

# Saidi Wang

Data Integration and Knowledge Discovery Lab, University of Central Florida

Mobile:407-675-0719|E-mail: tjwangsaidi@gmail.com

## RESEARCH INTEREST

---

- **Data Mining:** Develop novel data mining methods to analysis biological data such as sequences, gene expressions data.
- **Machine Learning Algorithms:** Develop statistical and computational methods for data integration, pattern mining and knowledge discovery from massive genomic datasets.
- **Computational biology:**
  - **Gene Regulation:** Develop methods to computationally predict TF binding interactions through binding site identification at the whole genome scale.
  - **Enhancer-Promoter Interaction Prediction:** Design pipelines to predict enhancer-promoter interactions.
  - **Metagenomics:** Develop tools to reconstruct bacterial strain genomes from metagenomic data.

## EDUCATION

---

**Ph.D., Computer Science**

Fall 2017 – Summer 2022

*University of Central Florida, USA*

**M.S., Software Engineering**

Fall 2014 – Spring 2017

*Tongji University, China*

**B.S., Electronic and Information Engineering**

Fall 2010 – Spring 2014

*Tianjin Polytechnic University, China*

## ACADEMIC POSITION

---

### Graduate Research Assistant

Fall 2017 - now

*Data Integration and Knowledge Discovery Lab, University of Central Florida, USA*

### Graduate Research Assistant

Fall 2014 – Spring 2017

*Department of Software Engineering, Tongji University, China*

## TEACHING EXPERIENCE

---

- Mentoring high school students with bioinformatics tools (Yuxiang Zhang, TRINITY PREPARATORY SCHOOL). Fall 2021
- Mentoring undergraduate students with bioinformatics project (Lan Wallace and Fabian Sanchez, UCF). Summer and Fall 2021
- Helping to teach PCB 5596(Biomedical Informatics: Sequence Analysis) about the Python Tutorial part. Fall 2021
- Teaching assistant of Embedded system design (homework grading and teaching). Fall 2014

## TECHNICAL SKILLS

---

- Professional skills
  - Proficient in Python, R, C, C++; Familiar with JAVA、WEB Development (JS、HTML、CSS)、SSH frameworks.
  - Proficient in Linux development environment.
  - Familiar with various types of Databases, such as Oracle, MySQL, SQLServer.
  - Familiar with PyTorch, NumPy, pandas, scikit-learn.
- Computer skills
  - Microsoft Office products; LabVIEW; MATLAB; Assembly Language.

## HONORS AND ACHIEVEMENTS

---

Full Award from University of Central Florida.	Fall 2017 - now
Graduate Research Award, University of Central Florida.	Fall 2017 - now
Graduate Research Scholarship, Tongji University.	Fall 2014 – Spring 2017

## ACADEMIC SERVICE

---

- Journal review for nature scientific reports

## SELECTED PROJECTS

---

### Projects in Data Integration and Knowledge Discovery Lab, University of Central Florida

--Fall 2017 - now

- **SMS Project:** Developed a novel computational approach called SMS to reconstruct bacterial Strain genomes in Multiple Samples.
- **Enhancer-promoter project:** Designed a pipeline to study the motif pairs in enhancer-promoter interactions.
- **RPG project:** Designed a pipeline to identify the potential distal regulatory regions of human ribosomal protein genes.
- **miRNA TSS project:** Analysis of miRNA TSS annotations from multiple representative studies and compiled a robust set of miRNA TSSs.
- **Bacterial strain tools project:** Investigated twenty-one computational tools that attempt to infer bacterial strain genomes from shotgun reads.
- **CCI project:** Investigated and evaluated nine computational tools that attempt to infer Ligand-Receptor interactions to decipher cell-cell communication.

### Projects in Department of Software Engineering, Tongji University

Fall 2014 – Spring 2017

- **Research on the adaptability of unfamiliar individuals in the human body's specific behavior detection algorithm**  
Spring 2016 – Spring 2017
  - Built data measurement module: connected acceleration transducer of MSP 430 with a 51 Chip and added an acceleration transducer onto it to collect data.

- Process data based on machine learning: Using neural network algorithm to train human body data and classify human behaviors.
- Recognizing different human behaviors based on Big Data Mining: By using time series similarity searching method to recognize human behaviors in a big database filled with collected behavior data.
- **The design and implementation of the system of fall detection and alarming device for old people** --Fall 2015 – Spring 2016
  - Design and implementation of a fall detection device more intelligently.
  - Realization of fall detection algorithm based on machine learning.
  - Using Decision Tree to classify the human behavior features.
  - Recognizing human behavior based on KNN.
- **Monitoring Network Traffic Abnormity Based on Big Data Mining** Fall 2016 – Spring 2016
  - Constructed SQL Query's syntax tree and make it meet specific requirements.
  - Parsed syntax tree to lay the foundation for further data analysis.
- **Basic theory and method of error tolerance calculation** Fall 2014 – Summer 2015
  - The research is mainly engaged in the direction of the test as a leader.
  - Research on the direction of adaptive testing in software.
  - Research on How to carry out the online test more intelligently.
  - Research on software and hardware to simultaneously online testing automatically.
- **Handwritten digit recognition based on pattern recognition** Fall 2104
  - Using PCA algorithm to reduce dimension.
  - Using neural networks algorithm to classify.
  - Using K-means algorithm to cluster.

### **Projects in Department of Electronic and Information Engineering, Tianjin Polytechnic University**

--Spring 2014

- **Control of DC motor based on PID algorithm** Spring 2014
  - Write PID algorithm more intelligently to control DC motor.
  - Research on how to control DC motor automatically.

## PUBLICATIONS

---

1. **Saidi Wang**, Haiyan Hu, Xiaoman Li. A systematic study of motif pairs that may facilitate enhancer-promoter interactions. *Journal of integrative bioinformatics*. 2022 Feb 7. doi: 10.1515/jib-2021-0038. (*IF* = 3.321)
2. Minerva Fatimae Ventolero\*, **Saidi Wang**\*, Haiyan Hu, Xiaoman Li. Computational analyses of bacterial strains from shotgun reads. *Briefings in Bioinformatics*. 2022 Feb 8. doi: 10.1093/bib/bbac013. (*IF*= 11.622, *co-first*)
3. **Saidi Wang**\*, Amlan Talukder\*, Mingyu Cha, Xiaoman Li, Haiyan Hu. Computational annotation of miRNA transcription start sites, *Briefings in Bioinformatics*, Volume 22, Issue 1, January 2021, Pages 380–392, <https://doi.org/10.1093/bib/bbz178>. (*IF* =11.622)
4. **Saidi Wang**, Haiyan Hu, Xiaoman Li, Shared distal regulatory regions may contribute to the coordinated expression of human ribosomal protein genes, *Genomics*, Volume 112, Issue 4, 2020, Pages2886-2893, <https://doi.org/10.1016/j.ygeno.2020.03.028>. (*IF* = 3.76)
5. **Saidi Wang**, Xiaoman Li, Haiyan Hu, A Systematic Evaluation of the Computational Tools for Ligand-Receptor Interaction-based cell-cell Communication Inference. (Accepted by *Briefings in Functional Genomics*, *IF* = 4.241)
6. **Saidi Wang**, Minerva Fatimae Ventolero, Haiyan Hu, Xiaoman Li. SMS: a novel approach for bacterial strain genome reconstruction in multiple samples. (Submitted to *Genomics*, *IF*= 3.76)
7. **Saidi Wang**, Minerva Ventolero, Haiyan Hu, Xiaoman Li. A revisit to the single-copy genes in bacterial genomes. (Accepted by *Nature Scientific Reports*, *IF* = 4.13)
8. Hansi Zheng\*, **Saidi Wang**\*, Xiaoman Li, Haiyan Hu. A computational modelling of the expression of primary-miRNAs. (Submitted to *Nature Scientific Reports*, *IF* = 4.13)
9. Hansi Zheng\*, **Saidi Wang**\*, Xiaoman Li, Haiyan Hu. INSISTC: Incorporating Network Structure Information for Single-Cell Type Classification. (Accepted by *Genomics*, *IF*= 3.76).
10. Minerva Fatimae Ventolero\*, **Saidi Wang**\*, Haiyan Hu, Xiaoman Li. Are known bacterial strains really present? (Submitted to *Nature Scientific Reports*, *IF* = 4.13)
11. Yuxiang Zhang\*, **Saidi Wang**\*, Haiyan Hu, Xiaoman Li. A systematic study of HIF1A cofactors in hypoxic cancer cells. (Submitted to *Nature Scientific Reports*, *IF* = 4.13)
12. **Saidi Wang**, Hansi Zheng, Xiaoman Li, Haiyan Hu. A computational modelling of the expression of primary-miRNAs. Poster, Annual Open House of the UCF Genomics and Bioinformatics Cluster. Apr 14<sup>th</sup>, 2022.

## SOFTWARE AND TOOLS

---

- **[SMS](#): To reconstruct bacterial Strain genomes in Multiple Samples.** The analysis of the bacterial strains is important for the understanding of drug resistance. Despite the existence of dozens of computational tools for bacterial strain studies, most of them are for known bacterial strains. Almost all remaining tools are designed to analyze individual samples or local strain regions. With multiple shotgun metagenomic samples routinely generated in a project, it is necessary to create methods to reconstruct novel bacterial strain genomes in multiple samples. SMS reliably identified the strain number, abundance, and polymorphisms. Compared with two existing approaches, SMS showed superior performance.
- **[EPmotifpair](#): The pipeline to predict enhancer-promoter motif pairs.** Pairs of interacting transcription factors (TFs) have previously been shown to bind to enhancers and promoters and contribute to their physical interactions. However, to date, we have limited knowledge about such TF pairs. To fill this void, we systematically studied the co-occurrence of TF-binding motifs in interacting enhancer-promoter (EP) pairs in seven human cell lines.
- **[StrainToolSurvey](#): The pipeline codes to do the strain tool comparison and analysis.** We investigated twenty-one computational tools that attempt to reconstruct bacterial strain genomes from shotgun reads. For the first time, we discussed the methodology behind these tools and systematically evaluated tools for de novo bacterial strain genome reconstruction on the same datasets. Because of their suboptimal performance, we discussed future directions that may address the limitations.
- **[MixtureS](#): Updated version to make it deal with more complicated mapped reads to predict strains coverage and SNPs.** MixtureS can de novo identify bacterial strains from shotgun reads of a clonal or metagenomic sample, without prior knowledge about the strains and their variations.