Jianyu Yang

☎ +1 8149968462 | ✉ yztxwd@gmail.com | yztxwd | Jianyu

Education

Bioinformatics and Genomics Program (Ph.D. candidate) *State College, US*
Pennsylvania State University *Aug. 2020 - May. 2026 (Planned)*

Developmental Biology (Master's Degree) *Guangzhou, China*
Southern Medical University *Sep. 2017 - Jun. 2020*

Preclinical Medicine (Bachelor's Degree) *Guangzhou, China*
Southern Medical University *Sep. 2012 - Jun. 2017*

Research Experience

Interpret Genomics Deep Learning Models via Concept Attribution
Skills: PyTorch, Captum/DeepLiftSHAP, TF-Modisco, Scikit-Learn *Oct. 2024 - Now*

- Improved global concept attribution method Testing with Concept Activation Vector (TCAV) and adapted it to Genomics
- Constructed a concept database for Genomics Deep learning Models
- Highly generalizable and flexible, proved reliability on large scale foundation models and various type of model inputs.

Nucleosome Calling with Bayesian Gaussian Mixture Models (SEM Algorithm)
Related skills: Java, Bayesian Methods, Cell Culture, Molecular Biology, CRISPR-Cas9 *Sep. 2017 - April. 2024*

- Constructed TET family mutant cell lines by CRISPR-Cas9.
- Designed and implemented a nucleosome calling package using Gaussian Mixture model
- Features the unique ability of distinguishing various nucleosome types

Training and interpreting deep learning model for FOXA1 binding partner in A549
Related skills: PyTorch, PyTorch-Lightning, Captum/DeepLiftSHAP, TF-Modisco *Jan. 2022 - Oct. 2023*

- Designed and Applied a bimodal neuron network to dissect the influence of sequence and chromatin to understand FoxA1 binding mechanism

Multimodal Deep Learning for induced Fox Factor Binding Prediction
Related skills: Tensorflow, PyTorch, PyTorch-Lightning, Captum/DeepLiftSHAP, TF-Modisco *Jan. 2022 - Sep. 2023*

- Applied deep neuron network to predict various Fox factors bindings in mouse embryonic stem cell
- Employed feature attribution techniques to understand preexisting chromatin determinants of Fox factors bindings

Regulatory Network Analysis on RUNX1 in leukemia cell
Related skills: R, DESeq, Snakemake *Sep. 2020 - April. 2021*

- Used automated Snakemake workflows to analyze differential expression and TF binding to identify RUNX1 downstream targets in leukemia cells
- Integrated multi-omics datasets to find CENPE as a RUNX1-regulated gene affecting proliferation

Publications

Link to Google Scholar

Systematic Dissection of Sequence Features Affecting the Binding Specificity of a Pioneer Factor Reveals Binding Synergy Between FOXA1 and AP1 *Molecular Cell*
10.1016/j.molcel.2024.06.022 *2024*
Cheng Xu, Holly Kleinschmidt, Jianyu Yang, Erik Leith, Jenna Johnson, Song Tan, Shaun Mahony, Lu Bai

SEM: size-based expectation maximization for characterizing nucleosome positions and subtypes *Genome Research*
10.1101/2023.10.17.562727 *2024*
Jianyu Yang, Kuangyu Yen, Shaun Mahony

Joint sequence & chromatin neural networks characterize the differential abilities of Forkhead transcription factors to engage inaccessible chromatin *Biorxiv*
10.1101/2023.10.06.56122 *2023*
Sonny Arora*, Jianyu Yang*, Tomohiko Akiyama, Daniela Q James, Alexis Morrissey, Thomas R Blanda, Nitika Badjatia, William KM Lai, Minoru SH Ko, B Franklin Pugh, Shaun Mahony

Comprehensive understanding of Tn5 insertion preference improves transcription regulatory element identification *NAR Genomics and Bioinformatics*
10.1093/nargab/lqab094 *2021*
Houyu Zhang, Ting Lu, Shan Liu, Jianyu Yang, Guohuan Sun, Tao Cheng, Jin Xu, Fangyao Chen, Kuangyu Yen

RUNX1 upregulates CENPE to promote leukemic cell proliferation *Frontiers in Molecular Biosciences*
10.3389/fmolb.2021.692880 *2021*
Shan Liu, Jianyu Yang, Guohuan Sun, Yawen Zhang, Cong Cheng, Jin Xu, Kuangyu Yen, Ting Lu

Presentations

2025	**Poster**, Explaining genomics deep learning models via concept attribution	*MLCB*
2024	**Proceeding talk (selected for Genome Research)**, SEM: size-based expectation maximization for characterizing nucleosome positions and subtypes	*RECOMB*
2024	**Talk**, Jointly characterizing the sequence and chromatin binding preferences of transcription factors using neural networks	*GLBIO*
2023	**Poster**, SEM: size-based expectation maximization for characterizing nucleosome positions and subtypes	*Penn State Summer Symposium*
2023	**Poster**, SEM: size-based expectation maximization for characterizing nucleosome positions and subtypes	*GLBIO*
2023	**Poster**, SEM: size-based expectation maximization for characterizing nucleosome positions and subtypes	*Keyston Symposium*
2022	**Poster**, Size-based expectation maximization for characterizing nucleosome positions and subtypes	*RSGDREAM*

Honors & Awards

2025	**Rising Researcher Collaborations Funding Opportunity**, Penn State ICDS	*State College, US*
2024	**Travel Fellowship**, 16th Great Lakes Bioinformatics conference	*Pittsburgh, US*
2020	**Graham Endowment Fellowship**, Bioinformatics and Genomics Program Recruitment	*State College, US*
2018	**1st Prize**, Excellent Graduate Student of Southern Medical University	*Guangzhou, China*
2018	**1st Prize**, National Scholarship for Graduate students	*Guangzhou, China*

Personal Projects

Seqchromloader (Training Data Toolkit for Genomic DL)

Skills: Python, Pytorch, Webdataset *github.com/seqcode/seqchromloader*

- Built a production-ready toolkit to construct training datasets for sequence/chromatin DL models. Optimized for high-throughput, distributed dataset streaming. Has been widely adopted by lab members.

HDF5-Backed Genome Coverage & Heatmap Engine

Related skills: Python, HDF5 *github.com/yztxwd/chiptoolkit*

- Developed a Deeptools-like plotting engine using HDF5 to pre-store genome-wide tracks, enabling extremely fast data retrieving and heatmap/composite plot generation for thousands of regions.

General Snakemake Pipelines for NGS Data

Skills: R, Python, Snakemake, Common packages used in NGS analysis, Slurm, HPC *github.com/yztxwd/snakemake-pipeline-general*

- Built modular snakemake pipelines for ATAC-seq, ChIP-seq, RNA-seq, BS-seq, and MNase-seq. Designed for reproducibility, portability, and HPC batch environments (Slurm). Adapted by lab members for routine preprocessing workflows.

Skills

Programming	Python, JAVA, R
Machine learning & Statistics	Scipy, Scikit-Learning, CuML General machine learning skills, specialized in Bayesian method
Deep learning	Tensorflow, DALI, Pytorch, Webdataset, Pytorch-lightning, Huggingface, Captum, TF-Modisco CNN, Transformer style model design, training, and interpretation
Genomics	ChIP-seq, ATAC-seq, RNA-seq, MNase-seq, BS/TAB-seq analysis
Pipelines & Infra	Snakemake, HDF5, Singularity/Apptainer, Docker, Conda, Slurm, HPC clusters, Proxmox
Wetlab	Molecular Cloning, CRISPR-Cas9, Cell Culture