Phu T. Van, PhD

email	pvan@alumni.cmu.edu
website	ptvan.github.io

WORK EXPERIENCE

2021/11–   Bioinformatics Solutions Manager

TwinStrand Biosciences

Served as bioinformatics expert for the company in approving new projects. Spearheaded project to offer "à la carte" bioinformatics services to customers. Provided bioinformatics guidance to customers with Field&Technical Support department. Assisted Custom Applications department in creating new genomic panels. Supervised R&D collaborations between bioinformatic and wetlab technicians. Oversaw development of new R data analysis packages. Initiated and later supervised data analyses that expanded TwinStrand DuplexSeq(TM) support to two new sequencing platforms. Spearheaded data mining collaborations with the Data Engineering department and co-led company's Machine Learning Interest Group. Produced whitepapers and reviewed marketing materials for technical accuracy. Co-authored customer communication SOPs.
Reference: Clint Valentine III, MS · clint@twinstrandbio.com

2021/01-10   Bioinformatics Scientist II

TwinStrand Biosciences

Performed analyses on duplex sequencing data for numerous mutagenesis and cancer diagnostic projects. Reviewed and contributed R code to internal bioinformatic pipelines and customer-facing reports. Prepared data products and presented findings to industry and academic clients.
Reference: Fang Yin Lo, PhD · mailto:fangyinlo@gmail.com

2014–2020   Postdoc → Bioinformatics Analyst

Fred Hutchinson Cancer Research 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.
Reference: Amy Paguirigan, PhD · apaguiri@fredhutch.org

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.
Reference: Jonathan Minden, PhD · minden@cmu.edu

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, POSTERS & PATENTS
complete list: https://scholar.google.com/citations?user=IDadFEkAAAAJ

Aug. 2022   Characterisation of Procarbazine-induced Mutation Spectrum in the Bone Marrow of MutaMouse Males using Duplex Sequencing

Environmental & Molecular Mutagenesis conference

This poster characterizes genetic mutations in procarbazine-exposed mice using Duplex Sequencing as a potential replacement for the gold-standard Ames test. As the contract analyst to HealthCanada, I performed sequence alignment, variant calling and downstream analyses on the data. I also delivered data products to the client and aided in their preparation of the poster.
Authors: Anette Dodge, Danielle LeBlanc, Andrew Williams, Phu T. Van, *et al.*

Jun. 2021   Monocyte metabolic transcriptional programs associate with resistance to tuberculin skin test/interferon- 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 aligned the raw transcriptomic data to the genome, quantified gene expression, created data visualizations and made recommendations on downstream statistical/machine learning methods. I also wrote the computational parts of the manuscript.
Authors: Jason Simmons, 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

Nov. 2018   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*)