William **Connell**

Research Scientist, Machine Learning · Ph.D.

(+1) 707-529-8516 | wconnell93@gmail.com | wconnell.github.io | wconnell | wilstc | @wilstc

Summary

As a consultant, I advise startups and investors on ML-driven drug discovery and platform strategy. At Vevo Therapeutics, I built the ML foundation for an scRNA-seq drug discovery platform, developing evaluation frameworks and scaling models, infrastructure, and teams from seed stage. My academic research in biological sequence-based representation learning guided our strategies for drug response prediction and interpretation. I thrive in collaborative environments and enjoy communicating industry advancements.

Education

University of California, San Francisco — *San Francisco, CA*

PhD in Pharmaceutical Sciences and Pharmacogenomics — *2018 - 2022*

University of California, Los Angeles — *Los Angeles, CA*

BS in Microbiology, Immunology and Molecular Genetics — *2012 - 2016*

Experience

BioxML, LLC — *Berkeley, CA*

Machine Learning Consultant — *Jun 2024 - Present*

- Advised platform drug discovery startups on ML strategy, model selection, and infrastructure for high-throughput screening
- Developed experimental frameworks for benchmarking protein design models and virtual chemical screening approaches
- IGSC member; red-teamed oligomer sequence screening tools using protein design models to identify biosecurity risks
- Collaborated with venture capital firms to analyze AI/biotech market trends, distilling technical insights for stakeholder engagement

Vevo Therapeutics — *San Francisco, CA*

Research Scientist, Machine Learning — *Sep 2022 - Jan 2024*

- Developed benchmarking suite for scRNA-seq foundation models (target retrieval, sensitivity prediction, biological recapitulation)
- Scaled models and data, expanding the pretraining corpus >50x and improving chemical retrieval top-k accuracy by 10x
- Led a proposal accepted into the AWS Gen AI Accelerator program (<1%), securing $300k compute credits and industry recognition
- Managed contractors, recruited three FTEs, and mentored an intern, growing the ML team by 50%

Department of Pharmaceutical Chemistry, UCSF — *San Francisco, CA*

Graduate Researcher advised by Michael Keiser — *Mar 2019 - Dec 2022*

- Pioneered the application of self-supervised learning to scRNA-seq, advancing the transfer learning framework in the field [code]
- Developed an *in silico* chemical probing model (+10% baseline), discovering new indication responses and ferroptosis drug targets [code]
- Identified a genomic biomarker to stratify mAb ustekinumab response in psoriasis, advancing personalized treatment decisions

AI Research Group, Invitae — *Remote*

Computational Biology Research Intern — *May 2021 - Aug 2021*

- Assisted in developing a hierarchical Bayesian model that improved polygenic risk scoring accuracy by 5%
- Engineered efficient software pipelines for managing and analyzing large-scale genomic datasets using tools like Python, R, and AWS

Selected Publications

Learning chemical sensitivity reveals mechanisms of cellular response
William Connell, Kristle Garcia, Hani Goodarzi, Michael J Keiser; *Commun. Biol.* 7.1 (Sept. 2024) p. 1149. Springer, 2024 DOI

Genome-wide association study of ustekinumab response in psoriasis
William Connell, Julie Hong, Wilson Liao; *Frontiers in Immunology* 12 (Jan. 2022). 2022 DOI

A single-cell gene expression language model
William Connell, Umair Khan, Michael J. Keiser; *LMRL Workshop, NeurIPS* (Oct. 2022). 2022 DOI

Extracurricular Activity

Communication — Behind BioML: thought leadership on AI in biology with 10k+ impressions, 175+ subscribers

Scientific community — *3rd place Bio x ML Hackathon*, Evolved 2024; *Foundation Models for Bio*, SynBioBeta 2024 (panel moderator); *MLGenX Workshop*, ICLR 2024 (reviewer); *DNA Diffusion*, OpenBioML (contributing member)

Skills

programming — python [numpy, pandas, pytorch, scikit-learn, huggingface, wandb, lightning, pytorch-geometric, rdkit, hydra], AWS, R, bash, git, plink, snakemake

OS contributions — OpenPlasmid, enformer-finetune, pytorch-metric-learning, scikit-learn sprint (WiMLDS)

model repos — ChemProbe, Exceiver