SONG WU

✉ wusong2019m@big.ac.cn | 🌐 https://wusong.pages.dev/ | 🔬 Google Scholar

Computational Cancer Biology | Bioinformatics | AI4Science | Spatial Transcriptomics

EDUCATION

- ✧ Ph.D. in Bioinformatics (09/2019 – 06/2025) Supervisor: Prof. Yiming Bao
  China National Center for Bioinformation, University of Chinese Academy of Sciences, Beijing, China
- ✧ Bachelor of Science in Bioinformatics (09/2014 – 06/2019)
  Harbin Medical University, Harbin, China

RESEARCH EXPERIENCE

[1] ASCancer Atlas: A Comprehensive Resource of Alternative Splicing in Human Cancers. (*First Author, 2021-2023, Published*)

In this study, I compiled experimentally validated cancer-associated splicing events and systematically analyzed large-scale datasets from TCGA pan-cancer and GTEx pan-tissue cohorts to characterize splicing patterns. Additionally, I developed multi-dimensional online tools that integrate multi-omics and clinical data to identify splicing regulators, assess clinical relevance, and evaluate functional impacts through drug and protein interaction analyses. This resource equips researchers with comprehensive knowledge, data, and tools to understand cancer-specific splicing dysregulation.

[2] SpliceAgent: An LLM-powered AI Agent System for Cancer Alternative Splicing Analysis. (*First Author, 2024-2025, Manuscript in Preparation*)

In this study, I exploratively applied large language models (LLMs) and AI Agent technologies to cancer aberrant splicing research, developing the SpliceAgent. It features three core modules: 1) an intelligent curation assistant based on LLMs; 2) a cancer splicing Q&A chatbot integrating RAG, Graph-RAG, and real-time web retrieval; 3) an agentic analysis workflow automating data-to-report processes.

[3] Characterization of the Splicing Landscape and Identification of Related Genetic Risk Variants in Colorectal Cancer. (*First Author, 2023–2025, In Submission*)

In this study, I identified *EXOC7* exon skipping as a key splicing abnormality linked to colorectal cancer (CRC) prognosis. Through integrated multi-omics analysis including GWAS and sQTLs, I uncovered how genetic variants disrupt RNA-Protein interactions to mediate aberrant splicing, providing new mechanistic insights and a potential therapeutic target in CRC.

[4] HALL: Human Aging and Longevity Study (*Co-first Author, 2023–2024, Rank 2, Published*)

The HALL database aims to collect comprehensive and multi-dimensional datasets from diverse human cohorts, spanning a wide range of ages from young and old individuals to centenarians. The HALL also functions as an open-accessed knowledgebase, featuring meticulously curated information regarding aging- and longevity-associated genes, biomarkers, and biological age clocks.

[5] CROST: Spatial Transcriptomics Repository (*Co-first Author, 2023–2024, Rank 2, Published*)

In this study, I gained extensive expertise in spatial transcriptomics through analyzing 1,313 diverse samples across five different spatial technologies. By creating a unified analytical platform for these datasets, I significantly enhanced research capabilities for investigating tissue structure and spatial gene expression patterns in life sciences.

[6] TargetGene: Cell-type Specific Target Genes for Genetic Variants (*Co-first Author, 2023–2024, Rank 2, Published*)

Through multiple SNP-to-gene linking strategies at different molecular levels, TargetGene identified cell-type-specific target genes. TargetGene comprised 1,276 GWAS summary statistics datasets collected from different sources (UKBB, BBJ, GWAS Catalog), and analyzed 574,279 traits-associated SNPs and discovered 23,838 target genes in 45 tissues and 539 cell types.

[7] CNCB-BLAST Platform: Basic Local Alignment Search Tool (*Co-first Contribution, 2022–2023*)

This online web server compares nucleotide or protein sequence to databases and infers their biological significance from sequence similarity, widely applicable in genomics research.

RESEARCH SKILLS

Multi-omics Data Analysis

- ✧ Genomic data (e.g., WGS, WES, GWAS)
- ✧ Transcriptomic data (e.g., Bulk RNA-Seq, scRNA-Seq, Spatial Transcriptomics)
- ✧ Epigenetic data (e.g., WGBS, EWAS, ChIP-Seq, ATAC-Seq, CLIP-Seq)

Programming Languages: Python (NumPy, Pandas, Seaborn) | R (Tidyverse Suit) | Shell

Full-stack and AI Agent Development:

- ✧ **Frontend & Backend & Database**: React.js | Next.js | Node.js | MongoDB | Neo4j
- ✧ **Chatbot / AI Agent**: LlamaIndex | LangGraph | AI-SDK

Vibe Coding: Familiar with Cursor | MCP Protocol | Docker | Git

SELECTED PUBLICATIONS (with 682 total citations in Google Scholar as of May 20, 2025)

[1] **Wu S**#, Huang Y#, Zhang M, Gong Z, Wang G, Zheng X, Zong W, Zhao W, Xing P, Li R, Liu Z, Bao Y. ASCancer Atlas: a comprehensive knowledgebase of alternative splicing in human cancers[J]. ***Nucleic Acids Research***, 2023, 51: D1196-D1204.

[2] Zheng G#, **Wu S**#, Zhang Z#, Xin Z, Zhang L, Zhao S, Wu J, Liu Y, Li M, Ruan X, Qiao N, Bao Y, Qu H, Fang X. EryDB: A Transcriptomic Profile for Erythropoiesis and Erythroid-related Diseases[J]. ***Genomics, Proteomics Bioinformatics***, 2024, doi: 10.1093/gpbjnl/qzae029.

[3] Wang G#, **Wu S**#, Xiong Z#, Qu H, Fang X, Bao Y. CROST: a comprehensive repository of spatial transcriptomics[J]. ***Nucleic Acids Research***, 2024, 52: D882-D890.

[4] Lin S#, **Wu S**#, Zhao W, Fang Z, Kang H, Liu X, Pan S, Yu F, Bao Y, Jia P. TargetGene: a comprehensive database of cell-type-specific target genes for genetic variants[J]. ***Nucleic Acids Research***, 2024, 52: D1072-D1081.

[5] Li H[#], **Wu S[#]**, Li J[#], Xiong Z[#], Yang K[#], Ye W[#], Ran J[#], Wang Q, Xiong M, Zheng Z, Zhang S, Han Z, Yang P, Jiang B, Ping J, Zuo Y, Lu X, Zhao Q, Yan H, Wang S, Ma S, Zhang B, Ye J, Qu J, Yang Y, Zhang F, Liu G, Bao Y, Zhang W. HALL: a comprehensive database for human aging and longevity studies[J]. ***Nucleic Acids Research***, 2024, 52: D909-D918.

HONORS & AWARDS

The “Outstanding Graduate Student Scholarship” of the Beijing Institute of Genomics (2021)
The “Scholarship of China National Center for Bioinformation” (2024)
The “President's Award of the Chinese Academy of Sciences” (2025)