VICTORIA CHEUNG

victoriakcheung@gmail.com LinkedIn@victoriacheung GitHub@vic-cheung website: vic-cheung.github.io

EDUCATION

University of California, San Francisco (UCSF) – PhD in Genetics with a concentration in Systems Neuroscience
University of California, San Diego (UCSD) – Bachelor of Science in Microbiology

TECHNICAL SKILLS

- **Programming**: Python, R, MATLAB, SQL (PostgreSQL), Linux (bash, zsh)
- **Machine Learning**: Scikit-learn, XGBoost, Torch
- **Virtual Environments**: Conda, Poetry
- **Cloud Computing**: GCP, AWS, Azure
- **Data Visualization**: Matplotlib, Seaborn, ggplot2
- **Bioinformatics/NGS Data Processing**: CellRanger, SAMtools/pysam, BedTools, DESeq, fgsea, Azimuth, ScanPy/Seurat
- **Multiomics/Multimodal Analysis**: Transcriptomics, Proteomics, Genomics, Fragmentomics, Methylomics
- **Additional Skills**: Arduino, Onshape, Cura, eMachineShop, CAD software, Animal Research, Image Processing, Histology/Immunohistochemistry, FIJI, Zen, PCR/qPCR

CAREER EXPERIENCE

Computational Biologist, Translational Science | *Freenome* | APR 2022 — PRESENT

- Applied bioinformatics and computational biology techniques to analyze high-throughput multi-omic datasets (genomic, transcriptomic, epigenetic) for liquid biopsy and cancer diagnostics.
- Developed and implemented computational tools and algorithms for analyzing complex data from various NGS platforms, including RNA-Seq, Whole Genome Sequencing (WGS), targeted sequencing, and methyl sequencing. Conducted analyses of raw NGS data (FASTQ and BAM files) from single-cell RNA-seq/bulk RNA-seq data.
- Utilized existing bioinformatics tools (e.g., SAMtools/pysam, BedTools, fgsea, ScanPy) alongside custom scripting in Python and Bash to uncover insights and drive innovation in diagnostic solutions.
- Developed 2 software packages for reproducible data analysis for the team: (1) Freenome's internal fragmentomics modeling architecture to predict gene activation scores from cfDNA. (2) Wrote distributed workflows (Flyte) to enhance the efficiency of NGS data processing, reducing the alignment and data aggregation time for scRNAseq from 8 days to half a day. Ensured the reproducibility and scalability of bioinformatics analyses.
- Led cross-functional collaborations with molecular and cancer biologists, computational researchers, and external partners to design, analyze, and interpret molecular experiments. Key contributions included building classifiers for early-stage cancer detection and identifying predictive biomarkers for treatment response.
- Multi-Omic Data Integration: Led the integration of multimodal data (fragmentomics, methylomics, and proteomics) for the development of early fusion and late fusion models for partnership work (computational lead). This work contributed to identifying novel cancer biomarkers, monitoring longitudinal disease progression, and characterizing cancer signatures.
- Prepared detailed reports and presentations for stakeholders, including abstracts and posters presented at major conferences such as AACR and ASH. Communicated complex scientific findings effectively to both technical and non-technical audiences.
- Mentored junior scientists and facilitated knowledge sharing within the team through technical support, leading journal club discussions, and fostering a collaborative research environment.
- Identified and developed collaborations with key opinion leaders (KOLs) to leverage Freenome's platform as well as develop new computational tools.
- Key Projects and Achievements:
  - Partnership with ADCT: 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 (ASH 2022 abstract, poster **2nd author**, AACR 2023 abstract **co-first author**)
  - Partnership with Siemens Healthineers: Modeled plasma DNA methylation, proteomics, and fragmentomics data to build classifiers for early-stage breast cancer detection.
  - Fragmentomics Model Validation: Served as the computational lead on a project that validated Freenome's fragmentomics model
    - Manuscript currently in preparation as first author
    - 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, poster **2nd author**)
  - Machine Learning in Liquid Biopsy: Applied machine learning models to detect colorectal cancer disease burden as well as perform longitudinal monitoring on patients through deep methylation sequencing of plasma, validated with imaging data (AACR 2024 abstract, poster **co-first author**)

Oncology Bioinformatics and Molecular Oncology PhD Intern | *Genentech* | SEP 2021 — APR 2022

- Characterized gene signatures for T cell signaling pathways in cancer models, optimizing for biomarker discovery.
- Built a data processing pipeline using Scanpy, Numpy, Pandas, scikit-learn, SciPy for statistical analyses and Matplotlib/Seaborn visualization of single-cell RNA-Seq data.
- Performed statistical analyses on different drug treatment populations: gene set enrichment analysis, differential gene expression analysis.
- Applied supervised batch correction and unsupervised clustering (UMAP, topic modeling) to analyze scRNA-seq data.

Graduate Researcher in Single-cell Omics, Systems Neuroscience | *UCSF @Evan Feinberg Lab* | JUL 2016 — SEP 2021

- **Project 1:** Developed VECTORseq, a high-throughput single-cell sequencing method for neurons, preserving connectivity information.
  - Developed a data processing pipeline using Python after genome alignment using Cellranger (10x Genomics) on an AWS EC2 instance.
  - Applied unsupervised ML techniques (t-SNE/UMAP clustering) to correlate molecular identities with neuronal function and behavioral output, using nearest neighbors algorithms to adjust for batch differences.
  - Enhanced neuron survivability yield by 100x through optimized brain dissociation techniques based on data-driven outcomes from clustering analyses.
- **Project 2**: Designed vision and audition-based behavioral paradigm to investigate sensorimotor integration in mice.
  - Explored sensory input representation and behavioral command transformation using custom software for behavioral and fiber photometry data.
  - Designed and 3D-printed (LulzBot) custom behavioral apparatuses using CAD software (Onshape).
  - Assembled infrared sensors to detect animal interactions in Skinner boxes, improving precision and efficiency in tracking behavioral responses.
  - Developed a rotary encoder system with Arduino to track mice responses to auditory cues and analyzed decision-making with MATLAB.
  - Wrote software to automate and parallelize data acquisition with MATLAB and Arduino, boosting productivity by 6-fold.
  - Conducted neural activity recordings via calcium-based imaging and fiber photometry, applying signal processing techniques (e.g., noise filtering, event detection) to analyze neuronal dynamics.
  - Improved surgical protocols for virus delivery and optogenetic tools, increasing survival rates by 20%.
  - Validated optogenetic and fiber photometry experiments through electrophysiology recordings.

Data Scientist, Health Data Fellowship | *Insight Data Science* | MAY 2020 – JULY 2020

- Developed a predictive clinical tool to assess Acute Kidney Injury (AKI) in hospitalized patients, improving care management and reducing hospital stays.
- Queried and processed data from the MIMIC-III database using PostgreSQL and Python (Pandas), generating over 3 million rows of data and 70 unique features from 25 tables of data and 46,000 patients. Features included lab tests, demographic information, thousands of diagnoses, and other clinical documentation. Analyses were performed on an AWS EC2 instance.
- Applied supervised machine learning models (scikit-learn, XGBoost) to forecast AKI, achieving a predictive accuracy of ~91%.
- Published work in ***Towards Data Science***: Predicting Acute Kidney Injury in Hospitalized Patients Using Machine Learning

PUBLICATIONS [* denotes equal contribution]

Herault, A, et al., **Cheung, V.** "NKG2D-bispecific enhances NK and CD8+ T cell antitumor immunity." Cancer Immunology, Immunotherapy 73.10 (2024): 1-16. DOI

Tang, A.D.*, Gupte, R.*, **Cheung, V.*,** Qing, T.*, et al "Noninvasive longitudinal monitoring of residual disease in chemotherapy-treated colorectal cancer patients." Cancer Research 84(6): 6258. DOI

Leff, E.*, Tseng, A.*, **Cheung, V.**, et al. "Inference of gene expression using fragmentation patterns from targeted high-depth sequencing of cell-free DNA." Cancer Research, 84(6_Supplement), 2024, 4788. DOI

Vallania, F.*, **Cheung, V.***, Tripathi, A., et al. "Discovery of plasma protein biomarkers associated with overall survival in R/R DLBCL patients treated with loncastuximab tesirine." Cancer Research, 83(7_Supplement), 2023, 5387. DOI

Vallania, F., **Cheung, V.**, Zamba, MD., Liu, J., Pasupathy, A., et al. "Identification of Predictive Biomarkers for Response of R/R DLBCL Patients Treated with Loncastuximab Tesirine Using Low Pass Whole-Genome Sequencing (WGS)." **Blood**, 140(Supplement 1), 2022, 3551-3552. DOI

Cheung, V., Chung, P., Feinberg, E.H. "Transcriptional profiling of mouse projection neurons with VECTORseq." STAR Protocols, 3(3), 2022, 101625. DOI

Cheung, V., Chung, P., Bjorni, M., Shvareva, V.A., Lopez, Y.C., Feinberg, E.H. "Virally Encoded Connectivity Transgenic Overlay RNA sequencing (VECTORseq) defines projection neurons involved in sensorimotor integration." Cell Reports, 37(12), 2021, 110131. DOI