

# JOHN VIVIAN

## EDUCATION

Doctor of Philosophy in Bioinformatics	2015–Present	University of California, Santa Cruz Department of Biomolecular Engineering Computational Genomics Lab, Genomics Institute. Distributed computing, RNA-seq, large-scale cancer genomics, statistical analysis, machine learning, and visualization. Advisors: DR. DAVID HAUSSLER, DR. BENEDICT PATEN
Masters of Science in Biomolecular Engineering & Bioinformatics	2013–2014	University of California, Santa Cruz Department of Biomolecular Engineering Capstone Paper: <i>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
Bachelors of Science in Bioengineering	2010–2012	University of California, Santa Cruz Department of Biomolecular Engineering Final Project: <i>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

## RESEARCH EXPERIENCE

UC Santa Cruz Genomics Institute	1/2015–Present	Graduate Researcher Working for 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. Reference: DR. BENEDICT PATEN · +1 (831) 459-5232 · <a href="mailto:benedict@soe.ucsc.edu">benedict@soe.ucsc.edu</a>
UC Santa Cruz Nanopore Lab	09/2013–12/2014	Graduate Researcher Ran and analyzed single-molecule biological nanopore experiments for determining the effectiveness of improving accuracy by “rereading.” Created a hidden markov model to detect differences in methylation state and quantified improved accuracy by rereading. Reference: DR. MARK AKESON · +1 (831) 459-1038 · <a href="mailto:makeson@soe.ucsc.edu">makeson@soe.ucsc.edu</a> Reference: DR. KEVIN KARPLUS · +1(831) 459-4250 · <a href="mailto:karplus@soe.ucsc.edu">karplus@soe.ucsc.edu</a>
UC Santa Cruz Nanopore Lab	06/2012–01/2013	Junior Specialist Ran and analyzed single-molecule biological nanopore experiments for distinguishing epigenetic modifications on cytosine. Created scientific figures and videos for publications. Designed and modeled prototype nanopore device. Reference: DR. ROBIN ABU-SHUMAYS · +1 (831) 459-5537 · <a href="mailto:roabu@soe.ucsc.edu">roabu@soe.ucsc.edu</a>

## PUBLICATIONS

### *April 2017*      **Toil enables reproducible, open source, big biomedical data analyses**

*Nature  
Biotechnology*

Toil is portable, open-source workflow software that supports contemporary workflow definition languages and can be used to securely and reproducibly run scientific workflows efficiently at large scale. We used Toil to process over 20,000 RNA-seq samples to create a consistent meta-analysis of five datasets free of computational batch effects. Nearly all the samples were analysed in under four days using a commercial cloud cluster of 32,000 preemptable cores.

Authors: **Vivian J**, *et al.*

### *November 2016*      **UCSC Genome Browser**

*Nucleic Acids  
Research*

I contributed over 7,500 wiggle tracks derived from samples from the Genotype Tissue Expression Consortium that were used to build signal graphs for the **GTEx Signal Hub**, which is an extension **GTEx Gene Expression Track**.

Authors: Tyner C, *et al.*

### *October 2013*      **Error rates for Nanopore Discrimination Among Cytosine, Methylcytosine, and Hydroxymethylcytosine Along Individual DNA Strands**

*Proceedings of the  
National Academy  
of Science*

Epigenetic modifications to nucleotides influence gene regulation and can be used to predict genetic disease. Current next-generation sequencing devices can only read canonical base identity. The M2MspA nanopore can distinguish between unmodified cytosine, methylcytosine, and hydroxymethylcytosine using machine learning methods. The accuracy of this discrimination is sequence dependent, and characterized for each sequence context.

Authors: Schreiber J, Wescoe ZL, Abu-Shumays R, **Vivian J**, Baatar B, Karplus K, Akeson M.

## TEACHING EXPERIENCE

### *Summer 2016*      **BD2K Summer Workshop Teacher**

*UC Santa Cruz*

Created and taught a week-long bioinformatics bootcamp to undergraduates from around the country. Taught the basics of Git, cloud computing, shell, Python, notebooks, visualization, and gave lectures on evolution, mutation, and next generation sequencing. Materials: [Github repository](#).

Reference: DR. JUDITH CANNER · [jcanner@csumb.edu](mailto:jcanner@csumb.edu)

### *Spring 2014*      **Teacher's Assistant**

*UC Santa Cruz  
Baskin  
Engineering*

Python Programming for Biologists (BME-160). I taught six hours of section a week, held office hours, and graded homework.

Reference: DR. DAVID BERNICK · +1 (831) 459-2061 · [dbernicky@soe.ucsc.edu](mailto:dbernicky@soe.ucsc.edu)

### *Winter 2014*      **Teacher's Assistant**

Lab manager and assistant for the senior design class of 2014 (BME-123B). Graded students' final projects and presentations.

Reference: DR. MARK AKESON · +1 (831) 459-1038 · [makeson@soe.ucsc.edu](mailto:makeson@soe.ucsc.edu)

## ACADEMIC ACCOMPLISHMENTS AND HONORS

<i>UC Santa Cruz</i>	Winter 2014 · Received an ‘Outstanding’ Rank (21) from a National Institute of Health panel for a National Human Genome Research Institute grant ( 1 R01 HG007827-01 ) which funded the nanopore reread project.
	Fall 2012 · Recognition for Achievement in Undergraduate Research
	Fall 2011 · Dean’s Honor List
<i>CSU Bakersfield</i>	Fall 2007 · Hawk Honors Scholarship and Honors Program

## SKILLS

<i>Languages</i>	PYTHON, GIT, DOCKER, R, JAVASCRIPT — acquainted with SCALA, MATLAB, RUBY
<i>Markup/Web</i>	LaTeX, MARKDOWN, HTML, CSS
<i>Computation</i>	HIDDEN MARKOV MODELS, PANDAS, MACHINE LEARNING
<i>Cloud</i>	AMAZON WEB SERVICES, MICROSOFT AZURE
<i>Visualization</i>	ADOBE ILLUSTRATOR, MATPLOTLIB, SEABORN, AUTOCAD
<i>Laboratory</i>	NANOPORE EXPERIMENTATION, GEL ELECTROPHORESIS, PCR, OLIGO DESIGN, PROTEIN CULTURE & EXPRESSION, DNA PURIFICATION

## POSTERS AND PRESENTATIONS

<i>1-30-17</i>	National Institute of Health BD2K Scientific Advisory Board, <i>Santa Cruz, CA</i>
<i>11-28-15</i>	National Institute of Health BD2K All Hands Meeting, <i>Bethesda, CA</i>
<i>11-4-15</i>	Festival of Genomics, <i>San Mateo, CA</i>
<i>10-5-15</i>	Center for Research in Storage Systems, <i>Santa Cruz, CA</i>
<i>9-7-15</i>	Presented to BD2K consortium, <i>Santa Cruz, CA</i>
<i>6-17-15</i>	PanCancer Analysis of Whole Genomes, <i>Aptos, CA</i>