Yaoyi Dai

✉yaoyi.dai@bcm.edu| +1(314) 601- 4999| Google scholar| yaoyidai

Education

Baylor College of Medicine — Houston, TX
Ph.D. candidate in Quantitative and Computational Biology — August 2020 - Current

Washington University in St. Louis, School of Medicine — St. Louis, MO
M.S. in Biostatistics — July 2018 - December 2019

Southwest University — Chongqing, China
B.E. in Food Quality and Safety — August 2014 - June 2018

Skills

Programming: R/R shiny, SAS, Python, Unix/Bash scripting, HPC clusters, MATLAB

Analytical: Next generation sequencing (NGS) analyses including WES/WGS, bulk RNA-seq and scRNAseq, data visualization, machine learning

Experience

Merck Sharp & Dohme LLC — Cambridge, MA
Data, AI, and Genome Sciences (DAGs) — June 2024 - August 2024
Summer Intern

Integrative multi-omics and drug response characterization of in vitro models for KRAS inhibitors

- Designed and implemented a customized R pipeline to analyze DepMap data from over 900 cancer cell lines screened with Merck internal drug compounds.
- Processed and integrated PRISM data, including genomic, transcriptomic, and cell viability information, to evaluate drug responses across different genotypes.
- Investigated mechanisms of responders by analyzing expression profiles, signaling pathways, and comparing mutation and CNV patterns across response groups.
- Generated and tested several biological hypotheses using multi-omics data, contributing to a deeper understanding of drug response mechanisms in KRAS inhibitor-treated models.
- Collaborated with cross-functional teams to present findings, providing insights that could inform future drug development strategies.

The University of Texas MD Anderson Cancer Center — Houston, TX
Wenyi Wang Lab, Department of Bioinformatics and Computational Biology — March 2021 - Current
Graduate Research Assistant

Integrative deconvolution of genomic and transcriptomic data across multi-ethnic primary triple negative breast cancer (TNBC) patients to predict long-term progression free survival outcomes

- Collected and processed data including bulk RNAseq, DNAseq, single-nuclei RNAseq and clinical information from five TNBC patient cohorts across multiple ethnicities (n = 1013).
- Performed deconvolution analysis (DeMixT, ASCAT, SEQUENZA and CIBERSORTx) in bulk tumor tissues.
- Estimated tumor cell mRNA expression scores to delineate transcriptional landscape.
- Stratified TNBC patients' risk of recurrence across multi-ethnic cohorts using 'recursive partitioning'.
- Annotated biological functions of differential expressed genes with GSEA and GO analysis.
- Built an elastic-net linear regression model to identify potential therapeutic targets for high-risk TNBC patients.

Washington University in St. Louis, School of Medicine St. Louis, MO

NeuroGenomics and Informatics Group, Department of Psychiatry February 2020 - July 2020

Bioinformatic Research Analyst

Genetic architecture of disease impacting molecular/cellular architecture and the differences between mendelian and sporadic Alzheimer Disease

- Compared different tools of integrating single nuclei RNA sequencing data.
- Integrated snRNA-seq data from different brain regions.
- Built a random effect mixed model for differential expression analysis of snRNA-seq data.
- Constructed pseudo-time trajectory for astrocytes cluster and discovered reactive pattern long AD progression.

Washington University in St. Louis, School of Medicine St. Louis, MO

Zhao Lab, Department of Neuroscience

Research Assistant

Transcriptomic analysis of mouse cerebellum during associated learning process August 2019 - January 2020

- Generated single cell expression count matrices by processing raw single-cell RNA sequencing data.
- Annotated cell clusters by marker identification.
- Analyzed differential expression of genes (DEGs) and inferred trajectory of cells' pseudo-time.
- Built classification model to predict developmental stage using random forest algorithm.
- Constructed co-expression network to identify gene modules and regulons.

Comparison of single-cell RNA sequencing data integration methods September 2018 - May 2019

- Inspected batch effect caused by integration of multiple datasets generated by different technologies.
- Assessed different algorithms for normalization (regularized negative binomial regression, mutual nearest neighbors' detection, panoramic stitching of heterogeneous single-cell transcriptomic data)

Washington University in St. Louis, School of Medicine St. Louis, MO

Chaudhuri Lab, Department of Radiation Oncology June 2019 - September 2019

Research Intern

Circulating tumor DNA analysis in metastatic castration resistant prostate cancer

- Performed survival analysis for a cohort of 41 patients.
- Measured sensitivity of AR-V7 assay and ctDNA experiment by proportional test.
- Derived regression model for tumor mutation burden captured by whole exome sequencing and Capp-seq.

Publications

Dai, Y., Cao, S., Ji, S., Lim, B., Echeverria, G., and Wenyi Wang. Deciphering transcriptional activity of the tumor microenvironment for robust stratification of chemotherapy response in triple-negative breast cancer. Manuscript under review at *Cell Reports Medicine*.

Dai Y, Guo S, Pan Y, Castignani C, Montierth MD, Van Loo P, Wang W. A guide to transcriptomic deconvolution in cancer. *under 3rd review at Nature Reviews Cancer.*

Wang JR, Zafereo ME, Cabanillas ME, Wu CC, Xu L, **Dai Y**, Wang W, Lai SY, Henderson Y, Erasmus L, Williams MD, Joshu C, Ray D. The association between thyroid differentiation score and survival outcomes in papillary thyroid carcinoma. J Clin Endocrinol Metab. 2024 Aug 1:dgae532. doi: 10.1210/clinem/dgae532.

Brase, L., You, S.-F., del Aguila, J., **Dai, Y.**, Novotny, B. C. et al. (2023). Single-nucleus RNA-sequencing of autosomal dominant Alzheimer disease and risk variant carriers. Nature Commun (2023). https://doi.org/10.1038/s41467-023-37437-5

Cao, S., Wang, J.R., Ji, S., Yang, P., **Dai, Y**. et al. Estimation of tumor cell total mRNA expression in 15 cancer types predicts disease progression. Nat Biotechnol (2022). https://doi.org/10.1038/s41587-022-01342-x

Presentations

Dai, Y., Cao, S., Ji, S., Johan, S., Lim, B. et al. Abstract 2041: Deciphering transcriptional activity of the tumor microenvironment for robust stratification of chemotherapy response in triple-negative breast cancer. **AACR 2025**.

Dai, Y., Cao, S., Ji, S., Johan, S., Lim, B. et al. Deciphering Transcriptional Activity of the Tumor Microenvironment for Robust Stratification of Chemotherapy Response in Triple-Negative Breast Cancer. Platform Talk, Joint Retreat of Department of System Biology & Bioinformatics and Computational Biology at MD Anderson Cancer Center. April 2025.

Dai, Y., Cao, S., Ji, S., Johan, S., Lim, B. et al. Deciphering Transcriptional Activity of the Tumor Microenvironment for Robust Stratification of Chemotherapy Response in Triple-Negative Breast Cancer. Platform Talk, Genetic and Epigenetic Retreat at MD Anderson Cancer Center. Oct 2024.

Dai, Y., Cao, S., Ji, S., Johan, S., Lim, B. et al. Tumor cell total mRNA level predicts chemotherapy response across multi-ethnic patient cohorts with triple-negative breast cancer. **Mathematical Methods in Cancer Biology, Evolution and Therapy - BIRS workshop**. May 2023.

Dai, Y., Cao, S., Ji, S. et al. Abstract 3311: Tumor cell total mRNA level predicts disease progression across multi-ethnic patient cohorts with triple-negative breast cancer. **AACR 2023**.

Dai, Y., Cao, S., Ji, S., Guo, S., Montierth, M., Lim, B., Brown, P., Speed, T., Van Loo, P., and Wang, W. Defining Triple Negative Breast Cancer Subtypes through Measurements of Patient-level Cancer Plasticity. **RECOMB 2022 poster session.** May 2022.

Dai, Y., Zhao, G. Purkinje cell development at single cell level. Invited Poster Presentation, **WashU DBBS Neuroscience retreat**. September 2019.

Harris, P., Chauhan, P., Feng, W., Qiao, M., **Dai, Y**., Usmani, A., Smith, G., Ellis, H., Qaium, F., Chaudhuri, A. Circulating tumor DNA analysis of metastatic castration resistant prostate cancer. **WashU Cancer Biology Division retreat poster session.** September 2019.

Patent

Wang, W., **Dai, Y.**, & Cao, S. (2024). Methods for selecting triple-negative breast cancer patients for targeted therapy using tumor-specific total mRNA expression. U.S. Provisional Patent Application No. 63/554,048.