



---

## Experience

### BIOINFORMATICS ENGINEER, EMPIRICO • MARCH 2021 - PRESENT

- Developed an app in React and Python that allows scientists to view GWAS results for phenotypes and variants of interest in a custom styled excel table that automatically gets emailed to the user, and backed up to AWS S3 and Dropbox. This app increased scientist productivity by saving them hours of time that would be spent manually gathering data and styling/filtering the table.
- Wrote a data engineering Python package in PySpark to ingest external GWAS summary statistics data. The package handles all file movement from AWS S3 or Dropbox, data standardization to our schema and data validation, and writes an optimally partitioned output.
- Dockerized and developed bioinformatics pipelines for the alignment and analysis of long-read sequencing, bulk cell RNA-seq, and single cell RNA-seq data.
- Continuing to work on data infrastructure, apps, and bioinformatics pipelines that aid scientists in finding drug candidates via translational genetics.

### RESEARCH TECHNICIAN, PETERS LAB, LA JOLLA INSTITUTE OF IMMUNOLOGY • SEPT 2019 - MARCH 2021

- Wrote a Python package to generate Peptide:MHC binding energy covariance matrices ( hosted on pypi for pip install at <https://pypi.org/project/pmbec/> ).
- Developed models for predicting the binding affinity of peptides to MHC class I molecules that reduce the inherent bias against Cysteine residues, producing more accurate predictions.

### SOFTWARE ENGINEER INTERN, MOMENTIVE.AI (PREV. SURVEYMONKEY) • JUNE 2019 - AUGUST 2019

- Created a survey parser tool to transfer surveys between databases or data centers.
- This tool helped Engineers and QA staff save hours of time by automatically transferring surveys between servers for them, and is still being used at the company.

### SOFTWARE ENGINEER INTERN, CITY OF SAN DIEGO • AUGUST 2018 - JUNE 2019

- Automated the collection, cleaning, and distribution of energy data from various IoT devices.
- Built a variety of databases and APIs for future data leverage and cost vs. power analysis.

### COMPUTER SCIENCE TUTOR, UCSD DEPT OF COMPUTER SCIENCE & ENGINEERING • SEPT 2018 - DEC 2018

- Tutored Introduction to Programming I in Java under Prof. Minnes and Prof. Politz.
- Held open lab hours to help students with programming assignments and in-class questions.

---

## Education and Awards

### UNIVERSITY OF CALIFORNIA, SAN DIEGO • Sept 2016 - December 2020

- B.S. in Computer Science with a specialization in Bioinformatics [ **3.68 GPA** ]

### Winner of the IGNITE | SCALE | Hackathon San Diego • April 2019

- Helped build a web app called Chalk, where communities could decide on public art.

---

## Skills & Projects [ Github: [jacobdp24](https://github.com/jacobdp24) ] [ Personal Website: <https://www.jacobdp.com> ]

- Python, React, Java, JavaScript, C++, C, HTML, CSS, Databricks, AWS, SQLAlchemy, Docker, Airflow, Flask, Django, Git and GitHub, shell scripting, unit testing, Jupyter Notebooks, REST API development and design. Eager to learn more languages, tools, and technologies.
- **Software:** Extensive experience using Matplotlib, pandas, NumPy, and PySpark libraries for data processing, cleaning, prediction, and visualization. Experience with delta and parquet files for data engineering use. Broad experience with scikit-learn, SciPy, and Keras for data science use, moderate experience with PyTorch and TensorFlow. I've written recommender systems and reinforcement learning algorithms for class projects. Vast experience with web scraping, specifically in Python. I've scraped and cleaned data from a wide variety of sources, ranging from experimental solubility data on PubMed to UCSD's professor evaluation system. Built apps/interfaces with React (one of them being my website).
- **Bioinformatics:** Experience with alignment and analysis pipelines for long-read sequencing, bulk cell RNA-seq, and single cell RNA-seq data. Experience with de bruijn graph assembly, BWA aligners, and various other bioinformatics data structures. I also have experience with GWAS, pQTL, and eQTL data and analysis.