Lihe (Leo) Liu

📱 352-328-2718 | ✉ lihe.liu.95@gmail.com | ⌨ liulihe954 | 💼 liulihe | 🎓 Google Scholar

EDUCATION

UW-Madison | Madison, WI | *Ph.D. in Animal Sciences (**computational biology** track)* | 08/2018 - **08/2023 (expected)**

UW-Madison | Madison, WI | *M.S. in Computer Sciences* | 08/2021 - 05/2023

Huazhong Agricultural University | Wuhan, China | *B.S. in Animal Sciences* | 08/2014 - 07/2018

EXPERIENCE

Graduate Research Assistant, UW-Madison   Madison, WI

Computational Biology   *08/2018 - Present*

- ***Technical: FastQC/MultiQC, STAR, Bowtie2, BWA, HiSAT2, SAMtools, RSEM, Bismark, HTSeq, bcftools, snpEff** | cloud computing, Nextflow, Docker, Conda, Slurm, python/R/Bash | R package development*
- Developed, maintained and improved internal pipelines for **NGS data processing & analysis** (bulk RNA-Seq & WGB-Seq) in a **high-performance computing cluster** (Linux), significantly achieved effective automation and high reproducibility
- Developed, tested and documented novel software (**R package**) to facilitate genomics research in common animal species
- Collaborated across multiple teams of biologists, independently planned, analyzed bench-generated data and and managed bioinformatics outputs; wrote scientific manuscripts; communicated scientific results at international academic conferences

Bioinformatics Software Developer, UW-Madison   Madison, WI

Web Application | ***demo**: http://enrichkit.info/* | ***code repo**: https://github.com/liulihe954/EnrichKitWeb*   *08/2022 - 12/2022*

- ***Technical: full stack development, data manipulation (Pandas, NumPy and Requests), database design and development***
- Parsed 10M rows of public data (e.g., *Ensembl, NCBI, KEGG, AnnotationHub, MSigDB, etc.*) using **Python**, optimized **MySQL** schema (e.g., smaller data types, persistent computed columns, indexing), improved data storage and query efficiency
- Independently designed and developed a **Django (python)** application for **gene ID conversion**, **genomic position matching** and **gene-set enrichment** analysis, significantly improved data interpretation accuracy and speed for common animal species
- Deployed app to cloud using **AWS stack** (EC2, RDS and S3), implemented asynchronous & parallel tasks using **Celery** and **Redis**

ACADEMIC PROJECTS

Beef cattle fetal programming: effects of maternal diet on offspring epigenome and transcriptome   *2021 - 2022*

- ***Technical: methylKit, WGCNA, DEXSeq, DESeq2, gene-set enrichment analysis, ggplot2, Matplotlib***
- Identified differentially methylated CpG sites across the whole genome using **logistic regression** from WGB-Seq data
- Constructed weighted correlation co-expression networks using **hierarchical clustering** and **PCA**, accessed module preservation statistics using **permutation testing**, performed **enrichment** (over-representation) analysis using **Fisher's exact test**
- Performed differential **isoforms expression** and **exon usage** analyses, identified candidates with divergent splicing patterns

Bioinformatics course project (representative collection) | https://github.com/liulihe954/BioinformaticsProjects   *2021*

- ***Technical: OOP in python, data structure, algorithms analysis, scRNA-seq analysis, network inference, machine learning***
- Implemented a set of example algorithms: **Hamiltonian graph** for sequence assembly; **dynamic programming** for pairwise alignment, **expectation-maximization** algorithm for *de novo* motif discovery (OOPS); **hidden markov model** (Baum–Welch algorithm) for exon-intron finding; network construction using **mutual information** from gene expression data
- Performed **dimension reduction**, **clustering**, cell type **annotation**, **differential expression** analysis, **trajectory inference** for sample scRNA-Seq data using *scanpy, singleR, Seurat*
- Surveyed a set of **probabilistic graphical models** for gene regulatory network learning, e.g., Bayesian network (sparse candidate), Gaussian graphical models, dependency network (GENIE3), etc.

TECHNICAL SKILLS

Bioinformatics & Data analysis: RNA-Seq, WGB-Seq, Bioconductor, Tidyverse, Pandas/NumPy/SciPy, scikit-learn, TensorFlow

Programming Languages: Python, Java, R, Shell (Bash, Zsh, SFTP/FTP), SQL (MySQL), C++, JavaScript/HTML/CSS, MATLAB

Frameworks: Django, Spring Boot, ReactJS, NodeJS, JUnit, CUDA, OpenMP

Developer Tools: Git, Docker, Conda, GCP, AWS (EC2, RDS, S3), Elasticsearch, Jira

PUBLICATION & AWARDS

- **1**. Amorín[*], **Liu, L**[*], *et al*. **Maternal diet induces persistent DNA methylation changes in the muscle of beef calves.** *Sci Rep*. 13, 1587 (2023). doi: 10.1038/s41598-023-28896-3
- **2**. **Liu, L**. *et al*. **Maternal methionine supplementation during gestation alters alternative splicing and DNA methylation in bovine skeletal muscle.** *BMC Genom*. 22, 780 (2021). doi: 10.1186/s12864-021-08065-4
- **3**. **Liu, L**. *et al*. **Differential network analysis of bovine muscle reveals changes in gene coexpression patterns in response to changes in maternal nutrition.** *BMC Genom*. **21**, 684 (2020). doi: 10.1186/s12864-020-07068-x
- ***Neal A. Jorgensen Genome Travel Awards***, Plant & Animal Genome Conference XXIX (San Diego, CA), 2022
- ***Study Abroad Scholarship*** | China Scholarship Council (CSC), 2017