

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

- | | |
|---|---|
| <i>Fred Hutchinson Cancer Research Center</i> | 2015–Present Systems Analyst, Analyst Group Leader Led the lab's PhD-level analysts in brainstorming sessions and journal club. Performed analyses on flow cytometry and transcriptomic datasets. Authored Statistical Analysis Plans & peer-reviewed manuscripts. Created bioinformatic analysis pipelines & interactive data visualizations. Mentored PhD students and bench scientists on statistics, performed reviews of junior analysts' code. Reference: Evan Henrich · ehenrich@fredhutch.org |
| <i>Fred Hutchinson Cancer Research Center</i> | 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. Reference: Evan Greene, PhD · palmergreene@gmail.com |
| <i>Carnegie Mellon University</i> | 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. Reference: Jonathan Minden, PhD · minden@cmu.edu |
| <i>Institute for Systems Biology</i> | 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 model regulators of stress response in microarray data. Reference: Serdar Turkarslan, PhD · serdar.turkarslan@isbscience.org |

TECHNICAL SKILLS

| | |
|---------------------------------------|---|
| <i>Programming</i> | R, Python, Matlab, C/C++, bash, JavaScript, Perl |
| <i>Data Analysis / Statistics</i> | dimension reduction, clustering, regression/classification, linear models, variable selection, hypothesis testing |
| <i>Databases</i> | MySQL/MariaDB, PostgreSQL, SQLServer |
| <i>Workflows</i> | Nextflow, WDL, Slurm, Docker, GitHub |
| <i>Bioinformatics</i> | 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>

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

in review

This manuscript reports a potential link between oleic metabolism and *Tuberculosis* resistance in humans. I was the primary bioinformatician on this project, responsible for analyzing data, creating visualizations and making statistical/machine learning recommendations. I also assisted in writing the manuscript.
Authors: Jason Simmons, PHU T. VAN, *et al.*

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

Bioinformatics

The ggCyto R package enables grammar-of-graphics plotting of high-dimensional flow cytometry and mass cytometry data. 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)

Sep. 2014 **High dynamic range proteome imaging with the structured illumination gel imager**

Electrophoresis

I built the Structured Illumination Gel Imager (SIGI), which enabled a 1000-fold increase in dynamic range while imaging 2D electrophoretic gels over then-current methods. 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

Nov. 2010 **Coordination of frontline defense mechanisms under severe oxidative stress**

Molecular Systems Biology

This manuscript reports a gene network that regulates microbial response to oxidative stress. I wrote the R code responsible for predicting the gene network, processed microarray data, and co-wrote the manuscript.
Authors: Amardeep Kaur, PHU T. VAN, *et al.*

Jul. 2019 **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