

Skills

Programming	R (dplyr, tidyr, ggplot), Python, Perl, bash, Shiny, RMarkdown, knitr, Jupyter, HTML, CSS Working knowledge of Airflow, Spark, Hadoop
Warehousing	SQL, Snowflake Working knowledge of Redshift, Hive, MongoDB, BaseX
Other technologies	Tableau, Jira, Confluence, Git, Unix/Linux environments, AWS, REDCap Working knowledge of Docker, Kubernetes

Professional Experience

Senior Bioinformatics Scientist, RTI International, Research Triangle Park, NC	Jun 2015 - present
<ul style="list-style-type: none">Supervise multiple teams of analysts and developers as program manager and technical task leadDesign and build extract, transform & load (ETL) pipelines to ingest and warehouse data to enhance data interoperabilityUse R to implement data transformations (dplyr, tidyr) and create novel visualizations (ggplot)Utilize Snowpipe to ingest Twitter JSON data into Snowflake, where data is queried and transformedCreate intermediate tables from raw data to support ad-hoc analyses and reportingDesign and create tools and pipelines to support analytics needs and facilitate analysis and insightsDesign and configure AWS framework to host ETL tools and Shiny applicationsInterface with both technical and non-technical stakeholders to resolve complex business and technical issues	
Data Engineer (Consultant), Animal Cancer DX, LLC, Raleigh, NC	Jun 2021 - present
<ul style="list-style-type: none">Designed and created an ETL pipeline to aggregate and warehouse experimental data on AWS S3, and automatically generate and distribute reports with statistical analyses to stakeholdersDesigned and configured AWS framework (EC2, S3, RDS) to host a Shiny applicationDesigned and built data browser dashboard using Shiny and ggplot	
Bioinformatics Data Scientist, Johns Hopkins Medical Institutions, Baltimore, MD	Aug 2013 - Jun 2015
<ul style="list-style-type: none">Completed 2-year data mining project in 10 weeks using R, SQL, and bashDesigned and created an ETL batch process to stage public health datasets and ingest into a SQL databaseUsed Python to implement data transformations (pandas) and create novel visualizations (matplotlib, ggplot)Designed and implemented analysis pipelines for large genomic datasets for exploratory analysis and publicationEstablished unsupervised machine learning methods into quality control process	
Bioinformatics Analyst, Johns Hopkins Medical Institutions, Baltimore, MD	Oct 2011 - Aug 2013
<ul style="list-style-type: none">Designed and created an ETL pipeline to ingest and transform public health data sources for analytic use, increasing the efficiency by 700% over the legacy workflow	
Bioinformatics Researcher, Putonti Lab, Loyola University, Chicago, IL	Jan 2010 - Jan 2011
<ul style="list-style-type: none">Created pipeline to aggregate and analyze genomic 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

Publications

Gern, J. E., Jackson, D. J., Lemanske, R. F., Seroogy, C. M., Tachinardi, U., Craven, M., **Hwang, S. Y.**, ... Bacharier, L. B. (2019). The Children's Respiratory and Environmental Workgroup (CREW) Birth Cohort Consortium: Design, methods, and study population. *Respiratory Research*, 20(1), 115. doi:10.1186/s12931-019-1088-9

Vaz, M., **Hwang, S. Y.**, Kagiampaki, I., Phallen, J., Patil, A., O'Hagan, H. M., ... Baylin, S. B. (2017). Chronic cigarette smoke-induced epigenomic changes precede sensitization of bronchial epithelial cells to single-step transformation by KRAS mutations. *Cancer Cell*, 32, 360–376. doi:10.1016/j.ccell.2017.08.006