

BRIAN J. ARNOLD, PHD

DATA SCIENTIST | COMPUTATIONAL BIOLOGIST

CONTACT

651-206-2522

brianjohnarnold@gmail.com

brian-arnold.github.io

SKILLS

Data pipelines

Bioinformatics

Python

Snakemake

R

Bash

C++

Docker

Git

AI/ML

Communication

Collaboration

Detail-oriented

EDUCATION

Harvard University

2009-2015

PhD in Organismic and
Evolutionary Biology

University of Minnesota – Twin
Cities

2004-2008

BS in Plant Biology

PROFILE

Cross-functional computational biologist looking to apply 9+ years of professional data science experience to drive the development of machine learning solutions in biology. Proven leader in steering projects from inception to completion, with 25 publications in top-tier scientific journals (Science, Nature Reviews, Genome Biology).

EXPERIENCE

Senior Data Scientist, Princeton University

2020-Present

- Enabled or accelerated data science research across Princeton through collaborations with 8 groups, designing and teaching 6 workshops, and serving as a consultant
- Developed, optimized, and deployed custom encoder-decoder model to reveal gene expression gradients and enhance segmentation of domains in spatial transcriptomics data
- Engineered and deployed machine-learning pipeline to more accurately detect DNA duplications/deletions in bulk tumor sequencing data
- Spearheaded image segmentation and machine learning analyses of 3D videos to quantify neural activity from animal behavior experiments
- Served as individual contributor or manager on additional projects, including whole-genome sequencing of ~700 samples (short- and long-read) and epigenetic profiling of enhancer elements of ~40 samples

Senior Bioinformatician, Harvard University

2018-2020

- Created and published pipeline to massively parallelize mutation detection in whole-genome sequences
- Designed machine learning classifier to detect horizontal gene transfer in whole-genome alignments

Postdoctoral Researcher, Harvard School of Public Health

2015-2018

- Designed fast, memory-efficient C++ code to simulate models of bacterial evolution and used Bayesian Optimization to fit these models to data