DANNY ANTAKI

dannyantaki@gmail.com ✦ 870-833-3770 ✦ dantaki.github.io ✦ San Diego, California ✦ US Citizen

Experience

Twist Bioscience

Senior Bioinformatics Scientist — Oct. '22 – Present

- Led development of an RNA-seq pipeline on a cloud-based platform, which was later deployed to production.
- Co-managed a wet-lab scientist to develop and optimize protocols for a new NGS library prep product line.
- Worked closely with data science team to improve existing ML scoring models for syn-bio gene products.

Bioinformatics Scientist — Feb. '21 – Oct. '22

- Built an ETL pipeline to capture and store daily updates of gene manufacturing outcomes predicted by machine learning.
- Played key role in the design and execution of new experiments aimed at improving MRD targeted enrichment panels, leveraging insights gained from mining internal databases.
- Improved and streamlined pipeline for design of synthetic Cot-1 blockers.

Gleeson Lab for Pediatric Brain Disease

Postdoctoral Researcher — Sep. '18 – Feb. '21

- Laid groundwork for image-based AI model for somatic variant detection (Nature Biotechnology).
- Analyzed deeply sequenced whole genomes for somatic variation analysis in postmortem human brain (Nature) and in sperm (Cell, Nature Medicine)
- Spearheaded study integrating rare variants and polygenic risk scores in an Autism cohort of more than 37,000 individuals (Nature Genetics)

Education

UCSD — Sep. '13 – Sep. '18

PhD, Biomedical Sciences — Advisor: Jonathan Sebat

Thesis: *Disparity between Maternal and Paternal Contributions to Inherited Risk for Autism*

- Led groundbreaking discovery that noncoding deletions inherited from fathers contribute risk for Autism (Science)
- Designed a novel ML tool for structural variation genotyping (Bioinformatics, GitHub)

Purdue University — Jan. '09 – May '13

BS, General Biology – Dean's List & Honor's Biology

Skills

Python – pandas, pyspark, mlflow, sklearn, keras, matplotlib, seaborn, scipy, pysam, pybedtools
Bioinformatics Tools – samtools, bcftools, bedtools, GATK, Picard, VEP, Plink
Bioinformatics Analysis – Structural variation, RNA-seq, Somatic variants, Long Reads, Microarrays
Platforms - AWS, Databricks, Snowflake, Seven Bridges Genomics, Jira, Confluence, Trello
Languages - R, Perl, SQL, Bash