RICK M. TANKARD PhD

rickmtankard@gmail.com ricktankard.com linkedin.com/in/ricktankard

Experienced Statistician and Bioinformatician with a PhD in Bioinformatics and a profound commitment to advancing human disease research. Specialises in developing robust, reusable code, with extensive expertise in R package development and complex bash scripting. Passionate about applying scientific rigour and an analytical approach to solve complex problems in human biology. Eager to contribute to multidisciplinary teams that work on innovative projects that impact disease care and treatment. Brings a blend of technical proficiency, scientific curiosity, and a collaborative spirit with aspirations of a career in professional bioinformatics and data science.

Employment

Freelance Data Scientist February 2024 – present
Benchmarking large language models (LLMs) for data science in Python.

Career break – health February 2023 – present

The Walter and Eliza Hall Institute of Medical Research (WEHI) **Victoria, Australia**
Bioinformatics Analyst 18 July 2022 – 17 January 2023

As a Bioinformatics Analyst at WEHI, I efficiently managed the enhancement and debugging of an R package and a Perl script, which are integral for genomic analysis. My focused efforts successfully resolved two-thirds of the identified bugs, aligning the software with the latest human genome updates. My strategic planning ensured the continuity and effectiveness of these tools for future genomic research endeavours.

Career break – Health January 2021 – July 2022

Murdoch University **Western Australia, Australia**
Research Associate 12 June 2017 – 20 December 2020

I led multiple bioinformatics research projects, collaborating effectively with interstate and overseas teams. My role involved analysing extensive datasets (60TB) using servers and high-performance computing. I skilfully utilised R 'tidyverse' packages and the 'glmnet' machine learning tool for identifying genetic markers related to Alzheimer's disease and epigenetic markers for age prediction. This work culminated in presentations at four scientific symposiums and contributed to eight peer-reviewed articles. Additionally, I imparted my knowledge in statistics as a tutor for first-year university students, blending research and educational roles.

The Walter and Eliza Hall Institute of Medical Research (WEHI) **Victoria, Australia**
Research Technician 2 December 2010 – 10 June 2017

I enhanced the efficiency of next-generation sequencing data processing by developing a bash script-based pipeline, significantly reducing analysis time for large sample sets. This improvement showcased my technical foresight, considering the potential future use of Nextflow for similar tasks. Additionally, I effectively managed our division's weekly seminar series for two years, demonstrating strong organisational skills.

Education

2013–2018 **Doctor of Philosophy (PhD)** in bioinformatics/medical research
The University of Melbourne / The Walter and Eliza Hall Institute of Medical Research (WEHI), VIC
Thesis title: *Identifying disease-causing short tandem repeat expansions in massively parallel sequencing data, focusing on ataxias.*
http://hdl.handle.net/11343/197796
I developed a quantitative algorithm (exSTRa (R and Perl), see software list below) to detect repeat expansions from next-generation sequencing data. These methods can speed up the diagnosis of genetic repeat expansion disorders when whole-genome sequencing is performed.

2006–2010 **Bachelor of Science (Degree with Honours)** (Mathematics and Statistics)
The University of Melbourne / WEHI, VIC
Grade: H1 – First class honours

2007–2009 **Diploma of Arts** (History and Philosophy of Science)
The University of Melbourne, VIC

Technical Skills

- R, R Studio, package development, unit testing and Rmarkdown (11 years)
- Foundational SQL, Python and Shiny (in R)
- High-performance computing (SLURM, Torque PBS) (4 years)
- Scientific/bioinformatics workflows with Nextflow and Snakemake (1.5 years)
- Statistical models and hypothesis testing, such as general linear models, elastic net regression, power calculations, ANOVA
- Machine learning: XGBoost, KNN, model validation, deep learning
- Version control with git and GitHub/GitLab (8 years)
- Next-generation sequencing analysis tools (including bowtie2, GATK, Picard)
- Programming in bash and Perl (11 years)
- Linux (command line), macOS (work/PhD) and Windows (hobby computer)
- Continuous Integration (Travis CI / GitHub Actions) (2 years)
- Containers (Docker and Singularity) (2 years)
- Adobe Illustrator and Indesign for figures and posters
- Microsoft Office, report writing in Word, finance tracking in Excel, PowerPoint

Volunteer activities

2018– Member of the *Statistical Society of Australia*
Held **Treasurer** position for the VIC branch from 2022 to present.
Held **Secretary** position for the WA branch in 2019 and 2020.

2020 **Organiser** for the *WA Young Statisticians Workshop 2020* (Statistical Society of Australia event), including preparing the website https://ysw2020.netlify.app/

2019 *EMBL ABR Australian BioCommons Hands-on Workshop: Implementing Scalable Bioinformatic Workshops in Snakemake & Nextflow*. **Perth hub facilitator**: helped researchers with their problems as they came up during the workshop. Participate in training during the weeks before in Adelaide.

2019 *Resbaz Perth 2019* **committee member**. I helped organise the Docker stream. I taught git stream as a last-minute replacement.

2008–2011 *Melbourne University Dancesport Club* **committee member**, responsible for the website, e-mail lists, class supervision and cash handling.

2007 & 2009 *Professor Harry Messel International Science School* **Staff**
Assisted in recording and publishing lectures and took responsibility for groups of high school students.

Scholarships

2013–17	Australian postgraduate award (APA)	$24,653 per year
2010	Alan W Harris Honours Scholarship	$5,000
2010	Maurice Belz Scholarship	$7,000
2006–09	Melbourne Access Scholarship	$4,161 per year and BSc course fee waiver

Open-source software

exSTRa R and Perl libraries to detect repeat expansions in next-generation sequencing data. Primary software output from PhD thesis at WEHI. Performed profiling post-PhD to optimise the package drastically. https://github.com/bahlolab/exSTRa

Linkdatagen and VCF2linkdatagen (software updates)
Performed bug fixes, updated to support more platforms, improved Illumina TOP/BOT interpretation to make more SNPs available for linkage analysis and updated documentation. http://bioinf.wehi.edu.au/software/linkdatagen/

rwarrior Game designed to teach the R language interactively.
https://github.com/trickytank/Rwarrior

Publications and presentations

Seven oral presentations and ten poster presentations at conferences. 21 peer-reviewed publications.
Notable publications:

Tankard RM, Bennett MF, Degorski P, Delatycki MB, Lockhart PJ, Bahlo M. Detecting Expansions of Tandem Repeats in Cohorts Sequenced with Short-Read Sequencing Data. *Am J Hum Genet*. 2018;103(6):858-873. doi:10.1016/j.ajhg.2018.10.015

Hildebrand MS, **Tankard R**, Gazina EV, et al. (2015), PRIMA1 mutation: a new cause of nocturnal frontal lobe epilepsy. *Ann Clin Transl Neurol*, 2: 821-830. doi:10.1002/acn3.224

Bahlo M, Bennett MF, Degorski P, **Tankard RM**, Delatycki MB, Lockhart PJ. Recent advances in the detection of repeat expansions with short-read next-generation sequencing. F1000Res. 2018;7:F1000 Faculty Rev-736. Published 2018 Jun 13. doi:10.12688/f1000research.13980.1

See https://scholar.google.com/citations?user=AKoK1swAAAAJ&hl=en for a list of all publications.