

# Ryan P Smith

linkedin.com/in/rpseq

github.com/rpseq

Salt Lake City, UT

ryan.smith.p@gmail.com

+1-319-899-0190

## PROFILE

---

Bioinformatics scientist turned cloud data infrastructure engineer. Strong background with 7 years of bioinformatics computing work in academia and 3 years of professional software engineering on the Data Engineering team at Amazon.

## TECHNICAL SKILLS

---

- **Languages:** Python, Bash, R (ggplot), awk, SQL, C++, golang, L<sup>A</sup>T<sub>E</sub>X, Mathematica, Java, Scala
- **Technologies:** Linux/Unix, Kubernetes, Helm, AWS (S3, EC2, Redshift, EKS, EMR), GCP(GKE), Grafana, Prometheus, Terraform, Looker, git, Docker, Airflow, CircleCI, GitLab CI, Databricks, Spark, HBase, Pinot, Druid, Avro/Parquet, Kafka, GATK, bedtools, bwa-mem

## EXPERIENCE

---

- **eero, an Amazon company** San Francisco, CA  
*Software Development Engineer, Data* *January 2019 – Present*
  - **Data Infrastructure and Operations:** Design, provision, maintain, and scale k8s clusters to power our data systems. All cloud infrastructure is provisioned with Terraform and deployments automated with GitLab. Databricks and Spark for custom batch and streaming transformations. Metrics, monitoring and alerting using Prometheus and Grafana. Manage a diverse collection of data services using GitLab and Helm deploying to k8s clusters, including multi-tenant ETL platform.
  - **Real-time Data Systems:** Build and maintain data systems powering customer-facing features and internal tools. These features provide customers with visual dashboards for observing device bandwidth usage as well as security and content filtering events across a fleet of more than 1.5 million eero routers. Kafka for event streaming and Kafka Connect running in k8s to sync up to 60k records per second to a large HBase cluster, while simultaneously serving 60k customer queries per second on average.
  - **Data Analytics Plaform:** Design and provision all cloud infrastructure for a data platform built around Pinot in k8s. This powers sub-second queries over very large datasets with continuous realtime ingestion of fleet data from Kafka, using Superset for data visualization and exploration.
- **Girihlet, Inc., Oakland Genomics Center** Oakland, CA  
*Bioinformaticist + IT Engineer* *Aug 2018 – Oct 2018*
  - **Bioinformatics:** Provide a quality control dashboard for Illumina short read data. General Bash, Python, and Perl scripting to automate mitochondrial genome variant calling.
  - **IT:** Upgrade and maintain the Oakland Genomics Center 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 – April 2018*
  - **Distributed Computing, Bioinformatics:** Process and analyze population-level whole-genome sequencing data to improve our understanding of structural variation in human genomes. Computation is done in a Linux environment with Bash pipelines, distributed across a high-performance computing cluster. Pipelines often use tools combining Bayesian statistics and machine learning approaches, packaged and deployed via Docker containers.
  - **Single-cell Sequencing:** Develop novel computational and molecular biology methods for whole-genome sequencing of single mammalian neurons. Process up to 10 TB of raw Illumina sequencing data in 36 hours.
  - **Teaching Assistant:** Advise students sequencing the genomes of unknown bacteriophages. Lab course using bioinformatics tools to identify genes, predict their function, and classify into phylogenetic groups.
- **University of Iowa** Iowa City, IA  
*Undergraduate Research Fellow, Adam Dupuy Lab, Cancer Genetics* *Aug 2010 – May 2014*
  - **Genome Editing, Bioinformatics:** Design a viral gene delivery vector for simulating human cancers in mice alongside a DNA sequencing method to detect resulting transgene insertions. Process whole-genome sequencing data using Linux bioinformatics tools followed by ad-hoc statistical analyses 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*