Kevin Wang Ph.D.

Data Scientist and Statistician

(+61) 404 955 255
kevinwangstats@gmail.com
Portfolio: kevinwangstats.com
Nationality: Australian

Experience

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

2021 Nov
- Leading the bioinformatics strategies for three pioneering in-vitro diagnostic (IVD) cancer companion diagnostic products on the TruSight Oncology platform. These strategies and analytical insights were instrumental in executing deliverables, highlighting a strong capability to enhance efficiency in critical projects.
- Engaging in collaboration with a team of interdisciplinary experts, including internal R&D scientists, software developers, regulatory bodies and external pharmaceutical partners. Navigated the projects to meet critical multi-million milestone payments, showcasing adeptness in team leadership and stakeholder/project management.
- Employing innovative machine learning algorithms to extract critical new insights from large scale genomics data, thereby significantly boosting product efficacy in oncology diagnostics and prenatal testing. Highlighting technical expertise and innovation in developing commercially viable solutions.
- Crafting and deploying robust, production-grade analytical tools and software using Python and R. Contributions towards efficient data management strategies were pivotal in securing the reliability and scalability of the company's data solutions.

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

2020 Jul
- Used statistical process control techniques to standardise manufacturing and quality control laboratory procedures with engineers and biochemists.
- Designed experiments and used innovative machine learning models and R-Shiny dashboards to understand the chemical stability of products, conform to regulatory requirements and ensure product effectiveness.
- Worked with high-value products like COVID vaccines and treatments, provided actionable insights to achieve savings of multiple product batches, some up to $1M.

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

2019 Sep
- The major research project developed an interpretable machine learning framework for predicting patient clinical outcome using omics data. Melanoma was a key focus of this research and the key biomarkers were validated and implemented on a cost-effective bioassay.
- R software packages associated with the research were developed and maintained in a reproducible environment.
- Visiting scholar at Cornell University for 2 months and delivered a workshop in open-source single-cell analytics using Cloud technology.
- 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` (ChatGPT) 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 under a reproducible environment that support business operations and software development requirements. Proficient in building docker images and deployment through Google Cloud. 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

2021-2023 **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 accounts maintenance.

Selected Awards

2019 **Statistical Society of Australia, NSW**, *JB Douglas Award (runner-up). Annual competitive prize awarded to the top statistics research PhD candidate in the State.*

2019 **Statistical Society of Australia**, *Golden Jubilee Travel Grant. Annual competitive travel funding for a statistics PhD candidate in Australia.*

2017 **International Biometric Society - Australasian Region**, *Best student talk at the biennial conference.*

2016-2020 **Australian Postgraduate Award**, *for the duration of the PhD program at the University of Sydney.*

2015 **International Biometric Society - Australasian Region**, *Honours scholarship award. Annual award to the top biostatistics student undertaking a Honours program in AU/NZ region.*

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*.