VICTORIA CHEUNG

email: victoriakcheung@gmail.com website: https://vic-cheung.github.io/

EDUCATION

University of California, San Francisco (UCSF) – PhD in Genetics, research focus in Systems Neuroscience
University of California, San Diego (UCSD) – BS in Microbiology

TECHNICAL SKILLS

- **Programming Languages:** Python, R, MATLAB, SQL
- **Machine Learning:** Supervised and unsupervised techniques (e.g., SVM, Random Forest, Boosted Trees, LDA, NMF), hyperparameter tuning, model evaluation
- **Data Science & Analytics:** Multi-omics data analysis, feature engineering, data preprocessing, statistical analysis, cross-validation, model deployment, data fusion
- **Bioinformatics:** CellRanger, BedTools, ScanPy, fgsea, pysam, samtools
- **Cloud Computing & Tools:** AWS, GCP, Azure, Flyte, Conda, Poetry, UV
- **Frameworks & Libraries:** PyTorch, scikit-learn, Pandas, NumPy, SHAP, Matplotlib
- **Operating Systems:** Linux, macOS

PROFESSIONAL EXPERIENCE

Senior Computational Clinical Scientist, Translational Medicine | *Revolution Medicines* | JUN 2025 — PRESENT

- Developed modular Python and R packages to automate ETL, visualization, and statistical analyses of clinical trial data, including API/CLI interfaces, Pydantic-based configuration, and SQL-driven weekly reporting for continuous trial monitoring.
- Built reproducible analytical infrastructure with standardized covariate and survival-modeling frameworks (scikit-learn, SciPy, lifelines), producing traceable and consistent Kaplan–Meier outputs with full documentation.
- Designed ML-driven workflows to predict progression-free survival (PFS), integrating baseline co-mutation feature engineering, multivariate modeling, and ML-based feature selection to identify statistically and biologically meaningful predictors.
- Established engineering and reproducibility standards across the translational data team, including shared codebases, virtual environments, Git workflows, structured documentation, and automated testing.
- Mentored scientists and engineers on code reviews, onboarding, and best practices, accelerating adoption of Python, reproducible workflows, and modern development practices.
- Conducted biomarker and clinical outcome analyses, integrating automated infrastructure with scientific insight to inform translational hypotheses, biomarker evaluation, and clinical trial strategy.

Computational Biologist, Translational Science | *Freenome* | APR 2022 — MAY 2025

- Leveraged high-dimensional multi-omic liquid biopsy datasets (genomic, transcriptomic, epigenetic, proteomic, methylomic) to derive actionable insights for cancer diagnostics, disease monitoring, and treatment response evaluation.
- Designed pipelines and implemented computational tools (supervised and unsupervised ML) for multimodal data analysis, including RNA-Seq, WGS, targeted sequencing, and methyl sequencing.
- Developed software for modeling cfDNA fragmentomic signatures and applied wavelet-based signal processing to preprocess data, enabling robust machine learning model development.
- Engineered scalable, reproducible NGS pipelines on distributed compute workflows (e.g., Flyte), reducing analysis time by 80%.
- Integrated fragmentomics, methylomics, and proteomics data to develop classifiers for early-stage cancer detection, predictive biomarker identification, and disease monitoring; applied custom feature engineering and SHAP-based model interpretation.
- Designed visualizations (Matplotlib, Seaborn) and presented results to biopharma partners and internal teams; authored scientific slide decks and conference posters.
- Led cross-functional teams of biologists and computational researchers; served as scientific lead for biopharma partnerships, mentoring junior scientists and driving computational tool adoption.
- Key Projects and Achievements:
  - Subtyped SCLC with plasma multi-omics using ML (ASCO 2025 abstract, poster)
  - Developed multimodal profiling for urothelial carcinoma (AACR 2025 abstract, **co-first author**)
  - Applied ML for CRC residual disease monitoring (AACR 2024 abstract, poster **co-first author**).
  - Validated fragmentomics platform; contributed to LOD quantification (AACR 2024 abstract, poster **2nd author**)
  - Patent pending US20250305061A1: “Methods and systems for inferring gene expression using cfDNA fragments” (submitted 3/27/2025)
  - Partnership with ADCT**:** Identified predictive biomarkers for response in R/R DLBCL and characterized responders vs. non-responders to Lonca treatment through multimodal data analysis. (ASH 2022 abstract**,** poster **2nd author**, AACR 2023 abstract **co-first author**).

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

- Contributed to early-stage drug development by characterizing gene signatures for T cell exhaustion after treatment in cancer models, providing insights into poorly understood biological mechanisms.
- Designed and implemented an efficient data processing pipeline for large-scale single-cell RNA-Seq analysis using Scanpy, Numpy, Pandas, scikit-learn, SciPy, and Matplotlib/Seaborn.
- Conducted comprehensive statistical analyses, including gene set enrichment analysis (pathway analysis) and differential gene expression, to identify key biological insights from T cells after administration of therapeutic agents.
- Applied supervised batch correction techniques and unsupervised clustering algorithms (UMAP, LDA, NMF) to reveal novel patterns in large-scale scRNA-Seq data.
- Collaborated effectively with cross-functional teams to deliver high-quality computational insights, driving the advancement of oncology research programs.

Data Science Fellow–Health Data Analytics | *Insight Data Science* | MAY 2020 – JULY 2020

- Developed a predictive clinical tool using supervised machine learning models (scikit-learn, XGBoost) to forecast Acute Kidney Injury (AKI) with ~91% accuracy, enhancing care management and potentially reducing hospital length of stay.
- Queried and processed over 3 million rows of data from the MIMIC-III database (publicly accessible collection of de-identified medical data from critically ill patients) using PostgreSQL and Python (Pandas), extracting 70 unique features from 25 tables and 46,000 patients. Features included lab tests, demographic information, thousands of diagnoses, and other clinical documentation.
- Leveraged AWS EC2 for efficient data processing and analysis, demonstrating proficiency in cloud computing.
- Published work in ***Towards Data Science***: Predicting Acute Kidney Injury in Hospitalized Patients Using Machine Learning

PhD Researcher–Neuroinformatics and Systems Neuroscience | *UCSF @Evan Feinberg Lab* | JUL 2016 — SEP 2021

- Method Development and Computational Biology**:**
  - Developed VECTORseq, a multiplexed single-cell sequencing method combining viral barcoding with scRNA-seq to simultaneously capture neuronal molecular identity and long-range connectivity, reducing experimental costs and labor.
  - Built Python-based pipelines for barcode detection, noise filtering, demultiplexing, QC, normalization, and automated connectivity-aware clustering; deployed workflows on AWS EC2 for scalable processing.
  - Improved brain-dissociation protocols using clustering-derived metrics, increasing viable neuron yield ~100×.
- Machine Learning & Single-Cell Analysis:
  - Applied unsupervised ML (Leiden, NMF, UMAP, t-SNE) to map transcriptomic structure and identify projection-defined neuronal populations at scale.
  - Integrated VECTORseq connectivity labels with transcriptomic features to define novel projection-enriched subtypes across cortico-midbrain circuits.
  - Validated the method by recovering known cell types and discovering new ones, including a $Pax6^{+}$ ZI→VM GABAergic projection population.
- Signal Processing & Experimental Automation:
  - Arduino/MATLAB systems for behavioral tracking and fiber photometry; automated acquisition pipelines, increasing throughput ~6×.
  - Performed signal processing on calcium imaging (denoising, event extraction) to link sensory-driven activity with behavior.
- Key Outcomes:
  - Enabled discovery of projection-defined neuronal populations with linked transcriptomic profiles.
  - Published findings in *Cell Reports* and methods in *STAR Protocols* journals.

PATENTS

Leff, E.; Tseng, A.; **Cheung, V.**; Gupte, R.; Tabari, E.; Vallania, F.; Donnella, H.; St. John, J.; Lovejoy, A.; Ulz, P. "Methods and systems for inferring gene expression using cell-free DNA fragments." US Patent Application US20250305061A1, filed March 27, 2025, pending. Freenome Holdings, Inc.

SELECT PUBLICATIONS [* denotes equal contribution], full list on Google Scholar

Reck, M., Gupte, R., Qing, T., **Cheung, V.**, Tang, A.D., et al. "Machine learning classification of small cell lung cancer using plasma multiomics in IMPower133." *Annals of Oncology*. DOI

Cauwels, A.*, Levy, E.*, Tang, AD.*, **Cheung, V.***, et al. "Plasma-based multimodal profiling of urothelial carcinoma to characterize tissue-based subtypes." *Cancer Research* 85(8_Supplement_1): 3679. 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

Cheung, V., et al. "Virally Encoded Connectivity Transgenic Overlay RNA sequencing (VECTORseq) defines projection neurons involved in sensorimotor integration." *Cell Reports*, 37(12), 2021, 110131. DOI