Sean Hackett

seanrhackett@gmail.com | GitHub | shackett.org | ☎ 607-592-2678

Summary of Qualifications:

- Actively plans and directs long-term strategy, near-term tactics and day-to-day operations
- Leverages statistics and programming to tackle challenging scientific problems Google Scholar
- Develops systems to improve the accessibility and interoperability of data
- Excels at communicating with diverse audiences in both written and oral formats

Statistics & ML Regression (GLMs, GAMs, non-linear, regularized), MCMC, causal inference, random forest, LDA, Optimization (LP/QP, EM)

Languages Python (pandas, igraph, TensorFlow), R (purrr, ggplot2, Shiny), SQL (BigQuery), GraphQL

Systems & Infrastructure Cloud (GCP), AI (Cursor, MCP), automation (Docker, Airflow, WDL), GitHub

Leadership Skills-based recruiting, planning & prioritization (Asana), cross-functional collaboration

Education

Princeton University - **Ph.D., Quantitative and Computational Biology**

DOE Office of Science Graduate Fellowship (SCGF)

Cornell University - **B.S., Biological Sciences**
Magna Cum Laude with Distinction in Research

Research

Calico Life Sciences LLC — S. San Francisco, CA

Director of Discovery Data Science — Jan 2023 - Apr 2025

- Established a new data science team to support research through cross-functional collaboration and methods development.
- Achieved a 2024 corporate goal by leading the development of a cloud-based genomics platform to process hundreds of internal datasets and contextualize them using thousands of external studies.
- Designed data integration systems that transformed heterogeneous biological information into structured knowledge resources. These platforms connected pathway databases with gene-centric information using a Data Vault model in BigQuery with dbt transformations.

Manager → Assoc Director — Feb 2018 - Jan 2023

- Managed 4-6 data scientists, prioritizing high-value projects in a problem-rich environment.
- Led initiatives around computational education, results sharing and de-duplication of efforts.
- Helped reorganize the Computing team to improve impact, collaboration and accountability as part of a CEO-led eight person working group.

Computational Biologist → Principal Data Scientist — Jan 2017 - Apr 2025

- Designed algorithms to represent genome-scale molecular regulation as a graphical network. Applied this framework to characterize the disease signatures of targets' molecular neighborhoods and to predict the regulators driving scRNAseq pseudotime trajectories.
- Pioneered approaches for discovering novel regulators from perturbation-driven transcriptomic time series using a combination of parametric modeling and LASSO.
- Built a scalable pipeline for metabolomics data processing with automated compound identification and batch correction. This enabled the analysis of multiple aging cohorts, each containing >1,000 samples.

Research

Research	Princeton University, Lewis-Sigler Institute	Princeton, NJ
Postdoctoral Associate	• Supervisor: John Storey, Director of the Center for Statistics and ML • Used Latent Dirichlet Allocation with Empirical Bayes priors to identify latent variables that affect sparse high-dimensional data.	2015 - 2017
Graduate Fellow	• Adviser: Josh Rabinowitz, Professor of Chemistry and Genomics • Supervised two systems biology graduate students. • Developed a scalable algorithm for combining metabolomics, proteomics and fluxes to identify novel allosteric regulators and dissect how metabolite and enzyme concentrations jointly control metabolism.	2010 - 2015

Selected Publications

• Sean R. Hackett, Majed Mohamed Magzoub, Tobias M Maile, Ngoc Vu, Kevin M Wright, Eugene Melamud, Wilhelm Haas, Fiona E McAllister, Gary A Churchill, Bryson D Bennett. *The Molecular Architecture of Variable Lifespan in Diversity Outbred Mice*. bioRxiv, 2023.

• Kevin G Hicks, Ahmad A Cluntun, Heidi L Schubert, Sean R. Hackett, ..., Jared Rutter. *Protein-metabolite interactomics of carbohydrate metabolism reveal regulation of lactate dehydrogenase*. Science, 379 (6636), 2023.

• Sean R. Hackett, Edward A. Baltz, Marc Coram, Bernd J. Wranik, Griffin Kim, Adam Baker, Minjie Fan, David G. Hendrickson, Marc Brendl, R. Scott McIsaac. *Learning causal networks using inducible transcription factors and transcriptome-wide time series*. Molecular Systems Biology, 16 (3), 2020.

• Sam S. Schoenholz, Sean Hackett, Laura Deming, Eugene Melamud, Navdeep Jaitly, Fiona McAllister, Jonathon O'Brien, George Dahl, Bryson Bennett, Andrew Dai, Daphne Kohler. *Peptide-spectrum matching from weak supervision*. ArXiv.

• Sean R. Hackett, Vito R.T. Zanotelli, Wenxin Xu, Jonathan Goya, Junyoung O. Park, David H. Perlman, Patrick A. Gibney, David Botstein, John D. Storey, Joshua D. Rabinowitz. *Systems-level analysis of mechanisms regulating yeast metabolic flux*. Science, 345, 2016.

• J Kamphorst, M Nofal, C Commisso, SR Hackett, W Lu, E Grabocka, G Miller, JA Drebin, MG Vander Heiden, D Bar-Sagi, CB Thompson, JD Rabinowitz. *Human pancreatic cancer tumors are nutrient poor and the tumor cells actively scavenge extracellular protein*. Cancer Research, 75, 2015.

• Jeffrey S. Bruenig, Sean R. Hackett, Joshua D. Rabinowitz & Leonid Kruglyak. *Genetic basis of metabolome variation in yeast*. PLoS Genetics, 2013.

• C Commisso, SM Davidson, RG Soydaner-Azeloglu, SJ Parker, JJ Kamphorst, SR Hackett, E Grabocka, M Nofal, JA Drebin, CB Thompson, JD Rabinowitz, CM Metallo, MG Vander Heiden & D Bar-Sagi. *Macropinocytosis of protein is an amino acid supply route in Ras-transformed cells*. Nature, 497, 2013.

• AJ Greenberg, SR Hackett, LG Harshman & AG Clark. *Environmental and genetic perturbations reveal different networks of metabolic regulation*. Molecular Systems Biology, 7:563, 2011.

Selected Talks

- 2024 Growing Together Conference (Zürich, Switzerland). Invited Talk
Aging as a Data Science Problem

- 2024 Cold Spring Harbor Mechanisms of Aging.
The Molecular Architecture of Variable Lifespan in Diversity Outbred Mice

- 2024 Winter QBio.
The Molecular Architecture of Variable Lifespan in Diversity Outbred Mice

- Future Tech Immersive: AI x Synthetic Biology Meetup. Invited Talk
Dissecting aging's causality with synthetic biology

- 2019 MaxQuant Summer School. Plenary Talk
Bootstrapping the Peptide-Spectrum Matching Problem with Deep Learning

- 2019 Cold Spring Harbor Cellular Dynamics and Models.
Expansive perturbation profiling reveals a causal transcriptional network

- 2017 MIT Sloane Sports Analytics Conference. Research Paper finalist.
Mixed Membership Martial Arts: Data-Driven Analysis of Winning Martial Arts Styles

- 2016 Genomic Sciences Program Annual PI Meeting.
Systems-Level Analysis of Mechanisms Controlling Yeast Metabolic Flux

- 2014 Agilent Emerging Omics Research Tour: 'Omics and Integrated Biology.
Exploring Metabolic Regulation Via Integrative 'Omics.

- 2014 Yeast Genetics Meeting. Plenary Talk: Environmental Sensing Networks.
An Integrated 'Omics Approach to Large-Scale Quantitative Analysis of Cellular Metabolic Regulation

- 2013 International Conference on Systems Biology. Parallel Session: Complex Genetic Traits
Genetic Basis of Metabolome Variation in Yeast