Dave **Tang**

Bioinformatician at BrightPath Biotherapeutics

Life Innovation Center 412, 3-25-22 Tonomachi, Kawasaki

me@davetang.org | davetang.org | davetang | davetang

I am a bioinformatician experienced in analysing high-throughput sequencing data. I have developed start-to-finish bioinformatics pipelines for the processing and analysis of RNA-seq and DNA-seq datasets. Recently, I started to work in the immunogenomics field and developed a fully fledged neoantigen prediction pipeline. I am familiar with machine/statistical learning and their best practices and have applied these techniques in predicting immunogenicity. I am also familiar with Linux systems and their administration, high-performance computing technologies, and cloud computing. I write about these topics and more in my bioinformatics blog.

Education

The University of Queensland — *Brisbane, Australia*

Bachelor of Science (Honours) — *2001-2005*

- Majoring in Biochemistry and Microbiology

Vrije University — *Amsterdam, the Netherlands*

PhD in Computational Biology — *2010-2015*

- PhD completed through a Marie Curie training network (BrainTrain) as a Marie Curie Early Career Researcher.

Academic and Work Experience

BrightPath Biotherapeutics — *Tokyo, Japan*

Bioinformatician — *April 2020 - Current*

- Developing the clinical neoantigen prediction pipeline for the personalised cancer peptide vaccine program.
- Providing bioinformatics support, systems administration of local HPC servers, and reviewing scientific publications.

RIKEN — *Yokohama, Japan*

Visiting Scientist — *May 2015 - March 2020*

- Elucidating the potential role of a human minisatellite as an alternative transcription start site.
- Examining the potential role of tissue and cell type specific expression of repetitive DNA elements.

The University of Western Australia — *Perth, Australia*

Post-doctoral Research Associate — *Jul. 2017 - Nov. 2019*

- Analysing accession specific expression patterns in Arabidopsis thaliana single cell RNA-seq data.
- Developing a single cell RNA-seq computational approach for demultiplexing mixed genotype datasets.

Telethon Kids Institute — *Perth, Australia*

Post-doctoral researcher — *May 2015 - June 2017*

- Developing and implementing high-throughput sequencing analysis workflows for diagnosing rare genetic diseases.
- Generating a reference genotype database of an Australian Aboriginal population.

RIKEN — *Yokohama, Japan*

Research Associate and PhD candidate — *April 2010 - May 2015*

- Characterising a new class of universal non-coding RNAs (ncRNAs) involved in DNA Damage Repair.
- Analysis on the expression of repetitive DNA elements and how they contribute to the genesis of ncRNAs.
- Discovery of reverse transcriptase artefacts and developing a computational method for the removal of these artefacts.

University of Queensland — *Brisbane, Australia*

Bioinformatics Assistant — *August 2008 - February 2010*

- The development and maintainence of a Laboratory Information Management System
- Providing server and database administration for the lab.

Commonwealth Scientific and Industrial Research Organisation — *Brisbane, Australia*

Bioinformatics Assistant — *September 2006 - July 2008*

- Generating a reference set of tRNAs for the Bovine Genome Sequencing and Analysis Consortium.
- Assisting with the design of the first ovine SNP chip for the International Sheep Genomics Consortium.

Honours and Awards

Current	Top 70 Bioinformatics Blogs and Websites for Bioinformaticians	*Feedspot*
2010	CSIRO Chairman's Medal	*CSIRO*
2008	CSIRO Partnership Excellence Award	*CSIRO*

Grants Awarded

2018	Western Australia Department of Health Near-Miss Merit Award (Chief Investigator)	*AUD 50,000*
2017	Dissecting the leukaemia microenvironment (Chief Investigator)	*AUD 43,395*
2016	Targeting four-stranded DNA conformations to modulate gene expression in breast cancer (Chief Investigator)	*AUD 47,669*
2015	Translating NextGen Sequencing for the Diagnosis of Developmental Anomalies and Rare Diseases (Associate Investigator)	*AUD 192,505*
2009	BrainTrain: Integrative neuroscience school on brain function and disease	*USD 4,400,000*

Presentations and Administrative Experiences

2019	Invited bioinformatics workshop presenter at BioC Asia	*Sydney*
2019	Session chair at BioC Asia	*Sydney*
2019	Invited speaker at Laboratory Medicine Congress and Exhibition (LMCE2019)	*Busan*
2018	UWA Hacky Hour member	*Perth*
2017	Invited presenter at the Australian Bioinformatics and Computational Biology Society (ABACBS) meeting	*Perth*
2015	Invited MODHEP workshop presenter on analysing CAGE data	*Yokohama*
2013	Organiser of the RIKEN Yokohama Chat with Guest sessions	*Yokohama*
2013	Organiser of the RIKEN Yokohama Student Journal Club	*Yokohama*
2013	Organising committee for the BrainTrain conference	*Yokohama*
2013	Organiser of the Identifying Regulatory Elements in the Genome workshop in RIKEN	*Yokohama*
2012	Session chair for the Patients and Medicines Forum	*Yokohama*
2012	Invited presenter at the RIKEN OSC bioinformatics course	*Yokohama*

Workshop and Course Attendances

2019	BioC 2019: Where Software and Biology Connect	*New York*
2016	AGHA Workshop: reducing morbidity and mortality from genetic disease through screening	*Perth*
2016	Consumer and Community Involvement in Research workshop	*Perth*
2015	UQ winter school in mathematical and computational biology	*Brisbane*
2014	RIKEN/KI doctoral course: Employing Genome-wide Technologies for Functional Regulation in Development and Disease	*Stockholm*
2013	AMATA conference ECR workshop	*Gold Coast*
2013	Data analysis course from Johns Hopkins University	*Coursera*
2013	BrainTrain courses and workshops	*Europe*
2013	RIKEN/KI doctoral course: Epigenomics: Methods and Applications to Disease and Development	*Yokohama*
2012	SISSA summer school on dopaminergic neurons	*Trieste*
2012	RIKEN/KI doctoral course: Functional Regulation in Development and Disease	*Stockholm*
2011	UQ winter school in mathematical and computational biology	*Brisbane*
2011	RIKEN/EBI bioinformatics roadshow	*Yokohama*

Bioinformatic Skills

A list of bioinformatic datasets and tools that I have experience analysing and employing.

Data Analysis: DNA-seq, RNA-seq, CAGE-seq, sRNA-seq, ChIP-seq, ATAC-seq, scRNA-seq, HLA typing, CCLE, GTEx, TCGA, ENCODE.

Software Knowledge: Docker, GATK4, Picard, WDL, Cromwell, Bpipe, SAMtools, [VB]CFtools, BEDTools, deepTools, GEMINI, VEP, SnpEff, SnpSift, R Shiny, Cell Ranger, Seurat, Monocle, HISAT2, StringTie, STAR, RSEM, HOMER, Conda, SPAdes, HISAT-genotype, Kourami, HLA-LA, IEDB related tools, phASER, WhatsHap, bam-readcount, RepeatMasker, BLAT, BLAST, UCSC Genome Browser.

R packages: Tidyverse packages, tidymodels packages, core Bioconductor packages, biomaRt, data.table, Gviz, plotly, grid, gridExtra, UpSetR, pheatmap, dendextend, fgsea, edgeR, workflowr, parallel, vcfR.

Programming Languages: Perl, R, Python, Bash scripting.

Editorial Activities

Referee/Reviewer for: National Health and Medical Research Council, BMC Genomics, PeerJ, and Scientific Reports.

Referees

Brian Dalrymple Honorary Research Fellow, Institute of Agriculture, The University of Western Australia, Australia, `brian.dalrymple@uwa.edu.au`

Piero Carninci Principal Investigator, RIKEN Yokohama Campus, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan, `carninci@riken.jp`

Peer Reviewed Publications

1. Oliva, M., Stuart, T., Tang, D., Pflueger, J., Poppe, D., Jabbari, J., Gigante, S., & …. (2022). An environmentally responsive transcriptional state modulates cell identities during root development. *bioRxiv*.
2. Jamieson, S., Fakiola, M., Tang, D., Scaman, E., Syn, G., Francis, R., & …. (2021). Common and rare genetic variants that could contribute to severe otitis media in an australian aboriginal population. *Clinical Infectious Diseases*.
3. Anderson, D., Skut, P., Hughes, A., Ferrari, E., Tickner, J., Xu, J., Mullin, B., & …. (2020). The bone marrow microenvironment of pre-b acute lymphoblastic leukemia at single-cell resolution. *Scientific Reports*.
4. Lassmann, T., Francis, R., Weeks, A., Tang, D., Jamieson, S., Broley, S., & …. (2020). A flexible computational pipeline for research analyses of unsolved clinical exome cases. *NPJ Genomic Medicine*.
5. Bertuzzi, M., Tang, D., Calligaris, R., Vlachouli, C., Finaurini, S., Sanges, R., & …. (2020). A human minisatellite hosts an alternative transcription start site for NPRL3 driving its expression in a repeat number-dependent manner. *Human Mutation*.
6. Rothacker, K., Ayers, K., Tang, D., Joshi, K., Van Den Bergen, J., & …. (2018). A novel, homozygous mutation in desert hedgehog (DHH) in a 46, XY patient with dysgenetic testes presenting with primary amenorrhoea: A case report. *International Journal of Pediatric Endocrinology*.
7. Tang, D., Fakiola, M., Syn, G., Anderson, D., Cordell, H., Scaman, E., Davis, E., & …. (2018). Arylsulphatase a pseudodeficiency (ARSA-PD), hypertension and chronic renal disease in aboriginal australians. *Scientific Reports*.
8. Hon, C., Ramilowski, J., Harshbarger, J., Bertin, N., Rackham, O., Gough, J., & …. (2017). An atlas of human long non-coding RNAs with accurate 5′ ends. *Nature*.
9. Roudnicky, F., Dieterich, L., Poyet, C., Buser, L., Wild, P., Tang, D., & …. (2017). High expression of insulin receptor on tumour-associated blood vessels in invasive bladder cancer predicts poor overall and progression-free survival. *The Journal of Pathology*.
10. Abraham, M., Li, D., Tang, D., O'Connell, S., McKenzie, F., Lim, E., & …. (2017). Short stature and hypoparathyroidism in a child with kenny-caffey syndrome type 2 due to a novel mutation in FAM111A gene. *International Journal of Pediatric Endocrinology*.
11. Baynam, G., Pachter, N., McKenzie, F., Townshend, S., Slee, J., Kiraly-Borri, C., & …. (2016). The rare and undiagnosed diseases diagnostic service–application of massively parallel sequencing in a state-wide clinical service. *Orphanet Journal of Rare Diseases*.
12. Tang, D., Anderson, D., Francis, R., Syn, G., Jamieson, S., Lassmann, T., & …. (2016). Reference genotype and exome data from an australian aboriginal population for health-based research. *Scientific Data*.
13. Hasegawa, Y., Tang, D., Takahashi, N., Hayashizaki, Y., Forrest, A., & …. (2014). CCL2 enhances pluripotency of human induced pluripotent stem cells by activating hypoxia related genes. *Scientific Reports*.
14. Tang, D., Plessy, C., Salimullah, M., Suzuki, A., Calligaris, R., Gustincich, S., & …. (2013). Suppression of artifacts and barcode bias in high-throughput transcriptome analyses utilizing template switching. *Nucleic Acids Research*.
15. Francia, S., Michelini, F., Saxena, A., Tang, D., De Hoon, M., Anelli, V., Mione, M., & …. (2012). Site-specific DICER and DROSHA RNA products control the DNA-damage response. *Nature*.
16. Saxena, A., Tang, D., & Carninci, P. (2012). piRNAs warrant investigation in rett syndrome: An omics perspective. *Disease Markers*.
17. Thiagarajan, R., Georgas, K., Rumballe, B., Lesieur, E., Chiu, H., & …. (2011). Identification of anchor genes during kidney development defines ontological relationships, molecular subcompartments and regulatory pathways. *PloS One*.

18. Thiagarajan, R., Cloonan, N., Gardiner, B., Mercer, T., Kolle, G., & …. (2011). Refining transcriptional programs in kidney development by integration of deep RNA-sequencing and array-based spatial profiling. *BMC Genomics*.

19. Chiu, H., Szucsik, J., Georgas, K., Jones, J., Rumballe, B., Tang, D., & …. (2010). Comparative gene expression analysis of genital tubercle development reveals a putative appendicular Wnt7 network for the epidermal differentiation. *Developmental Biology*.

20. Kijas, J., Townley, D., Dalrymple, B., Heaton, M., Maddox, J., McGrath, A., & …. (2009). A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PloS One*.

21. Georgas, K., Rumballe, B., Valerius, M., Chiu, H., Thiagarajan, R., & …. (2009). Analysis of early nephron patterning reveals a role for distal RV proliferation in fusion to the ureteric tip via a cap mesenchyme-derived connecting segment. *Developmental Biology*.

22. Cloonan, N., Xu, Q., Faulkner, G., Taylor, D., Tang, D., Kolle, G., & …. (2009). RNA-MATE: A recursive mapping strategy for high-throughput RNA-sequencing data. *Bioinformatics*.

23. Tang, D., Glazov, E., McWilliam, S., Barris, W., & Dalrymple, B. (2009). Analysis of the complement and molecular evolution of tRNA genes in cow. *BMC Genomics*.