

BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: Ka Yee Yeung-Rhee (also known as **Ka Yee Yeung**)

eRA COMMONS USER NAME (credential, e.g., agency login): kayee2

POSITION TITLE: Virginia and Prentice Bloedel Endowed Professor, School of Engineering and Technology, University of Washington Tacoma, WA. Adjunct Professor, Department of Microbiology, School of Medicine, University of Washington, Seattle, WA.

EDUCATION/TRAINING

INSTITUTION AND LOCATION	DEGREE (if applicable)	Completion Date MM/YYYY	FIELD OF STUDY
University of Waterloo, Ontario, Canada	B.Math	05/95	Actuarial Science & Computer Science
University of Waterloo, Ontario, Canada	M.Math	08/96	Computer Science
University of Washington, Seattle, WA	M.S.	12/98	Computer Science
University of Washington, Seattle, WA	Ph.D.	12/01	Computer Science

Personal Statement

My research focuses on the development of optimized methods and software tools that will facilitate scalable analysis of diverse biomedical data. My research expertise includes bioinformatics, data science, cloud computing, statistics, molecular biology and reproducible research. I have an established track record working with high-throughput data for various biomedical applications, including cancer diagnostics, systems biology, and precision nutrition.

I teach undergraduate and graduate level computer science courses at UW Tacoma, including data structures, algorithms, and bioinformatics. I also have extensive experience mentoring research scientists, post-doctoral fellows, Ph.D., Master's and undergraduate students. In particular, I am strongly committed to enhancing the diversity of the research workforce.

B. Positions, Scientific Appointments, and Honors**Positions and Employment**

7/2024- present Virginia and Prentice Bloedel Endowed Professor, UW

9/2019-present Adjunct Professor, Department of Microbiology, UW, Seattle, WA

9/2019-present Professor, School of Engineering and Technology, UW, Tacoma, WA

9/2020-present CEO, BioDepot LLC, Seattle, WA

9/2021-6/2023 Associate Dean for Research and Innovation, School of Engineering and Technology, UW, Tacoma, WA

7/2014-8/2019 Adjunct Associate Professor, Department of Microbiology, UW, Seattle, WA

7/2014-8/2019 Associate Professor, School of Engineering and Technology, UW, Tacoma, WA

7/2011-6/2014 Research Associate Professor, Department of Microbiology, UW, Seattle, WA

6/2004-6/2011 Research Assistant Professor, Department of Microbiology, UW, Seattle, WA

Other Experience and Professional Memberships (selected)

2023- present Co-chair, Outreach Working Group, NIH MorPhiC program

2023 Reviewer, Multi-omics Data Analysis Coordination Center, NIH Multi-omics for Health and Disease Consortium

2022	Ad hoc reviewer, BRAIN Initiative Cell Atlas Network (BICAN) (ZMH1 ERB-L(04), NIH.
2021	Ad hoc reviewer, Small Business: Computational Modeling, and Biodata Management, NIH.
2017-2019	Elected academic leadership experience. Serve as the Vice Chair (2017-2018) and Chair (2018-2019) of Faculty Assembly, University of Washington Tacoma
2016-2018	Member of the NIH Commons Framework Working Group on Workflow Sharing and Docker Registry
2014	Ad hoc reviewer, Computational Analyses Exploiting Reference Epigenomic Maps, NIH.
2015-2018	Member of the NIH BD2K-LINCS Data Working Group

Honors

2019	Distinguished Research Award, University of Washington Tacoma
2002	Fast moving fronts in Computer Science, "Model-based Clustering and Data Transformations for Gene Expression Data", ISI Essential Science Indicators.
2003	Fast breaking paper in Computer Science, "Validating Clustering for Gene Expression Data", ISI Essential Science Indicators.

C. Contributions to Science

1. My early publications focused on *unsupervised machine learning methods* (clustering) for gene expression data. We addressed the question "which clustering algorithm should I use?" These publications provided a systematic framework for comparing clustering algorithms using gene expression data. We also showed the merit of using a probabilistic framework to cluster gene expression data.
 - a. **Yeung K. Y.**, Haynor D. R., Ruzzo W. L. Validating Clustering for Gene Expression Data. *Bioinformatics* 2001, 17:309-318.
 - b. **Yeung K.Y.**, Ruzzo W.L. Principal Component Analysis for Clustering Gene Expression Data. *Bioinformatics* 2001, 17:763-774.
 - c. **Yeung K. Y.**, Fraley, C., Murua A., Raftery A. E., Ruzzo W.L. Model-based Clustering and Data Transformations for Gene Expression Data. *Bioinformatics* 2001, 17:977-987.
 - d. Young W. C., **Yeung K. Y.**, Raftery A. E. Model-based clustering with data correction for removing artifacts in gene expression data. *Annals of Applied Statistics* 2017, 11(4): 1998-2026. PMID: PMC6364860.

2. I also developed *supervised machine learning* methods for bioinformatics applications. We have built on the Bayesian Model Averaging (BMA) framework. BMA is a multivariate variable selection (or feature selection) method that accounts for dependencies between variables, as well as uncertainty of model selection by averaging over multiple models. We have developed BMA methods to infer gene networks integrating multiple data sources. Gene networks summarize the relationships between biological entities. We formulate network inference as a series of variable selection problems. In other words, we aim to identify candidate regulators (i.e. parent nodes) from high-dimensional biological data for each target gene. Our network inference methods guide the search of regulators in the model space using prior probabilities derived from multiple data sources.
 - a. **Yeung K.Y.**, Dombek K.M., Lo K., Mittler J.E., Zhu J., Schadt E.E., Bumgarner R.E. and Raftery A.E. Construction of regulatory networks using time series microarray data in genotyped yeast segregants. *PNAS* 2011, 108(48): 19436 - 41. PMID: PMC3228453
 - b. Young W.C., Raftery A.E., **Yeung K.Y.** Fast Bayesian inference for gene regulatory networks using ScanBMA. *BMC Systems Biology* 2014, 8:47. PMID: PMC4006459
 - c. Hung L.H., Shi K., Wu M., Young W.M., Raftery A.E., **Yeung K.Y.** FastBMA: Scalable network inference and transitive reduction. *Gigascience*. 2017, 6(10):1-10. PMID: PMC5632288
 - d. Liang X., Young W.C., Hung L.H., Raftery A.E., **Yeung K.Y.** Integration of multiple data sources for gene network inference using genetic perturbation data. *Journal of Computational Biology* 2019, 26(10). PMID: PMC6786343.

3. I developed precision medicine approaches to analyze and integrate various 'omics' data. In particular, I have extensive experience collaborating with clinicians and oncologists to develop efficient computational methods and cloud-based software tools to analyze multi-omics data in cancer.

- a. Oehler V.G.*, **Yeung K.Y.***, Choi Y.E., Bumgarner R.E., Raftery A.E., and Radich J.P. The derivation of diagnostic markers of chronic myeloid leukemia progression from microarray data. *Blood* 2009; 114, 3292-3298. *Co-first authors. PMID: PMC2759651
 - b. **Yeung K.Y.**, Gooley T.A., Zhang A., Raftery A.E., Radich J.P., and Oehler V.G. Predicting relapse prior to transplantation in chronic myeloid leukemia by integrating expert knowledge and expression data. *Bioinformatics* 2012, 28(6): 823-830. PMID: PMC3307121
 - c. Sala-Torra O., Reddy S., Hung H.L., Beppu L., Wu D., Radich J., **Yeung K.Y.**, Yeung C.C.S. Rapid detection of myeloid neoplasm fusions using Single Molecule Long-Read Sequencing. *PLOS Global Public Health* 2023, 3(9):e0002267. PMID: PMC10497132.
 - d. Hung H.L., Fukuda B., Schmitz R., Hoang V., Lloyd W., **Yeung K.Y.** Harmonizing and integrating the NCI Genomic Data Commons through accessible, interactive, and cloud-enabled workflows. bioRxiv 10.1101/2022.08.11.503660. [Pre-print] 2024.
4. I also develop containerized software tools that are easy-to-use and easy-to-deploy to enhance reproducibility of bioinformatics analyses. My lab brings the latest technology to biomedical scientists by adding a graphical user interface to container-based tools, featuring gene network inference, image processing, next generation and long-read sequencing analyses as proof-of-concept examples.
- a. Mittal V., Hung L.H., Keswani J., Kristiyanto D., Lee S.B. and **Yeung K.Y.** GUIdock-VNC: Using a graphical desktop sharing system to provide a browser-based interface for containerized software. *Gigascience* 2017, 6(4): 1-6. PMID: PMC5530313.
 - b. Hung H.L., Hu J., Meiss T., Ingersoll A., Lloyd W., Kristiyanto D., Xiong Y., Sobie E.A., **Yeung K.Y.** Building containerized workflows using the BioDepot-workflow-Builder (BwB). *Cell Systems* 2019, 9(5): 508-514.E3. PMID: PMC6883158.
 - c. Reddy S., Hung H.L., Sala-Torra O., Radich J., Yeung C.C.S., **Yeung K.Y.** A graphical, interactive and GPU-enabled workflow to process long-read sequencing data. *BMC Genomics* 2021, 22:626. PMID: PMC8381503.
 - d. Hung H.L., Straw E., Reddy S., Schmitz R., Colburn Z., **Yeung K.Y.** Cloud-enabled Biodepot workflow builder integrates image processing using Fiji with reproducible data analysis using Jupyter notebooks. *Scientific Reports* 2022, 12:14920. PMID: PMC9440253.
5. In addition to facilitating reproducible bioinformatics analyses, my lab optimizes performance of big biomedical data workflows by leveraging algorithm design, implementation, deployment and systems architecture of computationally intensive tasks.
- a. Zhang P., Hung L.H., Lloyd W., **Yeung K.Y.** Hot-starting software containers for bioinformatics analyses. *Gigascience* 2018, 7(8) giy092. PMID: PMC6131214.
 - b. Hung L.H., Lloyd W., Sridhar R.A., Ravishankar S.D.A., Xiong Y., Sobie E. and **Yeung K.Y.** Holistic optimization of RNA-seq workflow for multi-threaded environments. *Bioinformatics* 2019, btz169. PMID: PMC6792127.
 - c. Niu X., Kumanov D., Hung H.L. Lloyd W., **Yeung K.Y.** Leveraging Serverless Computing to Improve Performance for Sequence Comparison. *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics 2019*, Pages 683-687. DOI 10.1145/3307339.3343465.
 - d. Hung H.L., Niu X., Lloyd W., **Yeung K.Y.** Accessible and interactive RNA sequencing analysis using serverless computing. bioRxiv 576199v2. [Pre-print] 2020.

Complete list of publications in MyNCBI:

<https://www.ncbi.nlm.nih.gov/myncbi/ka%20yee.yeung-rhee.1/bibliography/public/>