

Pengyi Yang

Office address: 5W96, Charles Perkins Centre,
School of Mathematics and Statistics,
The University of Sydney, NSW 2006, Australia

Email: pengyi.yang@sydney.edu.au

Lab website: <http://pengyiyang.github.io>

PROFESSIONAL POSITIONS

2020 – 2024	NHMRC Investigator	The University of Sydney School of Mathematics and Statistics, NSW, Australia
2020 – 2023	USyd Robinson Fellowship	The University of Sydney School of Mathematics and Statistics, NSW, Australia
2019 – present	Group Leader	Children's Medical Research Institute Westmead, NSW, Australia
2017 – present	ARC DECRA Fellow & Senior Lecturer	The University of Sydney School of Mathematics and Statistics, NSW, Australia
2016 – 2017	Lecturer	The University of Sydney School of Mathematics and Statistics, NSW, Australia
2015 – 2017	USyd Fellow (DVCR)	The University of Sydney School of Mathematics and Statistics, NSW, Australia
2012 – 2015	Research Fellow	National Institutes of Health (NIH), Systems Biology Group, Epigenetics & Stem Cell Biology Laboratory; Biostatistics and Computational Biology Branch, NIEHS, NC, USA
2011 – 2012	Visiting Postgraduate	Garvan Institute of Medical Research, Diabetes and Obesity Program, NSW, Australia
2009 – 2012	NICTA Ph.D. student	National ICT Australia (NICTA), Australian Technology Park, NSW, Australia

EDUCATION

2009 – 2012	Ph.D., Bioinformatics	The University of Sydney, Australia
2005 – 2008	M.E., Computer Science	Southwest University, China
2001 – 2005	B.S., Computer Science	Southwest University, China

HONOURS AND AWARDS

2020 – 2024	Investigator (Emerging Leader 2), National Health and Medical Research Council (NHMRC)
2020 – 2023	Robinson Fellowship (DVC Research), USyd, Australia
2017	J G Russell Award, Australian Academy of Science, Australia
2017	Finalist of Eureka Prize for Outstanding Early Career Researcher, Australian Museum
2017 – 2019	Discovery Early Career Researcher Award (DECRA), Australian Research Council (ARC)

2015 – 2017	University of Sydney Postdoctoral Fellowship (DVC Research), USyd, Australia
2015	Fellows Award for Research Excellence, National Institutes of Health, USA
2014	Paper of the Year Award, National Institute of Environmental Health Sciences, National Institutes of Health, USA
2014	Fellows Award for Research Excellence, National Institutes of Health, USA
2012 – 2016	Research Fellowship, National Institutes of Health, USA
2009 – 2012	National ICT Australia (NICTA) Research Project Award (NRPA), Australia
2009 – 2012	National ICT Australia (NICTA) International Postgraduate Award (NIPA), Australia
2009	Best Thesis Award on Master of Engineering, Chongqing Education Commission, China
2008	Student Travel Award, International Conference on Pattern Recognition in Bioinformatics (PRIB 2008), Melbourne, Australia
2005 – 2008	National Scholarship Award on Master of Engineering, Ministry of Education of the People's Republic of China

SUMMARY OF RESEARCH INCOME

Source	Value	Duration	Role	Note
NHMRC Investigator (Emerging Leader 2)	1.5M (AUD)	2020 – 2024	Sole CI	The success rate of NHMRC Investigator at EL2 is 9.1% in the funding year of 2019
Robinson Fellowship	625K (AUD)	2020 – 2023	Sole CI	Ten such fellowships are awarded first ever in USyd in 2018.
National Health and Medical Research Council (NHMRC) Project Grant	479K (AUD)	2017 – 2019	CI B	A success rate of 15.2% in the funding year of 2017.
Discovery Projects (DP), ARC	354K (AUD)	2017 – 2019	CI D	A success rate of 17.8% in the funding year of 2017.
J G Russell Award, Australian Academy of Science	6K (AUD)	2017	Recipient	Four awardees in the entire nation for the highest quality of DECRA application judged in comparison to other DECRA applications in the year of 2017 (200 funded and 1197 considered).
Discovery Early Career Researcher Award (DECRA), Australian Research Council (ARC)	371K (AUD)	2017 – 2019	Sole CI	A success rate of 16.7% in the funding year of 2017. I relinquished my USyd Postdoctoral Fellowship in 2017 to take on this role.
University of Sydney (USyd) Postdoctoral Fellowship	273K (AUD)	2015 – 2017	Sole CI	Ten such fellowships are awarded each year in USyd. I relinquished my Research Fellowship in NIH and joined USyd in September 2015.
Fellows Award for Research Excellence (FARE), NIH, USA	1000 USD each	2014, 2015	Recipient	I won FARE consecutively in 2014 and 2015. It is a NIH-wide award compete by all research fellows across 27 institutes. The success rate was ~17% NIH-wide.
Research Fellowship, National Institutes of Health (NIH), USA	256K (USD)	2012 – 2016	Sole CI	In NIH, this is a senior postdoctoral position in contrast to Visiting Fellowship which is an entry level postdoctoral position.

INVITED ORAL PRESENTATIONS

- Oct 2019 Sydney University Bioinformatics Seminar, Charles Perkins Centre, USyd, Australia
- Sep 2019 Centenary Institute, Sydney, Australia
- Aug 2019 Children's Medical Research Institute (CMRI), Westmead, Australia
- Nov 2018 School of Biotechnology and Biomolecular Sciences, USNW, Australia
- Nov 2018 Computational Biology Fest II, Research School of Biology, ANU, Australia
- Oct 2018 Biological Domain Seminar, Charles Perkins Centre, USyd, Australia
- Sep 2018 Biomedical Sciences Seminar Series, College of Veterinary Medicine, Cornell University, USA
- Sep 2018 Single-Cell Seminar, Department of Mathematics and Statistics, University of Michigan, USA
- Aug 2018 Epigenetic & Stem Cell Biology Laboratory (ESCBL) Seminar, National Institutes of Health, USA
- Jul 2018 Oz Single Cell Conference, Garvan Institute of Medical Research, Sydney, Australia
- Jun 2018 Sydney Bioinformatics Research Symposium, University of Technology Sydney, Australia
- May 2018 John Curtin School of Medical Research, Australian National University, Canberra, Australia
- Nov 2017 Sydney-Fudan Symposium on Precision Medicine, Shanghai, China
- Sep 2017 Children's Medical Research Institute (CMRI), Westmead, NSW, Australia
- Aug 2017 The 26th International Joint Conference on Artificial Intelligence, Melbourne, Australia
- Apr 2017 Hunter Meeting on System, Cells and Development, Hunter Valley, NSW, Australia
- Feb 2017 Network of Mind Workshop, Panellist, University of Sydney, Sydney, Australia
- Nov 2016 Core Research Facilities Symposium, Panellist, Sydney Nanoscience Hub, Sydney, Australia
- Jul 2016 Walter and Eliza Hall Institute (WEHI) Bioinformatics Seminar, Melbourne, Australia
- Apr 2016 CSIRO Bioinformatics Seminar, Sydney, Australia
- Oct 2015 Victor Chang Cardiac Research Institute (VCCRI), Sydney, NSW, Australia
- Oct 2015 The Australian Bioinformatics And Computational Biology Society (ABACBS) conference, Garvan Institute of Medical Research, NSW, Australia
- Sep 2014 Laboratory of Molecular Carcinogenesis (LMC) Seminar Series, National Institutes of Health, USA
- Jun 2013 Chromatin, Metabolism, and Development (CMD) Meeting, National Institutes of Health, USA
- Apr 2013 The 17th Pacific-Asia Conference on Knowledge Discovery and Data Mining, Cold Coast, Australia
- Nov 2012 Bioinformatics User Group (BUG) Meeting, Garvan Institute of Medical Research, NSW, Australia
- Jan 2012 The 10th Asia Pacific Bioinformatics Conference, Melbourne, Australia
- May 2011 The 15th Pacific-Asia Conference on Knowledge Discovery and Data Mining, Shenzhen, China
- Jan 2011 The 9th Asia Pacific Bioinformatics Conference, Incheon, Korea
- Sep 2009 International Conference on Bioinformatics, Matrix, Biopolis, Singapore
- Oct 2008 The 3rd International Conference on Pattern Recognition in Bioinformatics, Melbourne, Australia

PUBLICATIONS

Full List in Reverse Chronological Order

Total peer-reviewed articles: **51**; First and/or corresponding author: **33/51** (65%)

Impact factor (IF) is derived from 5-year 2017 Journal Citations Reports (*Thomson Reuters*)

Notation:

Peer-Reviewed Journal Articles

1. Lin Y, Ghazanfar S, Strbenac D, Wang A, Patrick E, Lin D, Speed T, Yang J*, **Yang P*** (2019) Evaluating stably expressed genes in single cells. [GigaScience](#), 8, 10.1093/gigascience/giz106
2. Oldfield A, Henriques T, Kumar D, Burkholder A, Cinghu S, Paulet D, Bennett B, **Yang P**, Scruggs B, Lavender C, Rivals E, Adelman K, Jothi R (2019) NF-Y controls fidelity of transcription initiation at gene promoters through maintenance of the nucleosome-depleted region. [Nature Communications](#), 10, 3072.
3. **Yang P***, Humphrey S†, Cinghu S†, Pathania R, Oldfield A, Kumar D, Perera D, Yang J, James D, Mann M, Jothi R* (2019) Multi-omic profiling reveals dynamics of the phased progression of pluripotency. [Cell Systems](#), 8(5), 427-445.
 – Recommended by F1000 Prime, doi:10.3410/f.735727099.793563652
4. Lin Y, Ghazanfar S, Wang K, Gagnon-Bartsch J, Lo K, Su X, Han Z, Ormerod J, Speed T, **Yang P***, Yang J* (2019) scMerge leverages factor analysis, stable expression, and pseudoreplication to merge multiple single-cell RNA-seq datasets. [Proceedings of the National Academy of Sciences of the United States of America](#), 116(20), 9775-9784.
5. Kim T, Chen I, Parker B, Humphrey S, Crossett B, Cordwell S, **Yang P***, Yang J* (2019) QCMAP: An interactive web-tool for performance diagnosis and prediction of LC-MS Systems. [Proteomics](#), <https://doi.org/10.1002/pmic.201900068>
6. Parker B†, Calkin A†, Seldin M†, Keating M, Tarling E, **Yang P**, Moody S, Liu Y, Zerenturk E, Needham, E, Jayawardana K, Pan C, Mellet N, Weir J, Lazarus R, Lusi A, Meikle P, James D, Vallim T, Drew B (2018) A proteome- and lipidome-wide systems genetic analysis of hepatic lipid metabolism. [Nature](#), 567, 187-193 [IF 44.9]
7. Kim T, Chen I, Lin Y, Wang A, Yang J, **Yang P*** (2018) Impact of similarity metrics on single-cell RNA-seq data clustering. [Briefings in Bioinformatics](#) 10.1093/bib/bby076. [IF 7.1]
8. Ridder M, Klein K, Yang J, **Yang P**, Lagopoulos J, Hickie I, Bennett M, Kim J (2018) An uncertainty visual analytics framework for fMRI functional connectivity. [Neuroinformatics](#), 17(2), 211-223 [IF 3.5]
9. **Yang P***, Ormerod J, Liu W, Ma C, Zomaya A, Yang J (2018) AdaSampling for positive-unlabeled and label noise learning with bioinformatics applications. [IEEE Transactions on Cybernetics](#), 49(5), 1932-1943. [IF 8.8]
10. Fazakerley D, Chaudhuri R, **Yang P**, Maghzal G, Thomas K, Krycer J, Humphrey S, Parker B, Fisher-Wellman K, Meoli C, Hoffman N, Diskin C, Burchfield J, Cowley M, Kaplan W, Modrusan Z, Kolumam G, Yang H, Chen D, Samocha-Bonet D, Greenfield J, Hoehn K, Stocker R, James D (2018) Mitochondrial CoQ deficiency is a common driver of mitochondrial oxidants and insulin resistance. [eLIFE](#), 7, e3211. [IF 8.4]
11. Cinghu S†, **Yang P†**, Kosak J, Conway A, Kumar D, Oldfield A, Adelman K, Jothi R (2017) Intragenic enhancers attenuate host gene expression. [Molecular Cell](#), 68(1), 104-117. [IF 14.4]
 – Highlighted in *Nature Reviews Genetics*, doi:10.1038/nrg.2017.90, 2017
 – Highlighted in *Nature Reviews Molecular Cell Biology*, doi:10.1038/nrm.2017.111, 2017
12. Norris DM†, **Yang P†**, Krycer JR, Fazakerley DJ, James DE, Burchfield JG (2017) An improved Akt reporter reveals intra- and inter-cellular heterogeneity and oscillations in signal transduction. [Journal of Cell Science](#), 130, 2757-2766. [IF 5.2]
13. **Yang P***, Oldfield A, Kim T, Yang A, Yang J, Ho J* (2017) Integrative analysis identifies co-dependent gene expression regulation of BRG1 and CHD7 at distal regulatory sites in embryonic stem cells. [Bioinformatics](#), 33(13), 1916-1920. [IF 8.0]
14. Zheng X, **Yang P#**, Lackford B, Bennett B, Wang L, Li H, Wang Y, Miao Y, Foley J, Fargo D, Jin Y, Williams C, Jothi R, Hu G (2016) CNOT3-dependent mRNA deadenylation safeguards the pluripotent state. [Stem Cell Reports](#), 7(5), 897-910. [IF 7.5]

15. Minard A, Tan S, **Yang P[#]**, Fazakerley D, Domanova W, Parker B, Humphrey S, Jothi R, Stöckli J, James D (2016) mTORC1 is a major regulatory node in the FGF21 signaling network in adipocytes. [Cell Reports](#), 17(1), 29-36. [IF 8.7]
16. **Yang P**, Patrick E, Humphrey SJ, Ghazanfar S, Jothi R, James DE, Yang YH (2016) Phosphoproteomics data annotation using hypothesis driven kinase perturbation analysis. [Proteomics](#), 16(13), 1868-1871. [IF 3.8]
17. **Yang P^{*}**, Humphrey SJ, James DE, Yang YH, Jothi R^{*} (2016) Positive-unlabeled ensemble learning for kinase substrate prediction from dynamic phosphoproteomics data. [Bioinformatics](#), 32(2):252-259. [IF 8.0]
18. Lu C, Wang J, Zhang Z, **Yang P**, Yu G (2016). NoisyGOA: Noisy GO annotations prediction using taxonomic and semantic similarity. [Computational Biology and Chemistry](#), 65, 203-211. [IF 1.4]
19. Domanova W, Krycer J, Chaudhuri R, **Yang P**, Vafaee F, Fazakerley D, Humphrey S, James D, Kuncic Z, (2016). Unraveling kinase activation dynamics using kinase-substrate relationships from temporal large-scale phosphoproteomics studies. [PLoS One](#), 11(6), e0157763. [IF 3.4]
20. **Yang P^{*}**, Zheng X, Jayaswal V, Hu G, Yang YH, Jothi R^{*} (2015) Knowledge-based analysis for detecting key signaling events from time-series phosphoproteomics data. [PLoS Computational Biology](#), 11(8):e1004403. [IF 5.0]
21. Hoffman N, Parker B, Chaudhuri R, Fisher-Wellman K, Kleinert M, Humphrey S, **Yang P**, Holliday M, Trefely S, Fazakerley D, Stockli J, Burchfield J, Jensen T, Jothi R, Kiens B, Wojtaszewski J, Richter E, James DE (2015) Global phosphoproteomic analysis of human skeletal muscle reveals a network of exercise-regulated kinases and AMPK substrates. [Cell Metabolism](#), 22(5):922-935. [IF 19.8]
 - Recommended by Faculty of 1000 Biology
22. Pathania R, Ramachandran S, Elangovan S, Padia R, **Yang P[#]**, Cinghu S, Veeranan-Karmegam R, Fulzele S, Pei L, Chang C-S, Choi J-H, Shi H, Manicassamy S, Prasad PD, Sharma S, Ganapathy V, Jothi R, Thangaraju M (2015) DNMT1 is essential for mammary and cancer stem cell maintenance and tumorigenesis. [Nature Communications](#), 6:6910. [IF 13.1]
 - Recommended by Faculty of 1000 Biology
23. Oldfield AJ[†], **Yang P[†]**, Conway AE, Cinghu S, Freudenberg JM, Yellaboina S, Jothi R (2014) Histone-fold domain protein NF-Y promotes chromatin accessibility for cell type-specific master transcription factors. [Molecular Cell](#), 55(5):708-722. [IF 14.4]
24. **Yang P[†]**, Patrick E[†], Tan SX, Fazakerley DJ, Burchfield J, Gribben C, Prior MJ, James DE, Yang YH^{*} (2014) Direction pathway analysis of large-scale proteomics data reveals novel features of the insulin action pathway. [Bioinformatics](#), 30(6):808-814. [IF 8.0]
25. Ma X, **Yang P[#]**, Kaplan WH, Lee BH, Wu LE, Yang YH, Yasunaga M, Sato K, Chisholm DJ, James DE (2014) ISL1 regulates PPAR γ activation and early adipogenesis via BMP4-dependent and independent mechanisms. [Molecular and Cellular Biology](#), 34(19):3607-3617. [IF 4.8]
26. Lackford B, Yao C, Charles GM, Weng L, Zheng X, Choi E, Xie X, Wan J, Xing Y, Freudenberg JM, **Yang P**, Jothi R, Hu G, Shi Y (2014) Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal. [EMBO Journal](#), 33(8):878-889. [IF 9.9]
27. **Yang P^{*}**, Yoo PD, Fernando J, Zhou BB, Zhang Z, Zomaya AY (2014) Sample subset optimization techniques for imbalanced and ensemble learning problems in bioinformatics applications. [IEEE Transactions on Cybernetics](#), 44(3):445-455. [IF 8.8]
28. Humphrey SJ, Yang G, **Yang P[#]**, Fazakerley DJ, Stockli J, Yang YH, James DE (2013) Dynamic adipocyte phosphoproteome reveals Akt directly regulates mTORC2. [Cell Metabolism](#), 17(6):1009-1020. [IF 19.8]
29. **Yang P**, Humphrey SJ, Fazakerley DJ, Prior MJ, Yang G, James DE, Yang YH^{*} (2012) Re-Fraction: a machine learning approach for deterministic identification of protein homologs and splice variants in large-scale MS-based proteomics. [Journal of Proteome Research](#), 11(5):3035-3045. [IF 4.4]
30. **Yang P^{*}**, Ma J, Wang P, Zhu Y, Zhou BB, Yang YH^{*} (2012) Improving X!Tandem on peptide identification from mass spectrometry by self-boosted Percolator. [IEEE/ACM Transactions on Computational Biology and Bioinformatics](#), 9(5):1273-1280. [IF 1.9]

31. Wang P, **Yang P**, Yang YH (2012) OCAP: an open comprehensive analysis pipeline for iTRAQ. [*Bioinformatics*, 28\(10\):1404-1405. \[IF 8.0\]](#)
32. **Yang P**^{†,*}, Ho JWK[†], Yang YH, Zhou BB* (2011) Gene-gene interaction filtering with ensemble of filters. [*BMC Bioinformatics*, 12:S10. \[IF 3.5\]](#)
33. **Yang P**^{*}, Ho JWK, Zomaya AY, Zhou BB* (2010) A genetic ensemble approach for gene-gene interaction identification. [*BMC Bioinformatics*, 11:524. \[IF 3.5\]](#)
34. Wang P, **Yang P**, Arthur J, Yang YH (2010) A dynamic wavelet-based algorithm for pre-processing mass spectrometry data. [*Bioinformatics*, 26\(18\):2242-2249. \[IF 8.0\]](#)
35. Yoo PD, Ho YS, Ng J, Charleston M, Saksena NK, **Yang P**, Zomaya AY (2010) Hierarchical kernel mixture models for the prediction of AIDS disease progression using HIV structural gp120 profiles. [*BMC Genomics*, 11:S4. \[IF 4.3\]](#)
36. **Yang P**^{*}, Zhang Z, Zhou BB, Zomaya AY (2010) A clustering based hybrid system for biomarker selection and sample classification of mass spectrometry data. [*Neurocomputing*, 73:2317-2331. \[IF 3.2\]](#)
37. **Yang P**^{*}, Zhou BB, Zhang Z, Zomaya AY (2010) A multi-filter enhanced genetic ensemble system for gene selection and sample classification of microarray data. [*BMC Bioinformatics*, 11:S5. \[IF 3.5\]](#)
38. **Yang P**^{*}, Xu L, Zhou BB, Zhang Z, Zomaya AY (2009) A particle swarm based hybrid system for imbalanced medical data sampling. [*BMC Genomics*, 10:S34. \[IF 4.3\]](#)
39. **Yang P**^{*}, Zhang Z* (2009) An embedded two-layer feature selection approach for microarray data analysis. [*IEEE Intelligent Informatics Bulletin*, 10:24-32](#)
40. Zhang Z, **Yang P**, Wu X, Zhang C (2009) An agent-based hybrid system for microarray data analysis. [*IEEE Intelligent Systems*, 24\(5\):53-63. \[IF 4.0\]](#)
41. Zhang Z*, **Yang P**^{*} (2008) An ensemble of classifiers with genetic algorithm based feature selection. [*IEEE Intelligent Informatics Bulletin*, 9:18-24](#)

Peer-Reviewed Articles in Conference Proceedings

42. **Yang P**, Liu W, Yang J (2017) Positive unlabeled learning via wrapper-based adaptive sampling. [*Proceedings of the 26th International Joint Conference on Artificial Intelligence \(IJCAI\)*, 3273-3279.](#)
43. **Yang P**, Liu W, Zhou BB, Chawla S, Zomaya AY (2013) Ensemble-based wrapper methods for feature selection and class imbalance learning. [*Proceedings of the 17th Pacific-Asia Conference on Knowledge Discovery and Data Mining \(PAKDD\)*. Lecture Notes in Artificial Intelligence 7818, Springer, 544-555](#)
44. **Yang P**, Zhang Z, Zhou BB, Zomaya AY (2011) Sample subsets optimization for classifying imbalanced biological data. [*Proceedings of the 15th Pacific-Asia Conference on Knowledge Discovery and Data Mining \(PAKDD\)*. Lecture Notes in Artificial Intelligence 6635, Springer, 333-344](#)
45. Li L, **Yang P**, Qu L, Zhang Z, Cheng P (2010) Genetic algorithm-based multi-objective optimisation for QoS-aware web services composition. [*Proceedings of the 4th International Conference on Knowledge Science, Engineering and Management \(KSEM\)*. Lecture Notes in Artificial Intelligence 6291, Springer, 549-554](#)
46. **Yang P**, Tao L, Xu L, Zhang Z (2009) Multiagent framework for bio-data mining. [*Proceedings of the Fourth International Conference on Rough Set and Knowledge Technology \(RSKT\)*. Lecture Notes in Computer Science 5589, Springer, 200-207](#)
47. **Yang P**, Zhang Z (2008) A clustering based hybrid system for mass spectrometry data analysis. [*Proceedings of Pattern Recognition in Bioinformatics \(PRIB\)*. Lecture Notes in Bioinformatics 5265, Springer, 98-109](#)
48. **Yang P**, Zhang Z (2008) A hybrid approach to selecting susceptible single nucleotide polymorphisms for complex disease analysis. [*Proceedings of BioMedical and Engineering Informatics \(BMEI\)*. IEEE, 214-218](#)
49. **Yang P**, Zhang Z (2007) Hybrid methods to select informative gene sets in microarray data classification. [*Proceedings of the 20th Australian Joint Conference on Artificial Intelligence \(AI\)*. Lecture Notes in Artificial Intelligence 4830, Springer, 811-815](#)

Review Articles

50. **Yang P***, Yang YH, Zhou BB, Zomaya AY (2010) A review of ensemble methods in bioinformatics. [*Current Bioinformatics*](#), 5(4):296-308 [IF **0.78**]

Book Chapters

51. **Yang P**, Yang YH, Zhou BB, Zomaya AY (2013) Stability of feature selection algorithms and ensemble feature selection methods in bioinformatics. In [*Biological Knowledge Discovery Handbook: Preprocessing, Mining and Postprocessing of Biological Data*](#), Wiley, New Jersey, USA, 333-352

TEACHING

- Lecturer, BioInfoSummer 2019, Workshop on phosphoproteomics data analysis, Sydney, Australia
- Primary Lecturer, Molecular Systems Biology (QBIO3001), Semester 1, 2020 (forthcoming), School of Life and Environmental Sciences, USyd, Australia
- Primary Lecturer, Molecular Systems Biology (QBIO2001), Semester 1, 2018 and 2019, School of Life and Environmental Sciences, USyd, Australia
- Guest Lecturer, Data Science Capstone (DATA3888), Semester 2, 2019, School of Mathematics and Statistics, USyd, Australia
- Primary Lecturer, Computational Statistical Methods (STAT5003), Semester 2, 2016-2018, School of Mathematics and Statistics, USyd, Australia
- Lecturer, EMBL Australia PhD Course 2018, Predictive models and their bioinformatics applications, UNSW, Sydney, Australia
- Guest Lecturer, Nature of Systems (HTIN5001), Semester 1, 2017, School of IT, USyd, Australia
- Lecturer, BioInfoSummer 2015, Introduction to R Programming, Functions, and Visualisation, Sydney, Australia
- Lecturer, Introduction to Biostatistics & Bioinformatics Short Courses, 2014, NIH/NIEHS, USA
- Tutor, Knowledge Discovery and Data Mining (COMP5318), 2009 and 2010, School of IT, USyd, Australia
- Tutor, Systems Analysis and Modelling (INFO2110), 2009, School of IT, USyd, Australia

SUPERVISION

- Alan Burzevski, Dalyell Scholars Program (2019), Westmead, USyd
- Isaac Shipsey, Exchange student (2018-2019), SMS, USyd (Primary supervisor)
- Yue Cao, PhD student (2019-present), SMS, USyd (Auxiliary supervisor)
- Hani Kim, PhD student (2018-present), SMS, USyd (Primary supervisor)
- Taiyun Kim, PhD student (2018-present), SMS, USyd (Primary supervisor)
- Thomas Geddes, PhD student (2017-present), School of Life and Environmental Sciences (SoLES), USyd (Primary supervisor)
- Hao Wu, PhD candidate (2016-present), School of IT, USyd (Auxiliary supervisor)
- Mengbo Li, MPhil student (2018-present), SMS, USyd (Associate supervisor)
- Kukulege Dinuka Perera, Talented Student Program (2016 and 2017), SoLES, USyd
- Chendong Ma, Honours 2016, SMS, USyd (co-supervision with Prof. Jean Yee Hwa Yang)
- Kevin Shrestha, Honours 2016, School of IT, USyd (co-supervision with A/Prof. Bing B. Zhou)
- Taiyun Kim, Summer Scholarship 2015, Electrical Engineering, UNSW (co-supervision with Dr. Joshua Ho)
- Steven Phan, Summer Scholarship 2015, CPC, USyd (co-supervision with Dr. Rima Chaudhuri)
- Shila Ghazanfar, Summer Research Scholarship (The Centre of Obesity, Diabetes and Cardiovascular Disease) 2012, SMS, USyd (contributing to supervision with Prof. Jean Yang)

PROFESSIONAL SERVICE

Journal & Grant Reviewer

- Grants** **Australian Research Council (ARC):** Discovery Project (DP); Discovery Early Career Researcher Award (DECRA); Industrial Transformation Training Centres (IC); Linkage Infrastructure, Equipment and Facilities (LE)
- National Health and Medical Research Council (NHMRC):** Project Grants
- Breast Cancer Now, United Kingdom
- Terry Fox Research Grants, United Arab Emirates
- Journals** (Full list available from Publons: <https://publons.com/a/292352>)
- Bioinformatics; Briefings in Bioinformatics; BMC Bioinformatics;
- Cell Systems; Cell Reports; Nature Communications; Genome Biology;
- PLoS Computational Biology; Molecular BioSystems; Journal of Proteome Research;
- European Journal of Human Genetics; PLoS One; Knowledge and Information Systems;
- IEEE Transactions on Knowledge and Data Engineering; Neurocomputing;
- Artificial Intelligence in Medicine

Conference Program Committee & Reviewer

- 2019 **Program Committee Co-chair** – ABACBS conference, Dec 9-11, Sydney, Australia
- 2019 **Program Committee** – GIW conference, Dec 9-11, Sydney, Australia
- 2019 **Computational Challenge Chair** – Oz Single Cell Conference, July 21-23, University of Melbourne, Melbourne, Australia
- 2018 **Reviewer** – International Society for Stem Cell Research (ISSCR), Jun 20-23, Melbourne, Australia
- 2017 **Session Chair (Computational Biology)** – Australasian Genomic Technologies Association (AGTA), Oct 29 - Nov 1, Hobart, Australia
- 2014 **Program Committee Member** – Pattern Recognition in Bioinformatics (PRIB), Aug 21-23, Stockholm, Sweden
- 2010 **Reviewer** – Pacific Rim International Conferences on Artificial Intelligence (PRICAI), Aug 30-Sep 2, Daegu, Korea
- 2009 **Reviewer** – IEEE Congress on Evolutionary Computation (CEC), May 18-21, Trondheim, Norway
- 2008 **Reviewer** – Pattern Recognition in Bioinformatics (PRIB), Oct 15-17, Melbourne, Australia
- 2008 **Reviewer** – Pacific Rim International Conferences on Artificial Intelligence (PRICAI), Dec 15-19, Hanoi, Vietnam

SOFTWARE

- An **R** package for single-cell RNA-seq data integration
<https://bioconductor.org/packages/release/bioc/html/scMerge.html>
- An **R** package for single-cell RNA-seq data clustering
<https://sydneybio.github.io/scdney/>
- An **R** implementation of AdaSampling for positive-unlabeled and class label noise learning
<https://github.com/PengyiYang/AdaSampling/>
- A web application in **JavaScript** for integrative transcription factor analysis
<http://github.com/PengyiYang/PAD>
- A **Shiny** application for kinase perturbation analysis

<http://shiny.maths.usyd.edu.au/KinasePA/>

- A GUI application in **Java** for kinase substrate prediction from phosphoproteomics data
<https://github.com/PengyiYang/KSP-PUEL>
- An **R** package for knowledge-based clustering analysis
<https://github.com/PengyiYang/ClueR>
- An **R** package for multi-dimensional pathway analysis
<http://cran.r-project.org/web/packages/directPA/index.html>
- Sample subset optimization package in **Java** for building imbalanced ensemble
<https://code.google.com/p/sample-subset-optimization/> (download count: >40)
- A machine learning algorithm in **Java** for protein inference
<http://code.google.com/p/re-fraction/> (download count: >35)
- A boosted learning algorithm in **Java** for peptide filtering
<http://code.google.com/p/self-boosted-percolator/> (download count: >110)
- A parallel genetic algorithm in **Java** for SNP interaction detection
<http://code.google.com/p/genetic-ensemble-snp/> (download count: >180)
- An ensemble algorithm in **Perl** for SNP interaction filtering
<http://code.google.com/p/ensemble-of-filters/> (download count: >70)
- An open source mass spectrometry analysis pipeline in **R**
<http://code.google.com/p/ocap/> (download count: >195)
- A dynamic wavelet package in **C/C++** for mass spectrum modelling
<http://code.google.com/p/dywave/DyWave> (download count: >100)
- A particle swarm optimisation algorithm in **Java** for imbalanced data sampling
<http://code.google.com/p/imbalanced-data-sampling/> (download count: >220)