Kevin Wang Ph.D.

Data Scientist and Statistician

US: (+1) 619-215-3725
AU: (+61) 404-955-255
kevinwangstats@gmail.com
Portfolio: kevinwangstats.com

Experience

Now **Senior Data Scientist**, *Illumina*

2021 Nov
- Led bioinformatics research and development for multiple pioneering cancer companion diagnostics programs. Oversaw high-value data deliverables which directly contributed to multimillion dollar contract and regulatory milestones.
- Designed, deployed and maintained organizational-wide bioinformatics and data science tools in Python and R. These tools were foundational infrastructure for high-throughput clinical product development. These tools had significant improvements on the overall efficiency of product R&D and reduced the data verification processes from several months to hours.
- Spearheaded a comprehensive review of organizational data management practices for the clinical and medical operations teams. Developed and formalized a scalable framework to enhance data velocity and data standardization. This streamlined decision-making and data verification bottlenecks that ultimately fast-tracked critical data deliverables.
- Facilitated cross-functional collaboration among R&D scientists, software developers, regulatory authorities and pharmaceutical partners to achieve regulatory milestones for IVD products in both US and EU. This leadership ensured alignment across diverse teams, facilitating successful product approvals.
- Developed machine learning algorithms to analyze large-scale genomics datasets, uncovering clinically-actionable insights that significantly enhanced the performance of oncology and prenatal testing products.
- Led and authored a high-impact publication for a clinical oncology product's benefits for blood cancer.

2021 Oct **Statistician**, *CSL Behring*

2020 Jul
- Direct contributions to the R&D of high-value pharmaceutical products like the COVID-19 vaccines and treatments during the pandemic.
- Designed and formalized statistical process control techniques to standardize practices in the manufacturing and quality control laboratory. These tools are deployed through interactive dashboards which had a wide organizational impact.
- Provided actionable insights to multiple products which mitigated significant regulatory risks and achieved significant financial savings, with some up to multimillion dollars production batches.

2020 Jun **Research Associate**, *University of Sydney*

2019 Sep
- Developed an interpretable machine learning framework for predicting patient clinical outcomes using omics data. Melanoma was a key focus of this research and the key biomarkers were validated and implemented on a cost-effective assay.
- `R` software packages associated with the research were developed and maintained in a reproducible environment.
- Visiting scholar at Cornell University for 2 months. Orchestrated operations, designed contents and hosted a workshop in open-source single-cell analytics leveraging cloud computing.
- Published peer-reviewed articles in top-tier journals, including PNAS, Nature Digital Medicine and Nature Computational Science.

2019 Mar **Postgraduate Teaching Fellow**, *University of Sydney*

2016 Mar
- Delivered lectures (200+ students) and tutorials, covering 15 different courses, including statistics, mathematics and data science. Achieved 85% satisfactory rating for end-of-course surveys.
- Authored and designed course content at both the undergraduate and postgraduate levels.
- Mentored and trained junior statistics tutors.

Education

2016–2020 **Doctor of Philosophy in Statistical Bioinformatics**, *University of Sydney*

- Developed machine learning methods to solve statistical challenges in patient clinical outcome predictions and thereby enabling personalised medicine.
- Real-world biomedical data from collaborators were used to motivate and articulate important challenges in current statistical research. The value of this thesis is the synthesis of statistics, clinical implementation and practical deployment.

2012-2015 **Bachelor of Science (Adv. Mathematics) (Hon. I)**, *University of Sydney*

- Major in statistics and financial mathematics.
- The Honours thesis examined functional MRI data and the inferred brain connectivity networks through the use of different statistical association measures.

Skills

`python` Developed computational modules for regulated products at Illumina to support clinical studies. Highly proficient in writing scripts and notebooks to solve analytical problems for stakeholders. Experience in using and applying `pandas`, `numpy`, `scikit-learn`, `keras`, `pytorch`, and `openai` for a range of projects.

`R` Highly proficient in scripting and developing packages. Strong ability to use `tidyverse` and `shiny` for the data ETL process to produce visualisations and presentable insights. Author and maintainer of several open-source `R` packages, including `APES`, `learningtower`, `mcvis` and `scMerge`.

`git` & CI/CD & docker Highly proficient in a reproducible environment that supports business operations and software development requirements. Experience in holding reproducible workshops with excellent participant feedback.

Applied statistics Highly experienced at the research level and consulting level. Ability to transform business and cross-disciplinary problems into a statistical framework and deliver solutions. Strong experience with model selection to extract explainable insights from data.

Statistical modelling Highly experienced at the research level and consulting level. Specialisation in predictive modelling for high-dimensional data.

Other skills SQL, cloud and dashboard development for business use.

Volunteering

2021-Present **Mentor**, *Monash University*

Supervised and mentored three Master of Data Science students with the students successfully completed their research projects. Developed high-quality R-CRAN package and published academic journal article.

2021-2023 **Councillor**, *Statistical Society of Australia*

Served as the Assistant Secretary and Communication Officer in the NSW and Victorian Branch of the Society. Organised public events, designed automated management systems, responsible for member communications and social media account maintenance.

Publications

1 **Wang, K. Y. X.**, Pupo, G.M., Tembe, V., Patrick, E., Strbenac, D., Schramm, S.J., Thompson, J.F., Scolyer, R.A., Muller, S., Tarr, G. and Mann, G.J., Yang, J.Y.H. **2022**. Cross-Platform Omics Prediction procedure: a statistical machine learning framework for wider implementation of precision medicine. Nature Digital Medicine, 5(1) 1-10.

2 Kim J.H., **Wang, K. Y. X.**, Chen, C., Lin, Y., Tam, P.P.L., Lin, D.M., Yang, J.Y.H., & Yang, P. **2021**. Cepo uncovers cell identity through differential stability. *Nature Computational Science*, **1** 784–790.

3 Schafer, S., **Wang, K. Y. X.**, Sundling, F., Yang, J.Y.H., & Liu, A. **2021**. Modelling maternal and perinatal risk factors to predict poorly controlled childhood asthma. *PLOS ONE* 16(5): e0252215.

4 Lin, C., **Wang, K. Y. X.**, & Mueller, S. **2020**. mcvis: A new framework for collinearity discovery, diagnostic and visualization. *Journal of Computational and Graphical Statistics*, 1-13.

5 Hewavisenti, R., Ferguson, A., **Wang, K. Y. X.**, Jones, D., Gebhardt, T., Edwards, J., Zhang, M., Britton, W., Yang, J., Hong, A., & Palendira, U. **2020**. CD103+ tumour-resident CD8+ T cell numbers underlie improved patient survival in oropharyngeal squamous cell carcinoma. *Journal for ImmunoTherapy of Cancer*, 8:e000452.

6 **Wang, K.Y.X.**, Tarr, G., Yang, J.Y.H., Mueller, S. **2019**. Fast and approximate exhaustive variable selection for generalised linear models with APES, Invited paper to *Australia & New Zealand Journal of Statistics*, 61 (4) 445-465.

7 Lin, Y., Ghazanfar, S., **Wang, K.Y.X.**, Gagnon-bartsch, J.A., Lo, K.K., Han, Z., Ormerod, J.T., Speed, T.P., Yang, P., Yang, J.Y.H. **2019**. scMerge: Leveraging factor analysis, stable expression and pseudo-replication to merge multiple single-cell RNA-seq data, *Proceedings of the National Academy of Sciences of the United States of America*, 116 (20) 9775-9784.

8 Pires da Silva, I., **Wang, K.Y.X.**, Wilmott, J.S., Holst, J., Carlino, M.S., Park, J.J., Quek, C., Wongchenko, M., Yan, Y., Mann, G., Johnson, D.B., McQuade, J.L., Rai, R., Kefford, R.F., Rizos, H., Scolyer, R.A., Yang, J.Y.H., Long, G. V, Menzies, A.M. **2019**. Distinct molecular profiles and immunotherapy treatment outcomes of V600E and V600K BRAF-mutant melanoma. *Clinical Cancer Research*, 25 (4) 1272-1279.

9 **Wang, K.Y.X**., Menzies, A.M., Silva, I.P., Wilmott, J.S., Yan, Y., Wongchenko, M., Kefford, R.F., Scolyer, R.A., Long, G. V, Tarr, G., Mueller, S., Yang, J.Y.H. **2019**. bcGST - an interactive bias-correction method to identify over-represented gene-sets in boutique arrays. *Bioinformatics*, 35 (8) 1350-1357.

10 Strbenac, D., **Wang, K.Y.X.**, Wang, X., Dong, J., Mann, G.J., Mueller, S., Yang, J.Y.H. **2019**. Melanoma Explorer: a web application to allow easy reanalysis of publicly available and clinically-annotated melanoma omics datasets. *Melanoma Research*.