



# John Vivian

## Senior Bioinformatics Scientist

Oxnard, CA  
(661) 303-2717  
jtvivian@gmail.com



## Technical Skills

### Languages and Tools

- Python
- Rust
- Git
- Docker
- R
- Anaconda
- Jupyter
- Nextflow

### Data Science & Visualization

- Pandas
- Polars
- Numpy
- PyMC3
- Scipy
- SKLearn
- Matplotlib
- Seaborn
- Plotly
- Streamlit

### AWS and Cloud

- EC2
- S3
- RDS
- ECR
- Batch
- Familiar with Google Cloud and Azure

### Markup & Web

- Markdown
- LaTeX
- HTML
- CSS

## Select Publications

- Toil enables reproducible, open source, big biomedical data analyses
- Comparative tumor RNA sequencing analysis for difficult-to-treat pediatric and young adult patients with cancer
- The UCSC Genome Browser database: 2017 update
- Bayesian framework for detecting gene expression outliers in individual samples
- Error rates for Nanopore discrimination among Cytosine, Methylcytosine, and Hydroxymethylcytosine along individual DNA strands

## FOSS Contributions

- RNA-Seq quantification
- Gene expression modeling
- Scalable workflow engine
- Fast FASTQ pairing

## MOTIVATION

Three main principles drive my career aspirations: a desire to contribute meaningfully to societal well-being by engaging in work that positively impacts lives, a position that allows for continuous intellectual growth, and a team that is generous with both their knowledge and kindness. I believe my background in large-scale cloud computing, familiarity with genomic cancer data analysis, and passion for statistical inference and programming make me an ideal candidate for this position.

## EXPERIENCE

### 2019 - 2023 Senior Bioinformatics Scientist | Atreca, Inc. 📍 South San Francisco

I apply computational expertise to generate insights into the immune responses of oncology patients from one of the world's largest database of antibody and T-cell receptor repertoires. This includes analyzing hundreds of millions of antibody sequences, writing scalable bioinformatics pipelines in Python/Nextflow, and developing methods and statistical models for assays. I also lead analysis efforts on oncology projects that comprise terabytes of heterogeneous data ranging from sequencing reads to single-cell RNA-seq expression data and antibody probes.

### 2015 - 2019 Graduate Researcher | Computational Genomics Lab 📍 UC Santa Cruz

I worked within the Computational Genomics Lab with a focus on distributed computing for large-scale genomics projects and applications. Developed a series of analysis pipelines with emphases on portability, reproducibility, efficiency, and scalability. Collaborated with the Treehouse Childhood Cancer Initiative to develop novel statistical methods for identifying aberrant gene expression in individual cancer patients.

### 2013 - 2014 Graduate Researcher | Nanopore Lab 📍 UC Santa Cruz

I ran and analyzed single-molecule biological nanopore experiments for determining the effectiveness of improving accuracy through a "rereading" mechanism. Designed and implemented a hidden markov model to detect differences in methylation state and quantify improvements in accuracy from rereading.

## EDUCATION

### 2015 - 2019 Doctor of Philosophy | Bioinformatics 📍 University of California, Santa Cruz

Distributed computing, RNA-seq, large-scale cancer genomics, statistical analysis, machine learning, and visualization.

Advisors: DR. DAVID HAUSSLER, DR. BENEDICT PATEN

### 2013 - 2014 Masters of Science | Biomolecular Engineering & Bioinformatics 📍 UCSC

Capstone Paper: *Methods for Analysis of Nanopore Reread Data*

I used hidden markov models as a primary means of modeling "rereads" of a single molecule being translocated through a protein nanopore via a helicase motor.

Advisors: DR. MARK AKESON, DR. KEVIN KARPLUS

### 2010 - 2012 Bachelors of Science | Bioengineering 📍 UCSC

Final Project: *DNA Adapters for Nanopore Sequencing*

I created a protocol for attaching adapter oligos to both synthesized and genomic inserts that allow the adapted construct to interface with a biological nanopore with the help of  $\Phi$ 29 DNA polymerase.

Advisor: DR. MARK AKESON

## Personal Interests

**Academic** - Bayesian inference & ML, programming languages, and data visualization

**Other** - Drumming, Brazilian Jiu Jitsu, technical metal, and cats