William **Connell**

RESEARCH SCIENTIST, MACHINE LEARNING · PH.D.

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

Summary

At Vevo Therapeutics, I led machine learning initiatives from seed funding, establishing the technical foundations of an scRNA-seq drug discovery platform. Building on my graduate research in sequence-based representation learning, I developed strategies to integrate chemical and omic features for predicting and understanding drug responses. I thrive in collaborative environments and am passionate about communicating industry advancements to the broader community.

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

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: expanded the pretraining corpus 65x and trained/evaluated models up to 500M parameters
- Delivered a multimodal model aligning transcriptomes and chemical structures, achieving a 10x increase in top-k accuracy
- 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

Datacamp *Remote*
Project Developer *Nov 2019 - Jan 2020*

- Crafted and delivered an interactive data analysis course using R for a leading online educational platform
- Engaged 2,000+ learners with the course, achieving a rating of 4.7/5, contributing to the platform's educational content quality

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 3300+ views, 100+ subscribers

Scientific community *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 enformer-finetune, pytorch-metric-learning, scikit-learn sprint (WiMLDS)

model repos ChemProbe, Exceiver