*

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*

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-κB 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*