

Phu T. Van, PhD

Machine Learning Specialist

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

SUMMARY

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

TECHNICAL STRENGTHS

Programming	R, Python, Matlab, C/C++, bash, JavaScript, Perl
Data Analysis	dimension reduction, clustering, regression/classification
Statistics	linear models, variable selection, hypothesis testing
Databases	MySQL/MariaDB, PostgreSQL, SQLServer
Workflows	Nextflow, WDL, Slurm, Docker, GitHub
Markup/Interfaces	LaTeX, Markdown, Shiny, HTML
Bioinformatics	STAR, RSEM, BLAST, DEG, GSEA, KEGG, STRING

EXPERIENCE

Fred Hutchinson Cancer Research Center

October 2015 - Present

Systems Analyst, Bioinformatics Analyst Group Leader

Seattle, WA

- Performed differential gene expression and geneset enrichment analysis on transcriptomic datasets.
- Performed variable selection and sensitivity analyses on flow cytometry datasets.
- Led the lab's 3 PhD-level analysts in twice-monthly collaboration meeting and journal club.
- Coordinated analyses on data integration project with South African Tuberculosis Vaccine Institute.
- Authored Statistical Analysis Plans and peer-reviewed manuscripts.
- Created bioinformatic analysis pipelines using R, Nextflow and WDL.
- Created interactive data visualizations using RMarkdown, igrph and leaflet.
- Mentored PhD students and bench scientists on statistics, advised hiring and performed reviews of junior analysts' code.

Fred Hutchinson Cancer Research Center

October 2014 - October 2015

Postdoctoral Fellow

Seattle, WA

- Developed R code to normalize mass cytometry data across experiments using multi-mixture models.
- Developed R code for writing XML files to enable sharing of flow cytometry data.
- Worked with engineers from FlowJo Inc. (Ashland, OR) to enable XML parsing in FlowJo v10.
- Performed dimension reduction in mass cytometry data and identified correlates of disease progression in early-stage vaccine candidates.

Carnegie Mellon University

August 2009 - May 2014

PhD Student

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 acquisition of high-dynamic-range images.
- 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.

- 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 in microarray data.

EDUCATION

Carnegie Mellon University
Ph.D., Biological Sciences

2014

University of Washington

B.S., Biology (Physiology specialization), B.S., Wildlife Sciences

2007

REPRESENTATIVE PUBLICATIONS

1. 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
2. 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
3. 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
4. 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 Immunol Res*, 2(3):263–273, Mar 2014
5. 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. *Mol. Syst. Biol.*, 6:393, Jul 2010
6. 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. *Mol. Syst. Biol.*, 5:285, 2009
7. 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. *J. Proteome Res.*, 7(9):3755–3764, Sep 2008
8. 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 Res.*, 17(10):1399–1413, Oct 2007
9. 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