

Ryan P Smith

linkedin.com/in/rpseq

github.com/rpseq

San Francisco, CA

ryan.smith.p@gmail.com

+1-319-899-0190

ABOUT ME

Passionate about learning and developing new technologies as well as tinkering with complex systems—especially software infrastructure and data pipelines. After 7 years of bioinformatics and computing work in academia, I am seeking roles in data engineering/science. I am an extremely fast learner and given the opportunity, I can rapidly become an expert in any technology. This resume is a mini DevOps project, auto-built using CircleCI.

PROGRAMMING SKILLS

- **Languages:** Python, Bash, R (ggplot), awk, SQL, C++, golang, L^AT_EX, Mathematica
- **Technologies:** Linux/Unix, AWS (S3, EC2, Redshift), sed, git, Docker, Kubernetes, Airflow, CircleCI, Databricks

EXPERIENCE

- **Eero, Inc.** San Francisco, CA
Contractor January 2019 – Present
 - **Software Engineering, Data:** Developing robust automated data pipelines with Airflow, AWS, Databricks, and everything in between.
- **Circle Computer Resources, Inc.** Cedar Rapids, IA
Remote IT Procurement Specialist November 2018 – January 2019
 - **IT Procurement:** Temporary full-time contract job. Secure reliable internet services for small businesses in hard-to-reach areas. Negotiate with ISPs for service to meet our clients' needs.
- **Girihlet, Inc., Oakland Genomics Center** Oakland, CA
Data Analyst Aug 2018 – Sep 2018
 - **Bioinformatics:** Provide a quality control dashboard for Illumina short read data. General Bash, Python, and Perl scripting to automate mitochondrial genome variant calling.
 - **Information Technology:** Upgrade and maintain the Oakland Genomics Center wireless and LAN networks. Maintain two on-premises CentOS servers for data processing and web hosting (Apache, NGINX)
- **The McDonnell Genome Institute, Washington University** St Louis, MO
Graduate Research Scientist, Ira Hall Lab, Computational Genetics Feb 2015 – Present
 - **Distributed Computing, Bioinformatics:** Process and analyze population-level genome sequencing data to improve our understanding of structural variation in human genomes. Most work is done in a Linux environment with custom Bash pipelines, distributed across a Dockerized computing cluster. Pipelines often use tools combining Bayesian statistics and machine learning approaches, packaged in isolated Docker containers.
 - **Single-cell Sequencing:** Develop novel computational and molecular biology methods to sequence the genomes of single mammalian neurons, in collaboration with the Scripps Research Institute at University of California, San Diego. Typically process up to 10 TB of raw Illumina sequencing data within a 36 hour period.
 - **Teaching Assistant:** Advise students in characterizing the genome sequences of unknown bacteriophages. Hands-on computer lab course using a series of bioinformatics tools to identify genes and predict their function, as well as classifying novel phages into existing phylogenetic groups.
- **University of Iowa** Iowa City, IA
Undergraduate Research Fellow, Adam Dupuy Lab, Cancer Genetics Aug 2010 – May 2014
 - **Genome Editing:** Designed methods for viral genetic engineering in mouse models of human cancers and a sequencing method for detecting resulting transgene insertions.
 - **Bioinformatics:** Processed high-throughput genome sequencing data using Linux bioinformatics tools followed by ad-hoc statistical analyses and data visualization in Python and R.

EDUCATION

- **Washington University** St Louis, MO
MA, Molecular Genetics and Genomics Aug. 2014 – May 2018
- **University of Iowa** Iowa City, IA
BS, Microbiology and Informatics Aug. 2009 – May 2014