

Phu T. Van, PhD

email pvan@alumni.cmu.edu

website ptvan.github.io

SUMMARY

I'm seeking a scientific or technical position on collaborative teams working to solve data-intensive problems. I am particularly interested in projects that utilize statistical modeling or machine learning.

WORK EXPERIENCE

Fred Hutchinson Cancer Research Center	2015–Present	Systems Analyst, Analyst Group Leader
	Coordinated projects among the lab's 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. Performed reviews of junior analysts code, mentored PhD students and bench scientists on statistics. Reference: Evan Henrich · ehenrich@fredhutch.org	
Fred Hutchinson Cancer Research Center	2014–2015	Postdoctoral Research Fellow
	Developed R code to normalize mass cytometry data across experiments using multi-mixture models. Worked with engineers from FlowJo Inc. (Ashland, OR) to enable parsing of XML-encoded flow cytometry data in FlowJo v10. Performed dimension reduction on mass cytometry data and identified correlates of disease using regression. Reference: Evan Greene, PhD · palmergreene@gmail.com	
Carnegie Mellon University	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 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. Reference: Jonathan Minden, PhD · minden@cmu.edu	
Institute for Systems Biology	2006–2009	Research Associate
	Identified factors that affect peptide detectability in mass spectrometry experiments using R. Constructed a SQLServer database with web frontend for exploring mass spectrometry data and planning proteomic experiments. Developed a weighted regression algorithm in R to predict regulators of stress response in microarray data. Reference: Serdar Turkarslan, PhD · serdar.turkarslan@isbscience.org	

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, WDL, Slurm, Docker, GitHub
Bioinformatics	STAR, RSEM, BLAST, DEG, GSEA, KEGG, STRING

EDUCATION

2009-2014 Carnegie Mellon University

PhD, Biological Sciences

2001-2007 University of Washington

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

SELECTED PUBLICATIONS & PATENTS

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

in review Mar. 2020 Monocyte metabolic programs are associated with resistance to TST/IGRA conversion

This manuscript reports a potential link between oleic metabolism and *Tuberculosis* resistance in humans. As the primary analyst on the project, I aligned the raw transcriptomic data to the genome, quantified gene expression, created data visualizations and made recommendations on statistical/machine learning methods. I also assisted in writing the manuscript.
Authors: Jason Simmons, PHU T. VAN, *et al.*

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

Electrophoresis Sep. 2014 **High dynamic range proteome imaging with the structured illumination gel imager**
I built the Structured Illumination Gel Imager (SIGI), which enabled a 1000-fold increase in dynamic range over then-current methods while imaging 2D electrophoretic gels. I wrote the prototype imager software in Matlab, the production software in C/C++ and this manuscript, which informed the subsequent patent application for the imager.
Authors: PHU T. VAN, Victor Bass, Dan Shiwerski, Frederick Lanni, Jonathan Minden

Molecular Systems Biology Nov. 2010 **Coordination of frontline defense mechanisms under severe oxidative stress**
This manuscript reports a previously-undiscovered gene network that regulates cellular response to oxidative stress in microbes. I analyzed microarray data, wrote the R code for predicting the gene network, and co-wrote the manuscript.
Authors: Amardeep Kaur, PHU T. VAN, *et al.*

United States Patent Jul. 2019 **US10362237: Structured illumination system for increased dynamic range in quantitative imaging**
"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