# Phu T. Van, PhD

*email* pvan@alumni.cmu.edu

*website* ptvan.github.io

## WORK EXPERIENCE

*2021Nov -* **Bioinformatics Solutions Manager → Senior Manager**

*TwinStrand Biosciences*

Served as bioinformatics expert on company panel for approving new projects. Spearheaded *à la carte* bioinformatic service offering to customers to increase revenue. Provided analysis guidance to customers and co-authored customer communication SOPs with Field&Technical Support department. Mentored Bioinformatics Solutions scientists. Oversaw development of new R data analysis packages. Initiated and later supervised data analyses that expanded TwinStrand DuplexSeq™ support to two NGS sequencing platforms. Produced whitepapers and reviewed marketing materials for technical accuracy.

*2021Jan - Oct* **Bioinformatics Scientist II**

Performed analyses on duplex sequencing data for diverse projects including mutagenesis and cancer MRD. Reviewed and contributed R code to internal bioinformatic pipelines and customer-facing reports. Prepared data products and presented findings to industry and academic clients.

*2014 - 2020* **Postdoc → Bioinformatics Analyst**

*Fred Hutchinson Cancer Center*

Coordinated projects among PhD-level analysts, led brainstorming sessions and journal clubs. Performed analyses on flow cytometry and transcriptomic datasets. Created data analysis pipelines & interactive data visualizations. Authored Statistical Analysis Plans & peer-reviewed manuscripts. Reviewed junior analysts' code, mentored PhD students and bench scientists on statistics. Developed R code to normalize mass cytometry data across experiments using multi-mixture models. Performed dimension reduction on mass cytometry data and identified correlates of disease using regression.

*2009 - 2014* **Doctoral Student**

*Carnegie Mellon University*

Designed and built a patented high-dynamic-range protein gel imager with robotic gel cutting arm for capturing rare proteins. Developed **SIGILab**, a C++ GUI application controlling gel imager's acquisition of high-dynamic-range images. Developed bash scripts to quantify protein abundances in 2DE gel images. Lectured in Modern Biology course and mentored junior students.

## TECHNICAL SKILLS

*Programming* R, Python, Matlab, C/C++, bash, JavaScript, Perl

*Data Analysis / Statistics* dimension reduction, clustering, regression/classification, linear models, variable selection, hypothesis testing

*Databases* MySQL/MariaDB, PostgreSQL, SQLServer

*Workflows* Nextflow, Slurm, Docker, GitHub

*Cloud Computing* AWS EC2, S3, ECR

*Bioinformatics* BWA, STAR, RSEM, BLAST, SAMtools, BCFtools, VardictJava, varscan, delly, limma, edgeR

## EDUCATION

*2009 - 2014* **Carnegie Mellon University**

PhD, Biological Sciences

*2001 - 2007* **University of Washington**

BS, Biology (Physiology specialization); BS, Wildlife Sciences

## SELECTED PUBLICATIONS, SOFTWARE & PATENTS

*complete list: https://scholar.google.com/citations?user=IDadFEkAAAAJ*

*2023 Oct* **Comparison of the Transgenic Rodent Mutation Assay, Error Corrected Next Generation Duplex Sequencing, and the Alkaline Comet Assay to Detect Dose-Related Mutations Following Exposure to N-Nitrosodiethylamine**

*Mutation Research*

This collaboration with Pfizer, Gilead & Merck compared DuplexSeq against gold-standard genotoxic assays in detecting mutations caused by NDEA, a chemical carcinogen. I analyzed DuplexSeq data and co-wrote the manuscript.
Authors: Joel Bercu, Shaofei Zhang, Zhanna Sobol, Patricia Escobar, Phu T. Van, *et al.*

*2023 Jun* **Duplex Sequencing Provides Detailed Characterization of Mutation Frequencies and Spectra in the Bone Marrow of MutaMouse Males Exposed to Procarbazine Hydrochloride**

*Archives of Toxicology*

This HealthCanada collaboration studied mutations in procarbazine-exposed mice using DuplexSeq as a potential replacement for the gold-standard LacZ test. I performed sequence alignment, variant calling and analyzed mutation data. We also studied ENU-exposed human cells, published in Cho 2023.
Authors: Annette Dodge, Danielle LeBlanc, Andrew Williams, Phu T. Van, *et al.*

*2021 Jun* **Monocyte metabolic transcriptional programs associate with resistance to tuberculin skin test/interferon-$\gamma$ release assay conversion**

*Journal of Clinical Investigations*

This manuscript reports a potential link between oleic metabolism and *Tuberculosis* resistance in humans. I performed alignment, QC and transcript quantification, created data visualizations, made statistical recommendations and co-wrote the manuscript. My analysis code was also used in a follow-up project (Simmons 2022).
Authors: Jason Simmons, Phu T. Van, *et al.*

*2019 Jul* **US10362237: Structured illumination system for increased dynamic range in quantitative imaging**

*United States Patent*

"The systems disclosed herein employ an iterative image collection strategy that utilizes structured illumination to achieve greater than 1,000,000-fold dynamic range measurements, representing a dramatic improvement over the prior art."
Inventors: Jonathan Minden, Frederick Lanni, Phu T. Van

*2018 Nov* **ggCyto: next generation open-source visualization software for cytometry**

*Bioinformatics*

This R package enables plotting of high-dimensional flow cytometry and mass cytometry data in the grammar-of-graphics style. I contributed to the R codebase, performed testing and co-wrote the manuscript.
Authors: Phu T. Van*, Wenxing Jiang*, Raphael Gottardo, Greg Finak (**co-first authors*)