DAVID ZHANG

I am a bioinformatician who focusses on data analysis, visualisation and method development. The goal of my work is to improve the molecular diagnosis rate of patients with rare disorders. For this purpose, I specialise in developing algorithms that integrate large-scale genomic and transcriptomic datasets in order to detect aberrant, disease-causing events.

View this CV online with links at *dzhang32.github.io/cv/*

EDUCATION

Present | 2017

Research assistant, part-time PhD, Bioinformatics

University College London — London, UK

- **Thesis**: Using transcriptomics to improve the diagnosis rate of rare disease patients.
- The goal of my PhD is to develop and apply software that improve the genetic diagnosis rates using RNA-sequencing.

2016 | 2015

MSc, Neuroscience

University College London — London, UK

- **Thesis**: The role of mitochondrial dysfunction in Xerodoma pigmentosum
- **Grade**: Merit (68%)
- Awarded post-graduate support scheme bursary (£10,000)

2015 | 2012

BSc, Biomedical science

University College London — London, UK

- **Thesis**: Investigating the function of CYFIP1 in the development of rat hippocampal neurons.
- **Grade**: 2:1 (69%)

2012 | 2007

H.S.

Queen Elizabeth's School — Barnet, UK

- **Grade**: Maths (A*), Biology (A*), Chemistry (A*), Sociology (A).

RESEARCH EXPERIENCE

2017 | 2016

Research Technician

University College London — London, UK

- Used R and bash to investigate the effect of genetic variation on the age of onset of dementia and cognition within Down syndrome patients.

CONTACT

- dyzhang32@gmail.com
- dyzhang32
- dzhang32
- david-zhang32

LANGUAGE SKILLS

- R
- Bash
- Git/GitHub
- Python

*Made with the R packages **datadrivencv** and **pagedown**.*

The source code is available on github.com/dzhang32/cv.

Last updated on 2021-03-11.

SOFTWARE & PROGRAMMING

Present | 2020

Bioconductor packages

- dasper[1]: detection of aberrant splicing events in RNA-sequencing. Author and maintainer. 363 downloads.
- megadepth[2]: BigWig and BAM related utilities. An R wrapper for the megadepth software developed by Chris Wilks. Co-author and maintainer. 304 downloads.

2021

Data science blog

- Published chess-related blog post on Medium[3]. Post was curated by Towards Data Science and selected for their hands-on-tutorials column, which displays pieces that highlight best practices of data science.
- Applied popular data science packages in python to analyse[4] chess.com data.

2020

Kaggle-town

- Organised Kaggle-town club to study python and machine learning through tackling kaggle problems[5].

2018 | 2016

Data wrangling

Neuroimmunology & CSF Laboratory, NHS — London, UK

- Developer and maintainer of data wrangling pipelines that improved the efficiency and standardisation of monthly financial reports.

TEACHING EXPERIENCE

2020

Developing Bioconductor Packages

University College London — Virtual Event

- Hosted workshop[6] on best practices for developing Bionconductor using biocthis[7]

2020 | 2018

R fundamentals

Clinician Coders — London, UK

- Developed materials[8] and lead workshops that aimed to teach R fundamentals to clinicians.

2020 | 2017

RNA-sequencing for diagnostics

Kings College London — London, UK

- Lectured graduate level students about how transcriptomics can be applied in the diagnostic pipeline.

SELECTED PUBLICATIONS

2021

Megadepth: efficient coverage quantification for BigWigs and BAMs

Bioinformatics

- Wilks C, Ahmed O, Baker DN, **Zhang D**, Collado-Torres L, Langmead B. **2021**. *Megadepth: efficient coverage quantification for BigWigs and BAMs*. Bioinformatics.
- **Role**: R package developer.
- DOI: https://doi.org/10.1101/2020.12.17.423317

2021

Integration of eQTL and Parkinson's disease GWAS data implicates 11 disease genes

Jama Neurology

- Kia DA, **Zhang D**, Guelfi S, Manzoni C, Hubbard L, United Kingdom Brain Expression Consortium (UKBEC), International Parkinson's Disease Genomics Consortium (IPDGC), Reynolds RH, Botía JA, Ryten M, Ferrari R, Lewis PA, Williams N, Trabzuni D, Hardy J, Wood NW. **2021.** *Integration of eQTL and Parkinson's disease GWAS data implicates 11 disease genes.* Jama Neurology.
- **Role**: Co-first author.
- DOI: https://doi.org/10.1001/jamaneurol.2020.5257

2020

Incomplete annotation of disease-associated genes is limiting our understanding of Mendelian and complex neurogenetic disorders.

Science advances

- **Zhang D**, Guelfi S, Ruiz SG, Costa B, Reynolds RH, D'Sa K, Liu W, Courtin T, Peterson A, Jaffe AE, Hardy J, Botia JA, Collado-Torres L and Ryten M. **2020.** *Incomplete annotation of disease-associated genes is limiting our understanding of Mendelian and complex neurogenetic disorders.* Science Advances.
- **Role**: First Author.
- DOI: https://doi.org/10.1126/sciadv.aay8299

2020

Regulatory sites for known and novel splicing in human basal ganglia are enriched for disease-relevant information.

Nature Communications

- Guelfi S, D'Sa K, Botía J, Vandrovcova J, Reynolds RH, **Zhang D**, Trabzuni D, Collado-Torres L, Thomason A, Leyton PQ, Gagliano SA, Nalls MA, UK Brain Expression Consortium, Small KS, Smith C, Ramasamy A, Hardy J, Weale ME & Ryten M. **2020.** *Regulatory sites for known and novel splicing in human basal ganglia are enriched for disease-relevant information.* Nature Communications.
- **Role**: Analyst.
- DOI: https://doi.org/10.1038/s41467-020-14483-x

2019

Genetic variability in response to Aβ deposition influences Alzheimer's risk.

Brain Communications

- Salih DA, Bayram S, Guelfi S, Reynolds RH, Shoai M, Ryten M, Brenton JW, **Zhang D**, Matarin M, Botia JA, Shah R, Brookes KJ, Guetta-Baranes T, Morgan K, Bellou E, Cummings DM, Escott-Price V, Hardy J. **2019**. *Genetic variability in response to Aβ deposition influences Alzheimer's risk.* Brain Communications.
- **Role**: Analyst.
- DOI: https://doi.org/10.1093/braincomms/fcz022

2019

Duplication of 10q24 locus: broadening the clinical and radiological spectrum.

Eur J Hum Genet

- Holder-Espinasse M, Jamsheer A, Escande F, Andrieux J, Petit F, Sowinska-Seidler A, Socha M, Jakubiuk-Tomaszuk A, Gerard M, Mathieu-Dramard M, Cormier-Daire V, Verloes A, Toutain A, Plessis G, Jonveaux P, Baumann C, David A, Farra C, Colin E, Jacquemont S, Rossi A, Mansour S, Ghali N, Moncla A, Lahiri N, Hurst J, Pollina E, Patch C, Ahn JW, Valat AS, Mezel A, Bourgeot P, **Zhang D**, Manouvrier-Hanu S. **2019.** *Duplication of 10q24 locus: broadening the clinical and radiological spectrum.* Eur J Hum Genet.
- **Role**: Analyst.
- DOI: https://doi.org/10.1038/s41431-018-0326-9

2019

Genetic variation within genes associated with mitochondrial function is significantly associated with later age of onset of Parkinson disease and contributes to disease risk.

NPJ Parkinson's Disease

- Billingsley KJ, Barbosa IA, Bandrés-Ciga S, Quinn JP, Bubb VJ, Deshpande C, Botía JA, Reynolds RH, **Zhang D**, Simpson MA, Blauwendraat C, Nalls MA, Singleton A, International Parkinson's Disease Genomics Consortium (IPDGC), Ryten M, Koks S. **2019.** *Genetic variation within genes associated with mitochondrial function is significantly associated with later age of onset of Parkinson disease and contributes to disease risk.* NPJ Parkinson's Disease.
- **Role**: Data provider.
- DOI: https://doi.org/10.1038/s41531-019-0080-x

2018

Variation at the TRIM11 locus modifies Progressive Supranuclear Palsy phenotype.

Annals of Neurology

- Jabbari E, John W, Tan MMX, Maryam S, Pittman A, Ferrari R, Mok KY, **Zhang D**, Reynolds RH, de Silva R, Grimm MJ, Respondek G, Muller U, Al-Sarraj S, Gentleman SM, Lees AJ, Warner TT, Hardy J, Revesz T, Hoglinger GU, Holton JL, Ryten M and Morris HR. **2018.** *Variation at the TRIM11 locus modifies Progressive Supranuclear Palsy phenotype.* Annals of Neurology.
- **Role**: Analyst.
- DOI: https://doi.org/10.1002/ana.25308

CONFERENCES

2021 **The British Society for Genetic Medicine (BSGM)**
Virtual Event

- **Talk**: dasper: detecting aberrant splicing events in RNA-sequencing data

2020 **EuroBioc**
Virtual Event

- **Talk**: dasper: detecting aberrant splicing events in RNA-sequencing data

2019 **Genomics England Research Conference**
London, UK

- **Poster**: Predicting disease-causing genes using machine learning

2019 **Genomics of Rare Disease**
Cambridge, UK

- **Poster**: The use of transcriptomics to improve gene annotation
- **Poster**: Using machine learning to understand and predict genes causing rare neurological disorders
- Awarded prize for the best poster (£100)

2019 **International Parkinson's Disease Genomics Consortium (IPDGC)**
Lisbon, Portugal

- **Talk**: Incomplete annotation of disease-associated genes is limiting our understanding of Mendelian and complex neurogenetic disorders

2018 **European Society of human genetics (ESHG)**
Milan, italy

- **Poster**: Incomplete annotation of OMIM genes is likley to be limiting the diagnostics yield from genetic tests.

2018 **International Parkinson's Disease Genomics Consortium (IPDGC)**
Reykjavik, Iceland

- **Poster**: Incomplete annotation of OMIM genes is limiting the diagnostic yield from genetic tests.

2015 **World Science Conference Israel (WSCI)**
Jerusalem, Israel

- 1 of 11 UK participants chosen to attend.

LINKS

1: https://bioconductor.org/packages/release/bioc/html/dasper.html

2: https://bioconductor.org/packages/release/bioc/html/megadepth.html
3: https://medium.com/@dzhang32
4: https://github.com/dzhang32/chess
5: https://github.com/dzhang32/kaggling
6: https://dzhang32.github.io/biocthis_workshop/
7: https://bioconductor.org/packages/release/bioc/html/biocthis.html
8: https://github.com/ClinicianCoders/ClinicianCoders