

Phu T. Van, PhD

Bioinformatic Scientist

<https://ptvan.github.io>, pvan@alumni.cmu.edu

EXECUTIVE SUMMARY

- * Bioinformatics expert with 20 years experience analyzing and integrating biological data (genomic, proteomic and flow+mass cytometry).
- * Strong track record of applying machine learning and statistical methods (*eg.* random forest, GLMs, network analysis, *etc.*) to model diseases and make predictions.
- * Experienced in managing scientific teams in commercial, non-profit and academic environments.
- * Communicate effectively with programmers, wetlab scientists and non-experts; excel in interdisciplinary teams working on complex, evolving projects.
- * Agile learner able to master new topics and procedures quickly.

TECHNICAL STRENGTHS

Programming	R, Python, bash, Matlab, C/C++, JavaScript, Perl
Data Analysis	dimension reduction, clustering, regression/classification
Statistics	linear models, variable selection, survival analysis, hypothesis testing
Databases	MySQL/MariaDB, PostgreSQL, SQLServer
Workflows	Nextflow, Slurm, Docker, GitHub
Cloud computing	EC2, ECR, S3
Markup/Interfaces	Streamlit, Shiny, LaTeX, Markdown, HTML
Bioinformatics	BWA, STAR, RSEM, BLAST, samtools, bcftools bedtools, deeptools, VardictJava, varscan, delly fastp, FastQC/MultiQC, GATK, IGV picard, bowtie2, limma, edgeR, MACS SRA, SIGNAL, COSMIC

EXPERIENCE

TwinStrand Biosciences <i>Senior Manager, Bioinformatics Solutions</i>	September 2023 - May 2024 <i>Seattle, WA</i>
--	---

- Spearheaded company's new bioinformatic services to customers with Sales&Marketing department.
- Analyzed NGS data, presented oncology and mutagenesis findings to customers, co-authored manuscripts.
- Served as bioinformatics expert on company committee that approves new research contracts.
- Provided bioinformatics guidance to company's Technical Support department and major customers.
- Developed and delivered internal bioinformatic training for wetlab scientists.
- Collected bioinformatic Voice of Customers to guide software development.
- Authored customer-facing whitepapers and marketing materials for Sales&Marketing department.
- Co-authored data-use policy for customer cloud platform with Legal and Sales&Marketing departments.

TwinStrand Biosciences <i>Bioinformatics Solutions Manager</i>	November 2021 - August 2023 <i>Seattle, WA</i>
--	---

- Analyzed NGS data that expanded company's DuplexSeqTM technology to 2 new sequencer platforms.
- Mentored junior bioinformatic scientists on R programming and data analysis.
- Oversaw development of new R data analysis software packages, contributed code.

- Performed QC and troubleshooting of new genomic panels with Custom Application department.
- Analyzed NGS data, presented oncology and mutagenesis findings to customers, co-authored manuscripts.
- Served as bioinformatics expert on company committee that approves new research contracts.
- Provided bioinformatics guidance to Sales&Marketing department and major external customers.
- Co-authored company's customer communication procedure with Technical Support department.

TwinStrand Biosciences
Bioinformatics Scientist II

January 2021 - October 2021
 Seattle, WA

- Performed end-to-end analysis of NGS data for 4 projects in mutagenesis and oncology.
- Prepared data products, presented findings and co-authored manuscripts for customers.
- Reviewed and contributed R code to internal bioinformatic pipelines and customer-facing reports.

Fred Hutchinson Cancer Center
Bioinformatics Analyst, Analyst Group Leader

October 2015 - December 2020
 Seattle, WA

- Coordinated analyses on multimodal data (proteomics+flow cytometry+transcriptomics) for predicting *Tuberculosis* progression.
- Created interactive data visualizations using Shiny, igraph and leaflet that identified novel *Tuberculosis* pathways.
- Performed differential gene expression and geneset enrichment analysis on RNASeq datasets.
- Performed variable selection and sensitivity analyses on flow-cytometry data from clinical trials.
- Led the lab's PhD-level analysts in brainstorming sessions and journal club.
- Authored Statistical Analysis Plans and reviewed manuscripts.
- Created bioinformatic analysis pipelines using R and Nextflow.
- Mentored PhD students and bench scientists on statistics, performed reviews of junior analysts' code, advised analyst hiring.

Fred Hutchinson Cancer Center
Postdoctoral Fellow

October 2014 - October 2015
 Seattle, WA

- Developed R code to normalize mass cytometry data across experiments using multi-mixture models.
- Worked with engineers from FlowJo Inc. (Ashland, OR) to develop code to parse XML-encoded flow cytometry data in FlowJo v10.
- Performed dimension reduction on mass cytometry data and identified correlates of disease for early clinical trial data.

Carnegie Mellon University
PhD Student

August 2009 - May 2014
 Pittsburgh, PA

- 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 imaging and protein extraction.
- Developed processes that improved reduction and alkylation of complex protein samples in preparation for mass spectrometric sequencing.
- Developed workflows to quantify protein abundances in 2DE gel images using bash.

Institute for Systems Biology
Research Associate

May 2006 - August 2009
 Seattle, WA

- 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 regression algorithm in R to model regulators of stress response using microarray data.

University of Washington

Computer Support Analyst

May 2006 - April 2009

Seattle, WA

- Designed and maintained public health questionnaires for the University of Washington Twin Registry (UWTR) and Native People for Cancer Control (NPCC).
- Provided computer tech support for 30+ users: installing software, troubleshooting problems, training new users.
- Managed organization's data security through Group Policy and ActiveDirectory Access Control Lists.

US Forest Service

Wildlife Biologist Trainee

May 2004 - August 2005

Colville, WA

- Performed field surveys in the Colville National Forest for habitats of Threatened, Endangered or Sensitive (TES) species.
- Upgraded field data recorders to latest USFS standards, introduced new software packages to District Planning Office and created accompanying manuals.
- Collaborated with USFS District specialists to complete an environmental assessment mandated by the National Environmental Protection Act.

EDUCATION

Carnegie Mellon University

PhD, Biological Sciences

2014

University of Washington

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

2007

CONFERENCES, PATENTS AND PUBLICATIONS

1. Shaofei Zhang, Stephanie L Coffing, William C Gunther, Michael L Homiski, Richard A Spellman, Phu Van, and Maik Schuler. Assessing the genotoxicity of n-nitrosodiethylamine with three in vivo endpoints in male big blue® transgenic and wild-type c 57 bl/6 n mice. *Environmental and Molecular Mutagenesis*, 65(6-7):190-202, 2024
2. Irina G Minko, Michael M Luzadder, Vladimir L Vartanian, Sean PM Rice, Megan M Nguyen, Monica Sanchez-Contreras, Phu Van, Scott R Kennedy, Amanda K McCullough, and R Stephen Lloyd. Frequencies and spectra of aflatoxin b1-induced mutations in liver genomes of neil1-deficient mice as revealed by duplex sequencing. *NAR Molecular Medicine*, page ugae006, 2024
3. B Kopp, A Khawam, K Di Perna, D Lenart, M Vinette, R Silva, TB Zanoni, C Rore, G Guenigault, E Richardson, et al. Liver-on-chip model and application in predictive genotoxicity and mutagenicity of drugs. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, 896:503762, 2024
4. Joel P Bercu, Shaofei Zhang, Zhanna Sobol, Patricia A Escobar, Phu Van, and Maik Schuler. 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/Genetic Toxicology and Environmental Mutagenesis*, 891:503685, 2023
5. Annette E Dodge, Danielle PM LeBlanc, Gu Zhou, Andrew Williams, Matthew J Meier, Phu Van, Fang Yin Lo, Charles C Valentine III, Jesse J Salk, Carole L Yauk, et al. Duplex sequencing provides detailed characterization of mutation frequencies and spectra in the bone marrow of

- mutamouse males exposed to procarbazine hydrochloride. *Archives of Toxicology*, pages 1–15, 2023
6. Eunnara Cho, Carol D Swartz, Andrew Williams, Miriam Rivas, Leslie Recio, Kristine L Witt, Elizabeth K Schmidt, Jeffry Yaplee, Thomas H Smith, Phu Van, et al. Error-corrected duplex sequencing enables direct detection and quantification of mutations in human tk6 cells with strong inter-laboratory consistency. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, page 503649, 2023
 7. Annette E Dodge, Danielle LeBlanc, Andrew Williams, Phu Van, Jake Higgins, Fang Yin Lo, Jeffry Yaplee, Clint C Valentine, Jesse J Salk, Carole L Yauk, et al. Characterisation of procarbazine-induced mutation spectrum in the bone marrow of mutamouse males using duplex sequencing. In *ENVIRONMENTAL AND MOLECULAR MUTAGENESIS*, volume 63, pages 92–93. WILEY 111 RIVER ST, HOBOKEN 07030-5774, NJ USA, 2022
 8. E Cho, F Marchetti, P Van, J Higgins, FY Lo, J Yaplee, CC Valentine III, J Salk, and CL Yauk. Optimal methodological design for duplex sequencing (tm) in tk6 cells determined through a time and concentration response analysis following enu treatment. In *Environmental and Molecular Mutagenesis*, volume 62, page 36, 2021
 9. D LeBlanc, M Meier, A Williams, J Buick, J Higgins, J Yaplee, P Van, FY Lo, CC Valentine, JJ Salk, et al. Duplex sequencing reveals an attenuated mutation frequency increase in the germ cells of mutamouse males exposed to n-ethyl-n-nitrosourea and benzo [a] pyrene relative to the tgr lacz assay. In *Environmental and Molecular Mutagenesis*, volume 62, page 87, 2021
 10. AE Dodge, D LeBlanc, A Williams, P Van, J Higgins, FY Lo, J Yaplee, CC Valentine, JJ Salk, and CL Yauk. Duplex sequencing (tm) reveals increases in mutation frequencies and c_i t transitions in the bone marrow of mutamouse males exposed to procarbazine. In *Environmental and Molecular Mutagenesis*, volume 62, page 87, 2021
 11. Jason D Simmons, Kimberly A Dill-McFarland, Catherine M Stein, Phu T Van, Violet Chihota, Thobani Ntshiq, Pholo Maenetje, Glenna J Peterson, Penelope Benchek, Mary Nsereko, et al. Monocyte transcriptional responses to mycobacterium tuberculosis associate with resistance to tuberculin skin test and interferon gamma release assay conversion. *Mosphere*, 7(3):e00159–22, 2022
 12. Gemma Moncunill, Jason Carnes, William Chad Young, Lindsay Carpp, Stephen De Rosa, Joseph J Campo, Augusto Nhabomba, Maxmillian Mpina, Chenjerai Jairoce, Greg Finak, et al. Transcriptional correlates of malaria in rts, s/as01-vaccinated african children: a matched case–control study. *Elife*, 11:e70393, 2022
 13. J. Simmons, P. Van, C. Stein, Chihota V., Velen K., Fielding K., Grant A., Ntshiq T., Peterson G., Graustein A., Nguyen F., Gottard R., Mayanja-Kizza H., Wallis R., Churchyard G., Boom W.H., and Hawn T.R. Monocyte metabolic transcriptional programs associate with resistance to tuberculin skin test/interferon- release assay conversion. *Journal of Clinical Investigations*, June 2021
 14. M. Blundon, V. Ganesan, B. Redler, P. T. Van, and J. S. Minden. Two-Dimensional Difference Gel Electrophoresis. *Methods Mol. Biol.*, 1855:229–247, 2019
 15. D. F. E. Ker, S. Eom, S. Sanami, R. Bise, C. Pascale, Z. Yin, S. I. Huh, E. Osuna-Highley, S. N. Junkers, C. J. Helfrich, P. Y. Liang, J. Pan, S. Jeong, S. S. Kang, J. Liu, R. Nicholson, M. F. Sandbothe, P. T. Van, A. Liu, M. Chen, T. Kanade, L. E. Weiss, and P. G. Campbell. Phase contrast time-lapse microscopy datasets with automated and manual cell tracking annotations. *Scientific Data*, 5:180237, 11 2018
 16. J. S. Minden, F. Lanni, and P. T. Van. USPTO US10362237: Structured illumination system for

increased dynamic range in quantitative imaging, July 2019

17. P. Van, W. Jiang, R. Gottardo, and G. Finak. ggCyto: next generation open-source visualization software for cytometry. *Bioinformatics*, 34(22):3951–3953, 11 2018
18. P. T. Van, V. Ganesan, V. Bass, A. Parthasarathy, D. Schlesinger, and J. S. Minden. In-gel equilibration for improved protein retention in 2DE-based proteomic workflows. *Electrophoresis*, 35(20):3012–3017, Oct 2014
19. P. T. Van, V. Bass, D. Shiwerski, F. Lanni, and J. Minden. High dynamic range proteome imaging with the structured illumination gel imager. *Electrophoresis*, 35(18):2642–2655, Sep 2014
20. U. K. Iheagwara, P. L. Beatty, P. T. Van, T. M. Ross, J. S. Minden, and O. J. Finn. Influenza virus infection elicits protective antibodies and T cells specific for host cell antigens also expressed as tumor-associated antigens: a new view of cancer immunosurveillance. *Cancer Immunology Research*, 2(3):263–273, Mar 2014
21. A. Kaur, P. T. Van, C. R. Busch, C. K. Robinson, M. Pan, W. L. Pang, D. J. Reiss, J. DiRuggiero, and N. S. Baliga. Coordination of frontline defense mechanisms under severe oxidative stress. *Molecular Systems Biology*, 6:393, Jul 2010
22. T. Koide, D. J. Reiss, J. C. Bare, W. L. Pang, M. T. Facciotti, A. K. Schmid, M. Pan, B. Marzolf, P. T. Van, F. Y. Lo, A. Pratap, E. W. Deutsch, A. Peterson, D. Martin, and N. S. Baliga. Prevalence of transcription promoters within archaeal operons and coding sequences. *Molecular Systems Biology*, 5:285, 2009
23. P. T. Van, A. K. Schmid, N. L. King, A. Kaur, M. Pan, K. Whitehead, T. Koide, M. T. Facciotti, Y. A. Goo, E. W. Deutsch, D. J. Reiss, P. Mallick, and N. S. Baliga. Halobacterium salinarum NRC-1 PeptideAtlas: toward strategies for targeted proteomics and improved proteome coverage. *Journal of Proteome Research*, 7(9):3755–3764, Sep 2008
24. A. K. Schmid, D. J. Reiss, A. Kaur, M. Pan, N. King, P. T. Van, L. Hohmann, D. B. Martin, and N. S. Baliga. The anatomy of microbial cell state transitions in response to oxygen. *Genome Research*, 17(10):1399–1413, Oct 2007