William **Connell**

RESEARCH SCIENTIST, MACHINE LEARNING · PH.D.

☎ (+1) 707-529-8516 | ✉ wconnell93@gmail.com | ⌂ wconnell.github.io | ⌨ wconnell | in will-connell-26412352 | ✦ @wilstc

Summary

At Vevo Therapeutics, I led the initial machine learning initiatives from seed funding. I developed the core scientific hypotheses around discriminative and generative modeling for a phenotype-based drug discovery platform. I am interested in computational methods that practically influence drug discovery, particularly through a systems perspective. I enjoy working with quickly moving, collaborative teams.

Education

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

PHD IN PHARMACEUTICAL SCIENCES AND PHARMACOGENOMICS   *2018 - 2022*

- Advisor: Michael Keiser
- Committee: Hani Goodarzi, Luke Gilbert

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   *September 2022 - January 2023*

- Launched and steered ML directions: a scRNA-seq foundation model (scFM) and a multimodal model for phenotype-based chemical design
- Oversaw the creation of a model evaluation suite, establishing key performance benchmarks for scFMs in drug discovery
- Led a proposal accepted into the AWS Gen AI Accelerator program (<1%), securing resources and recognition for our team
- Managed contractors, recruited three FTEs, and mentored an intern, fostering team development and expansion

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

GRADUATE RESEARCHER ADVISED BY MICHAEL KEISER   *March 2019 - December 2022*

- Pioneered the application of self-supervised learning to scRNA-seq, advancing the pretrain/finetune framework in biological modeling
- Developed an *in silico* model for chemical probing, finding new indication responses and potential ferroptosis drug targets
- 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 - August 2021*

- Aided in developing a hierarchical Bayesian model to enhance polygenic risk scoring accuracy
- Involved in comprehensive software and pipeline engineering for efficient management and analysis of large-scale datasets

UCSF Innovation Ventures   *San Francisco, CA*

CATALYST AWARDS INTERNSHIP PROGRAM   *September 2018 - September 2019*

- Assessed scientific translation potential of academic research, collaborating with corporate strategy experts
- Headed a team of four to develop a Target Product Profile for a diagnostic gene expression biomarker panel
- Performed comprehensive product development feasibility studies, focusing on commercialization and market competitiveness

Datacamp   *Remote*

PROJECT DEVELOPER   *November 2019 - January 2019*

- Crafted and delivered a compelling data science project for a premier online educational platform
- Developed an interactive data analysis course in R, *Data Science for Social Good: Crime Study*
- Successfully engaged over 1000 learners, achieving a course rating of 4.7/5

Publications

JOURNAL ARTICLES

Learning chemical sensitivity reveals mechanisms of cellular response

William Connell, Kristle Garcia, Hani Goodarzi, Michael J. Keiser

bioRxiv (under review) (Aug. 2023). Cold Spring Harbor Laboratory, 2023. DOI: 10.1101/2023.08.26.554851

Genome-wide association study of ustekinumab response in psoriasis

William Connell, Julie Hong, Wilson Liao

Frontiers in Immunology 12 (Jan. 2022). 2022. DOI: 10.3389/fimmu.2021.815121

A single-cell gene expression language model

William Connell, Umair Khan, Michael J. Keiser

Learning Meaningful Representations of Life Workshop, NeurIPS (Oct. 2022). 2022. DOI: 10.48550/arXiv.2210.14330

Predicting cellular drug sensitivity using conditional modulation of gene expression

William Connell, Michael J. Keiser

Learning Meaningful Representations of Life Workshop, NeurIPS 2020 (Dec. 2020). 2020. DOI: 10.1101/2021.03.15.435529

Helical antimicrobial peptides assemble into protofibril scaffolds that present ordered dsDNA to TLR9

Ernest Y. Lee, Changsheng Zhang, Jeremy Di Domizio, Fan Jin, Will Connell, Mandy Hung, Nicolas Malkoff, Veronica Veksler, Michel Gilliet, Pengyu Ren, Gerard C.L. Wong

Nat. Commun. 10.1 (Mar. 2019) p. 1012. 2019. DOI: 10.1038/s41467-019-08868-w

A single-cell transcriptomic atlas of human neocortical development during mid-gestation

Damon Polioudakis, Luis Torre-Ubieta, Justin Langerman, Andrew G. Elkins, Xu Shi, Jason L. Stein, Celine K. Vuong, Susanne Nichterwitz, Melinda Gevorgian, Carli K. Opland, Daning Lu, William Connell, Elizabeth K. Ruzzo, Jennifer K. Lowe, Tarik Hadzic, Flora I. Hinz, Shan Sabri, William E. Lowry, Mark B. Gerstein, Kathrin Plath, Daniel H. Geschwind

Neuron 103.5 (Sept. 2019) 785–801.e8. 2019. DOI: 10.1016/j.neuron.2019.06.011

PRESENTATIONS

Quantifying the similarity of transcriptomic states in cancer

William Connell, Michael J. Keiser

CZI Neurodegeneration Challenge Network 2020 Annual Meeting. Poster. 2020. Virtual.

Target deconvolution across phenotypic space

William Connell, Garrett Gaskins, Michael J. Keiser

Northern California Computational Biology Symposium. Oral. 2019. Davis, CA.

Extracurricular Activity

OpenBioML   *Remote*

DNA DIFFUSION, CONTRIBUTING MEMBER   *August 2023 - Present*

- Contributed to OS research focusing on *in silico* generative sequence design
- Led the creation of an oracle model for assessing DNA sequences, aiming to boost genetic payload expression
- Coordinated scientific and technical efforts among a diverse group of global collaborators

Skills

programming	python [numpy, pandas, pytorch, scikit-learn, huggingface, wandb, lightning], AWS, R, bash, git, plink, snakemake
OS contributions	enformer-finetune, pytorch-metric-learning, scikit-learn sprint (WiMLDS)
featured blog posts	Blockchain: Tumber Network Analysis; Kanye West: Lyric Analysis