

# VICTORIA CHEUNG

[victoriakcheung@gmail.com](mailto:victoriakcheung@gmail.com)

LinkedIn@[victoriacheung](#)

GitHub@[vic-cheung](#)

website: [vic-cheung.github.io](https://vic-cheung.github.io)

## SUMMARY

Interdisciplinary computational biologist with a proven track record of leveraging advanced algorithms and computational techniques to tackle complex biological problems. Proficient in analyzing multimodal large-scale biological data sets and utilizing EHR data to develop predictive models, and to identify novel patterns to drive breakthrough discoveries. Skilled in programming languages such as Python with a deep understanding of statistical analysis, machine learning, and data visualization. Experienced in collaborating with interdisciplinary teams to translate biological questions into computational solutions, leading to impactful research outcomes. Passionate about harnessing the power of computational biology to revolutionize healthcare and advance scientific understanding.

## EDUCATION

(UCSF) University of California, San Francisco  
(UCSD) University of California, San Diego

PhD  
BS

Genetics conc. Systems Neuroscience  
Microbiology

## TECHNICAL SKILLS

Data Analysis (Python, R, MATLAB)  
Linux (bash, zsh)  
Cloud compute (GCP, AWS)  
Experimental Design  
SQL (PostgreSQL)  
Multiomics (transcriptomics, proteomics, genomics, fragmentomics, methylome)

Adobe Creative Suite (Ai, Lr, Ps )  
Arduino  
CAD (Onshape, Cura, eMachineShop)  
Histology/Immunohistochemistry  
NGS

Animal Research/Surgery  
PCR/qPCR  
Microscopy  
Image Processing (FIJI, Zen)

## CAREER EXPERIENCE

### Computational Biologist, Translational Science | [Freenome](#) | APR 2022 — PRESENT

- Apply bioinformatics, data science, and computational methods (including AI/ML techniques) to analyze multi-omic data to reveal, model, and interpret changes in both the cancer (pathways, gene activities, proteins) and the immune system (composition, activity, and repertoires) associated with clinical outcomes.
- Partner cross-functionally in the scientific planning and execution of collaborative projects, such as molecular and cancer biologists, computational biologists, medical affairs, commercial, business development.
- Developed 2 software packages for reproducible data analysis for the team.
  - (1) Iterated on Freenome's internal fragmentomics modeling architecture to predict gene activation scores from cfDNA.
  - (2) Wrote distributed workflows (Flyte) to increase efficiency of scRNAseq alignment and data aggregation from a scale of running for 8 days to half a day.
- Worked with: scRNAseq, RNAseq, ATACseq, proteomic, methylation, and fragmentomic (cfDNA) data
- Computational Lead on an internal project that validates and characterizes features of Freenome's fragmentomics model
- Execute research projects to model various biological changes resulting from diseases such as cancer, autoimmune disease, and infection with various business partners. Worked collaboratively on:
  - Modeling multi-omics plasma data (DNA methylation, proteomics, fragmentomics) to build classifiers for early stage breast cancer detection in partnership with SIEMENS Healthineers
  - Determining predictive biomarkers for response in R/R DLBCL and characterized responders vs non responders to treatment of Lonca using fragmentomic signatures from whole genome sequencing (WGS) as well as identifying predictive biomarkers for overall survival from plasma proteomics in partnership with ADCT ([ASH 2022 abstract](#) **2nd author**, [AACR 2023 abstract](#) **co-first author**)
- Worked with Freenome's machine learning models to detect colorectal cancer disease burden from deep methylation sequencing of plasma, performing longitudinal monitoring on patients validated with imaging data ([AACR 2024 abstract](#) **accepted, co-first author** pending publication)
- LOD quantification of Freenome's computational fragmentomics approach. Successfully determined the minimum quantity of input mass required to obtain reliable and accurate readouts to establish analytical sensitivity. ([AACR 2024 abstract](#) **accepted, 2nd author** pending publication)
- Mentored other scientists on the team through technical support, infrastructure support, as well as leading journal club discussions.

### Oncology Bioinformatics and Molecular Oncology PhD Intern | [Genentech](#) | SEP 2021 — APR 2022

- Characterized gene signature development and refinement for T cell signaling pathways in cancer models
  - Wrote a data processing pipeline utilizing Scanpy, Numpy, Pandas, scikit-learn, SciPy
  - Performed statistical analyses on different drug treatment populations: gene set enrichment analysis, differential gene expression analysis.
  - Utilized supervised batch correction techniques and unsupervised clustering algorithms (UMAP, topic modeling) to visualize and analyze single cell RNA seq data outputs.

- Wrote custom plotting functions using Matplotlib to better visualize the effect of drug treatments.

#### Graduate Researcher in Single-cell Omics, Systems Neuroscience | [UCSF @Evan Feinberg Lab](#) | JUL 2016 — SEP 2021

- **Project 1:** Developed a multiplexed, high-throughput, single-cell sequencing method for neurons that preserve connectivity information in addition to obtaining molecular identity (VECTORseq).
  - Wrote the data processing pipeline using Python after genome alignment using Cellranger (10x Genomics) on an AWS EC2 instance.
    - Used unsupervised machine learning techniques such as t-SNE/UMAP clustering to match molecular identities to cellular function and role in behavioral output.
    - Implemented nearest neighbors algorithms to account for batch differences when merging datasets.
  - Streamlined brain dissociation techniques and increased neuron survivability yield 100-fold based on data-driven outcomes from clustering analyses.
  - Validated clustering results of single-cell sequencing against the [2020 10x sequencing dataset from the Allen Atlas](#) and that the methodology was functional.
    - Evaluated range of highly variable genes expressed per cluster for the validation of cell identity.
  - Managed collaborations with the Chan-Zuckerberg Biohub (Spyros Darmanis Group, now @ Genentech)
- **Project 2:** Designed an audition-based behavioral paradigm to study sensorimotor integration in the context of mice.
  - Wrote custom software to support custom-built hardware using serial communication between MATLAB and an Arduino microprocessor, which increased productivity by 6-fold from the parallelization and automation of data acquisition, storage, and analysis.
    - Used this system in exploring how sensory input is represented in the brain and transformed into behavioral commands, using mice as the model organism.
  - Wrote custom analyses software to automate, refine, and interpret both raw behavioral data and fiber photometry signals. Used CAD software to design and 3D print custom behavioral apparatuses.
  - Refined surgical protocols to increase survival surgery success by 20%. Delivery of viruses, drugs, and organic dyes into the mouse brain.
  - Performed physiology recordings on brain slices to validate optogenetic and fiber photometry experiments.
  - Assembled fiber photometry and optogenetic manipulation equipment to record and perturb neuronal activity in the context of quantitative behavioral assays.

#### Data Scientist, Health Data Fellowship | [Insight Data Science](#) | MAY 2020 – JULY 2020

- Developed a predictive clinical calculator to assess Acute Kidney Injury in hospitalized patients, which would result in better management, care/medication dosing, injury prevention, and reduced hospital length of stay, thus freeing up occupied resources and minimizing financial costs to both patient and hospital.
- Utilized PostgreSQL querying to gather relevant data from the MIMIC-III database and manipulated the data with Python Pandas from 25 tables of data, 46,000 patients, thousands of diagnoses and lab tests, and clinical documentation--generating over 3 million rows of data and 70 unique features comprising lab tests and demographic information. Analyses were performed on an AWS EC2 instance.
- Used supervised machine learning in Python such as regression models from scikit-learn and XGBoost to forecast Acute Kidney Injury, with a predictive accuracy of ~91%.
- *Medium Article in Towards Data Science:* [Predicting Acute Kidney Injury in Hospitalized Patients Using Machine Learning](#)

#### SELECTED PUBLICATIONS

- Vallania, F.\*, **Cheung, V.\***, Tripathi, A., Louie, M., Snyder, T., Lin, J., Havenith, K., Qin, Y., Pantano, S., Wuerthner, J., van Berkel P.H.; (2023) [Discovery of plasma protein biomarkers associated with overall survival in R/R DLBCL patients treated with loncastuximab tesirine](#). Cancer Res 1 April 2023; 83 (7\_Supplement): 5387.  
<https://doi.org/10.1158/1538-7445.AM2023-5387>
- Vallania, F., **Cheung, V.**, Zamba, MD., Liu, J., Pasupathy, A., Donnell, H., Bailey, M., Louie, M., Lin, J., Havenith, K., Qin, Y., Pantano, S., Wuerthner, J., van Berkel, PH.; [Identification of Predictive Biomarkers for Response of R/R DLBCL Patients Treated with Loncastuximab Tesirine Using Low Pass Whole-Genome Sequencing \(WGS\)](#). Blood 2022; 140 (Supplement 1): 3551–3552.  
doi: <https://doi.org/10.1182/blood-2022-168993>
- Cheung, V.**, Chung, P., and Feinberg, E.H. (2022) [Transcriptional profiling of mouse projection neurons with VECTORseq](#). STAR Protocols, 3(3):101625
- Cheung, V.**, Chung, P., Bjorn, M., Shvareva, V.A., Lopez, Y.C., and Feinberg, E.H. (2021) [Virally Encoded Connectivity Transgenic Overlay RNA sequencing \(VECTORseq\) defines projection neurons involved in sensorimotor integration](#). Cell Reports, 37(12):110131
- Cheung, V.** “[Predicting Acute Kidney Injury in Hospitalized Patients Using Machine Learning](#)” Towards Data Science. Medium, 20 Jun. 2020. Web.