

# Phu T. Van, PhD

email [pvan@alumni.cmu.edu](mailto:pvan@alumni.cmu.edu)  
website [ptvan.github.io](http://ptvan.github.io)

## WORK EXPERIENCE

TwinStrand Biosciences	2021Nov -	Bioinformatics Solutions Manager	Served as bioinformatics expert on company panel for approving new projects. Spearheaded "à la carte" bioinformatic offering to customers to increase revenue. Provided bioinformatics 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 new 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.
Fred Hutchinson Cancer Research Center	2014 - 2020	Postdoc → Bioinformatics Analyst	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	Designed and built a patented high-dynamic-range protein gel imager with robotic gel cutting arm for capturing rare proteins. Developed <b>SIGILab</b> , 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.
Carnegie Mellon University			

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

*Archives of Toxicology*

We characterized 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 count and spectra data. This collaboration with HealthCanada also resulted in the next publication (Cho 2023). Authors: Annette Dodge, Danielle LeBlanc, Andrew Williams, PHU T. VAN, *et al.*

2023 July      Error-corrected Duplex Sequencing enables direct detection and quantification of mutations in human TK6 cells with strong inter-laboratory consistency

*Mutation Research*

As part of the same HealthCanada collaboration above, we also studied ENU-exposed cultured human cells with a focus on dose-response and reproducibility. I made similar contributions to this project. Authors: Eunnara Cho, Carol Swartz, Andrew Williams, Miriam Rivas, Leslie Recio, Kristine Witt, Elizabeth Schmidt, Jeffery Yapple, Thomas Smith, 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. As the primary analyst on the project, I performed alignment, QC and transcript quantification, created data visualizations, made statistical recommendations and wrote parts of the manuscript. My analysis code was also used in a subsequent follow-up manuscript (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)