Kyle McGovern, PhD

Resume

Email: kyle.mcgovern93@gmail.com
Website: kyle-mcgovern.github.io/
Google Scholar: kyle mcgovern
Github: kyle-mcgovern
Phone Number: 925-330-8491

Walnut Creek, CA, 94596
USA

Summary

Bioinformatics researcher and data scientist with more than 10 years experience developing statistical, machine learning, and computational tools/pipelines for bioinformatics analyses. Extensive experience working with NGS data including data from 16S rRNA-seq, metagenomic, CRISPR editing, RNA-seq, and nanopore sequencing. Author of the ALDEx3 R software package for robust mixed effects modeling in microbiomics experiments. Introduced a novel framework for machine learning of multi-omics using attention mechanisms. Developed a computational pipeline for SITE-Seq, a novel CRISPR-Cas9 off-target assay, and a patented directed evolution computational workflow for AAV protein engineering.

Experience

Pennsylvania State University | *Graduate Researcher* | State College, PA | 2020-2025

- Authored the ALDEx3 R software package, the only mixed effects modeling method for sequence count data that consistently controls false positives while maintaining power across diverse experiments
- Developed novel statistical theory for improved hypothesis testing in differential expression/abundance analyses
- Developed a bioinformatics and statistical pipeline for predicting chromosomal sex from contaminant host reads in metagenomic sequencing data

Sangamo Therapeutics | *Bioinformatics Engineer II* | Richmond, CA | 2018-2020

- Built a patented pipeline for tissue-targeted AAV protein engineering, automating AAV viral efficiency ranking, sequence homology clustering, and uncertainty quantification
- Developed Nanopore and Illumina sequencing pipeline for barcoding AAV variants for multiplexed engineering of AAV capsids

Caribou Biosciences | *Bioinformatics Programmer* | Berkeley, CA | 2015-2017

- Developed SITE-Seq: a custom CRISPR off-target detection tool in Python for an in vitro assay, leveraging statistical read clustering and motif analysis to assess genome editing specificity and actionable target identification
- Developed a Flask and D3.js web tool to visualize NHEJ, MMEJ, and (micro) homology directed repair patterns from amplicon sequencing data in gene editing experiments

Education

PhD, Bioinformatics & Genomics | Penn State, PA | 2020-2025

Dissertation Title: Statistical and Computational Methods for Addressing the Scale Limitation of Sequence Count Data
Advisor: Dr. Justin Silverman

BS, Bioinformatics | UC Santa Cruz, CA | 2011 - 2015

Publications

McGovern, KM, Silverman, J.D. **Scale reliant mixed effects models enhance microbiome data analysis.** *Microbiome* (2026). doi: 10.1186/s40168-026-02377-x | jsilve24/ALDEx3

McGovern, KM, Silverman, J. **Replacing normalizations with interval assumptions enhances DE/DA**. *BMC Bioinformatics*, 2025. doi: 10.1186/s12859-025-06177-2 | Silverman-Lab/INDExA.

McGovern, KM, Nixon, M., Silverman, J. **Addressing erroneous scale assumptions in gene set enrichment**. *PLoS Computational Biology*, 2023. doi: 10.1371/journal.pcbi.1011659.

Nixon, M., McGovern, KM, et al. **Scale Reliant Inference**. *arXiv*, 2022. doi: 10.48550/arXiv.2201.03616.

Amrani, N., Gao, X., . . ., McGovern, KM, et al. **NmeCas9 is an intrinsically high-fidelity genome-editing platform**. *Genome Biology*, 2018. doi: 10.1186/s13059-018-1591-1.

Select Articles

Kyle McGovern, Dec 12, 2025. **Exploring A Simple Attention Mechanism for Multi-omic Microbiomics**
https://kyle-mcgovern.github.io/attn-multiomic-micro/

Open Source Software

jsilve24/**ALDEx3** | 2025
R package for scale-reliant differential abundance including mixed effects modeling
Role: Co-author, Contributor

Silverman-Lab/**INDExA** | 2024
Interval Null Differential Expression and Abundance Analysis
Role: Author

Patents

ENGINEERING AAV | 2022
Patent # 18630996
David Ojala, McGovern, KM.

Skills

Languages
R, Python, Javascript, Java, C++

Software Packages
Pytorch, sk-learn, pandas, numpy, TensorFlow, SciPy, matplotlib, seaborn, ggplot2, tidyverse, Scanpy, Plotly, flask, django, D3.js, unittest, Pytest

Bioinformatics Tools
BLAST, samtools, Snakemake, BWA, FastQC, DADA2, QIIME, MetaPhlAn

Computational Tools
Linux, HPC, git, LaTeX, airflow, docker

Databases
MySQL, PostgreSQL, MongoDB

Statistics & Machine Learning Attention Mechanisms, Deep Representational Learning, Transformers, VAE, Partially identified Bayesian Models, Bayesian neural networks, conformal prediction, causal inference, random forests, xgboost, Gaussian & Student's T-processes, mixed effects models, hypothesis testing, PCA, SVMs, t-SNE, UMAP, MCMC, naive Bayes, Gaussian mixture models, generalized linear models (GLMs)