JIE WU, Ph.D.
Lexington MA
Jie.wu.phd@gmail.com
https://dna-geek.com

Summary

- Demonstrated expertise in leading bioinformatics, data science and machine learning activities across pre-clinical and clinical programs
- 20-year in bioinformatics; 15-year in genomics/NGS; 8-year in healthcare/pharmaceutical industry
- Track record of 49 publications in bioinformatics/oncology/genomics with 10,000+ citations
- Proven capabilities of establishing bioinformatics and data science platforms from scratch
- Developed a variety of bioinformatics workflows/frameworks and wrote three software

Experience

Accent Therapeutics, Lexington MA **2021-2025**
Associate Director, Bioinformatics

- Lead all bioinformatics and data science/machine learning activities to support all pre-clinical and clinical stage programs, including target identification/evaluation, mechanism of action study, disease biology, translational medicine and clinical trial design
- Established cloud based (AWS) computing and bioinformatics infrastructure from scratch, brought in and boosted internal data analysis capabilities, reduced analysis turnaround time from weeks to days
- Developed Accent proprietary database, knowledge-base and data science platform, streamlined biomarker discovery by utilizing public data and internally generated *in vitro* and *in vivo* compound response data
- As the data science expert, collaborated with other functions to develop scientific & business strategies
- Managed internal bioinformatics resources and external CROs to effectively deliver project goals, demonstrated team management and project management skills

Philips Research North America, Cambridge MA **2016-2021**
Senior Scientist, Oncology Informatics, Precision Diagnostics & Image Guided Therapy

- Led genomics data analysis efforts in collaborative projects with academia and hospitals
- Copy number variants analysis with shallow whole genome sequencing and single cell sequencing on blood, breast and pancreatic cancers (Collaboration with MSKCC, *Nature 2022 and Elife 2020*)
- Blood based gene expression signature discovery to predict immunotherapy response in advanced lung cancers (Collaboration with NYMC)
- Hands-on experience with Copy number signatures, mutational signatures and Homologous Recombination Deficiency (HRD), Tumor heterogeneity assessment and subpopulation identification

Massachusetts Institute of Technology, Cambridge MA **2013-2016**
Bioinformatics Scientist, BioMicro Center |Bioinformatics & Computing Facility at Koch Institute

- Developed a reliable approach to detect copy number variants using single cell sequencing *(Genome Research 2016 and PNAS 2014)*
- Genome assembly, annotation and comparative genomic and transcriptomic analysis of three *Pichia pastoris* strains (*BMC genomics 2016*)
- Bioinformatics work for dozens of collaborative projects (e.g. *NAR 2015*) involving mutation calling, RNA-Seq, ChIP-Seq, and other NGS data analysis
- Developed automatic quality control pipelines for RNA-Seq and DNA-Seq data for the facility

Columbia University, New York NY **2012-2013**
Visiting Graduate Student, Department of Systems Biology

- Developed OLego algorithm, allowing fast, accurate and sensitive splice mapping of RNA-Seq reads and enables novel micro-exon discovery (*NAR 2013*)
- Systematic discovery and conservation analysis of alternative spliced exons from mammalian RNA-Seq data which led to discovery of evolution pattern and constraints for NMD-exons (*PNAS 2015*)

Cold Spring Harbor Laboratory, Cold Spring Harbor NY **2008-2013**
Graduate Student, Michael Zhang Lab & Adrian Krainer Lab

- Developed SpliceTrap algorithm: efficient alternative splicing detection and quantification from RNA-Seq data and differential splicing analysis (*Bioinformatics 2011*)
- Established comprehensive splicing focused framework to analyze RNA-Seq data for the lab
- Collaborative projects with other CSHL labs on genomic and transcriptomic analysis such as gene expression, non-coding RNA, alternative splicing, etc.

Shanghai Institute of Materia Medica, Shanghai China **2007-2008**
Research Bioinformatician, Laboratory of Combinatorial Chemistry

- BSSF: Ultrafast functional protein-ligand binding sites predictions based on pharmacophore database and geometry similarity to assist drug design (*BMC Bioinformatics 2010*)

Nankai University, Tianjin China **2004-2007**
Graduate Student, Bioinformatics

- Non-linear analysis of electrophysiological signals, e.g., wavelet analysis on event-related potentials.

Education

2008-2013	**Ph.D. (Computational Biology)**	
	Cold Spring Harbor Laboratory & Stony Brook University, New York, USA	
	Advisors: Prof. Michael Q. Zhang and Prof. Adrian R. Krainer	
2004-2007	**M.Sc. (Bioinformatics)**	Nankai University, China
2000-2004	**B.Sc. (Biophysics)**	Nankai University, China

Software and Workflows

Established workflows	Biomarker discovery, visualization and evaluation workflows Omics workflows: RNA-seq (gene expression, splicing, pathway), variant calling and annotation, Copy Number Variation
OLego (2013)	A fast and sensitive spliced mapping tool for RNA-Seq
SpliceTrap(2011)	A splicing quantification tool for paired-end RNA-Seq data
BSSF(2007)	A binding site prediction tool based on similarity score.

Selected Publications

(For the full list, please visit my google scholar page, * Co-first Author)

- Kristin A. Knouse, **Jie Wu** and Angelika Amon, Assessment of megabase-scale somatic copy number variation using single cell sequencing. *Genome Research*, 2016. gr. 198937.115 (Cover Story)
- Kerry R. Love, Kartik A. Shah, Charles A. Whittaker, **Jie Wu**, M. Catherine Bartlett, Duanduan Ma, Rachel L. Leeson, Margaret Priest, Jonathan Borowsky, Sarah K. Young and J. Christopher Love, Comparative genomics and transcriptomics of *Pichia pastoris*. *BMC Genomics*, 2016. 17.
- Shiou-chi Chang, Bogdan I Fedeles, **Jie Wu**, James C Delaney, Deyu Li, Linlin Zhao, Plamen P Christov, Emily Yau, Vipender Singh and Marco Jost. Next-generation sequencing reveals the biological significance of the N2, 3-ethenoguanine lesion in vivo. *Nucleic acids research*, 2015. gkv243. (Breakthrough Article)
- Qinghong Yan, Sebastien M Weyn-Vanhentenryck, **Jie Wu**, Steven A Sloan, Ye Zhang, Kenian Chen, Jia Qian Wu, Ben A Barres and Chaolin Zhang. Systematic discovery of regulated and conserved alternative exons in the mammalian brain reveals NMD modulating chromatin regulators. *Proceedings of the National Academy of Sciences*, 2015. 112(11):3445-3450.
- Kristin A Knouse, **Jie Wu**, Charles A Whittaker and Angelika Amon. Single cell sequencing reveals low levels of aneuploidy across mammalian tissues. *Proceedings of the National Academy of Sciences*, 2014. 111(37):13409-13414.
- Wangzhi Li, **Jie Wu**, Sang-Yong Kim, Ming Zhao, Stephen A Hearn, Michael Q Zhang, Marvin L Meistrich and Alea A Mills. Chd5 orchestrates chromatin remodelling during sperm development. *Nature*

communications, 2014. 5.

- **Jie Wu**, Olga Anczuków, Adrian R Krainer, Michael Q Zhang and Chaolin Zhang. OLego: fast and sensitive mapping of spliced mRNA-Seq reads using small seeds. *Nucleic acids research*, 2013. 41(10):5149-5163.
- **Jie Wu***, Martin Akerman*, Shuying Sun, W. Richard McCombie, Adrian R. Krainer, Michael Q. Zhang, SpliceTrap: a method to quantify alternative splicing under single cellular conditions, *Bioinformatics*, 2011.27(21): 3010-3016.
- Manli Shen, Eduardo Eyras, **Jie Wu**, Amit Khanna, Serene Josiah, Mathieu Rederstorff, Michael Q. Zhang and Stefan Stamm, Direct cloning of double-stranded RNAs from RNase protection analysis reveals processing patterns of C/D box snoRNAs and provides evidence for widespread antisense transcript expression. *Nucleic Acid Research*, 2011.39(22): 9720–9730.
- Bing Xiong*, **Jie Wu***, David L. Burk, Mengzhu Xue, Hualiang Jiang, Jingkang Shen, BSSF: a fingerprint based ultrafast binding site similarity search and function analysis server, *BMC Bioinformatics*, 2010, 11:47
- **Jie Wu**, Ning Zhang, Zhuo Yang, Tao Zhang, Wavelet coherence and its application in analyzing auditory and motor task event-related potentials. *ACTA BIOPYSICA SINICA*, 2007, 23(6): 482-487.

Expertise

Infrastructure and Management Expertise

- Cloud based computing infrastructure for bioinformatics
- People and project management, resource management to balance priority and workload, budgeting
- CRO evaluation and management
- Code and data management

Bioinformatics Skills

- Extensive experience on RNA-Seq data analysis (quality control, alignment, gene expression, isoform discovery, splicing detection, splicing quantification, differential gene expression and splicing, pathway analysis, gene set enrichment)
- Extensive analysis experience with WES/WGS/ChIP-Seq, copy number variations, mutation calling and annotations, peak calling, motif analysis, genome assembly
- Familiar with oncology global biomarkers (COSMIC mutational signature, chromosome instability, Aneuploidy, whole genome doubling, HRD, MSI, TMB etc.)
- Extensive hands-on biomarker discovery and evaluation experience (enrichment of mutation, copy number, gene expression, global biomarkers etc. based on cell line data)
- Extensive experience with public database including DepMap/CCLE, Sanger cell passport, TCGA, gnomAD etc.
- Molecular dynamics; protein-ligand docking, optimization algorithms.

Programming Skills and Data Analysis Skills

- Programming in Perl, C++, Shell, R, Python and MATLAB
- Statistical analysis and Machine Learning (hypothesis test, linear regression, Bayesian inference, random forest, PLS-DA etc.)
- Cloud computing experience on AWS/Google Cloud/Azure
- Parallel computing experience on clusters with Qsub and MPI
- Familiar with workflow management system (e.g., Nextflow/SnakeMake)
- DevOps and environment management (e.g., Git, Docker, Conda)
- Signal processing (Fourier transformation, wavelet methods)
- Webserver techniques (MySQL, HTML, ASP, JSP and PHP)
- Software for data analysis (Spotfire, GSEA etc.)