

Professional Experience

**Senior Bioinformatics Scientist, Program Manager**  
GenOmics, Bioinformatics, & Translational (GOBOT) Research Center

RTI International, Research Triangle Park, NC  
Jun 2015 - present

- Supervise multiple teams of data analysts
- Responsible for designing and building "Extract, Transform, and Load" (ETL) pipelines to clean and aggregate data, enhance data interoperability & automate warehousing solutions
- Utilize Snowpipe to ingest Twitter data into Snowflake to directly query and transform JSON data
- Design and create analysis pipelines to clean and query experimental data, automatically generate daily reports using RMarkdown and distribute findings of statistical testing to stakeholders
- Design and configure Amazon Web Services (AWS) Elastic Compute Cloud (EC2) & Relational Database Service (RDS) to host ETL pipelines and Shiny web applications
- Create meaningful visualizations from large data volumes (examples can be found in 2019 publication)
- Interface with both technical and non-technical stakeholders to resolve complex business and technical issues
- Migrated legacy Oracle systems to AWS EC2 and RDS

**Data Engineer (Consultant)**

Animal Cancer DX, LLC, Raleigh, NC  
May 2021 - Sep 2021

- Designed and created an ETL pipeline to aggregate, clean and warehouse experimental data on AWS RDS, automatically generate daily reports with statistical analyses and distribute the reports to stakeholders
- Designed and configured AWS framework to host a Shiny application
- Designed and created data browser dashboard built using Shiny and ggplot

**Bioinformatics Data Scientist**  
Sidney Kimmel Comprehensive Cancer Center

Johns Hopkins Medical Institutions, Baltimore, MD  
Aug 2013 - Jun 2015

- Served as systems administrator for high-performance computing cluster
- Completed 2-year data mining project in 10 weeks using R, SQL, and bash
- Designed and created an ETL pipeline to stage public health datasets and ingest into a SQL database
- Designed and implemented analysis pipelines for large genomic datasets generated from clinical trials
- Incorporated machine learning methods to quality control processes where applicable
- Created novel visualizations from large genomic datasets (examples can be found in 2017 publication)

**Bioinformatics Analyst**  
Pediatrics Infectious Diseases

Johns Hopkins Medical Institutions, Baltimore, MD  
Oct 2011 - Aug 2013

- Individually designed and created an ETL pipeline, increasing the efficiency of the legacy workflow by over 700%

**Bioinformatics Researcher**  
Putonti Lab

Loyola University, Chicago, IL  
Jan 2010 - Jan 2011

- Created ETL pipeline to aggregate and analyze sequence data extracted from NCBI repository

Education

**Johns Hopkins University**  
M.Sc. Bioinformatics

Baltimore, MD  
Aug 2013

**Loyola University**  
B.Sc. Bioinformatics

Chicago, IL  
May 2011

Skills

**Programming**

R (dplyr, tidyr, ggplot), SQL, Python, Perl, bash, Shiny, RMarkdown, knitr, Jupyter, HTML, CSS  
Working knowledge of BaseX, MongoDB

**Technologies**

Tableau, Snowflake, Hive, Jira, Confluence, AWS (EC2, RDS, S3), Git/Subversion, REDCap  
Working knowledge of Airflow, Spark, Hadoop, AWS EMR, CDE/LOINC/OMOP standards

References

Available upon request