Curriculum Vitae

Gali Bai

Email: gali.bai@hotmail.com
Twitter: Gali Bai (@gali_bai)
GitHub: https://github.com/baigal628
Homepage: galib.io
Address: One Jimmy Fund Way, Boston, MA, 02115

Employment

10/2020 - present   Computational Biologist I
Department of Data Science, Dana Farber Cancer Institute, Boston, MA, USA
Advisor: X. Shirley Liu, Tommy (Ming) Tang

Education

08/2018 - 08/2020   M.S. in Genomics and System Biology
Department of Soil and Crop Sciences, Texas A&M University, College Station, TX, USA
Advisor: Hongbin Zhang, Sing-Hoi Sze
Key Courses: Complex Genomics, Bioinformatics, Applied Biostatistics and Data Analysis

09/2014 - 07/2018   B.S. in Agriculture and Biotechnology
China Agricultural University, Beijing, China
Key Courses: Advanced Math (Calculus), Statistics, Biochemistry

Publications

Len Taing*, **Gali Bai***, Clara Cousins*, Paloma Cejas, Xintao Qiu, Zachary T. Herbert, Myles Brown, Clifford A. Meyer, X. Shirley Liu, Henry W. Long, Ming Tang. CHIPS: A Snakemake pipeline for quality control and reproducible processing of chromatin profiling data. *F1000Research* 2021, **10**:517

Presentations

11/2021   **CSHL Single-Cell Analyses Meeting**, **Poster presentation**
Topic: Model-based AnalysEs of Transcriptome and RegulOme (MAESTRO) for single-cell analysis.

11/2021   **Current Topics in Bioinformatics**, Harvard Chan Bioinformatics Core & The Cancer Data Sciences Program at DF/HCC, **Oral presentation**
Topic: Application of MAESTRO in scATAC-seq Analysis

Research Experience

2021- present   **Investigating Killer-cell Immunoglobulin-like Receptors (KIRs) allelic diversity from the public datasets using T1000**
Dana-Farber Cancer Institute, Supervisor: X. Shirley Liu, Bo Li, Li Song

- Compared T-1000 genotyping results with popular genotyping tools such as arcasHLA and HISAT.
- Evaluated T-1000 HLA genotyping robustness by comparing allele variations in paired tumor and normal adjacent samples from TCGA RNA-seq data.
- Improved T-1000 sensitivity in KIR genotyping by tuning parameters for screening dominant and recessive alleles.

2020 – present   **CHIPS: A snakemake pipeline for quality control and reproducible processing of chromatin profiling data**
Dana-Farber Cancer Institute, CIMAC-CIDC Immuno-Oncology Biomarkers Network
Supervisor: X.Shirley Liu, Tommy (Ming) Tang
Source Code: https://github.com/liulab-dfci/CHIPS
- Analyzed high-throughput chromatin profiling data by performing read alignment, peak calling, peak annotations, motif finding, and regulatory potential calculation for all genes.
- Developed a snakemake pipeline for reproducible data analysis of ATAC-seq, ChIP-seq, and DNase-seq.
- Implemented an interactive HTML report that visualizes all CHIPS analysis results.
- Built the google cloud computing environment for large cohort data analysis.
- Processed ATAC-seq data of 81 GD2+ solid tumors treated with anti-GD2 CAR-T cells.

2020 - present   **Model-based AnalysEs of Transcriptome and RegulOme (MAESTRO) for single-cell analysis**
Dana-Farber Cancer Institute, Supervisor: X. Shirley Liu, Tommy (Ming) Tang, Tao Liu
Source Code: https://github.com/liulab-dfci/MAESTRO
- Implemented analysis pipeline for multi-sample scRNA-seq, multi-sample scATAC-seq, and multiome data.
- Integrated Chromap as the default scATAC-seq aligner.
- Built a comprehensive MAESTRO documentation using Workflowr.
- Improved 7x time and memory usage by optimizing alignment and count matrix generating steps for the whole.

2018 – 2020   **Genome-wide association study of seven agricultural traits in Maize**
Texas A&M University, Supervisor: Hongbin Zhang, Sing-Hoi Sze
- Performed QC, alignment and variant calling in 264 whole genome sequencing data.
- Conducted BLUP, PCA, and K-mer clustering to determine subpopulation structure.
- Constructed mixed linear model and generalized linear models for genome-wide association studies.

2017 – 2018   **Discovering aging resistant genes based on Mu transposon elements**
China Agricultural University, Supervisor: Li Li
- Constructed a DNA library with 10,000 samples using a high throughput DNA extraction method.
- Performed BLAST based on an online database and designed PCR primers.
- Screened over 2,000 candidate mutants and identified 1 target sequence potentially related to aging resistance.

Honors and Awards

2016 - 2017   Chao Da Industrial Scholarship, ChaoDa Modern Agriculture Group
2016 - 2017   Third Prize Academic Performance Scholarship, China Agricultural University
2015 - 2016   Third Prize Academic Performance Scholarship, China Agricultural University
2015 - 2016   Cultural Activities Scholarship, China Agricultural University
2017   First Prize in 2017 College Dance Festival

Technical Skills

Bioinformatics   scRNA-seq, scATAC-seq, CHIP-seq, ATAC-seq, RNA-seq, TCR-seq, KIR
Programming   R, Python, Bash
Pipeline   Snakemake
Cloud Computing   Google Cloud Computing, Amazon Web Services
Version Control   Anaconda, Docker, Singularity
Wet Lab   DNA/RNA Extraction, PCR, Electrophoresis, Next-generation Sequencing

Teaching Experience

07/2021- 08/2021   **National Summer Undergraduate Research Project**
Trainee: Lucas Clement
Topic: Identifying pathways involved in colon inflammation induced by cancer immunotherapy by scRNAseq analysis
- Mentored single-cell RNA-seq analysis pipeline.
- Prepared scRNA-seq Seurat object for the student to do further explorations.
- The student won the best presentation in the group.

Certifications

Tidyverse Skills for Data Science in R Specialization, John Hopkins University, Coursera

- Introduction to the Tidyverse

- Importing Data in the Tidyverse

- Wrangling Data in the Tidyverse

Python for Everybody Specialization, University of Michigan, Coursera

- Python Data Structure

- Using Python to Access Web Data