DAVID ZHANG

Experienced bioinformatician who enjoys learning and applying the best practices for method development. I focus on developing robust, user-friendly tools that harness transcriptomic data to improve the rate of genetic diagnosis.

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

EDUCATION

Present | 2017

PhD, Bioinformatics

University College London — London, UK

- **Thesis**: Using transcriptomics to improve the genetic diagnosis rate of rare disease patients.

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

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

WORK EXPERIENCE

2021

Bioinformatician internship (2 months)

Verge Genomics — Remote, US

- **Goal**: Set up an aberrant splicing detection pipeline for drug target discovery in C9orf72 ALS patients.
- Used **docker** to setup a reproducible workflow[1] for running aberrant splicing analyses on an **AWS** instance.

CONTACT

- dyzhang32@gmail.com
- GitHub
- LinkedIn
- ResearchGate

LANGUAGES SKILLS

- R
- Python
- Git/GitHub
- Bash
- docker

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

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

Last updated on 2021-12-21.

2017 | 2016

Research Technician

University College London — London, UK

- **Goal**: Investigate the impact of genetic variation on the age of onset of dementia and cognition within Down syndrome patients.

SOFTWARE & PROGRAMMING

Present | 2020

Bioconductor packages

- dasper[2]: Detection of aberrant splicing events in RNA-sequencing. **Author** and **maintainer**.
- megadepth[3]: BigWig and BAM related utilities. An R wrapper for the megadepth software developed by Chris Wilks. **Co-author** and **maintainer**.
- ODER[4]: Optimising the definition of Expressed Regions. Submitted to Bioconductor. **Co-author** and **maintainer**.

Present | 2021

Python packages

- codino[5] converts a codon design to the expected amino acid frequencies, and vice versa. **Author** and **maintainer**.

2021

Web scraping

- Applied the python packages **Beautiful Soup** and **Selenium** to web scrape[6] information on all UK biotechnology companies.

2021

Data science blog

- Chess-related blog post[7] was selected for the hand-on-tutorials column in Towards Data Science, which displays pieces that highlight best practices of data science.
- Applied popular data science packages in **python** to analyse[8] chess.com data.

TEACHING EXPERIENCE

2020

Developing Bioconductor packages

University College London — Virtual Event

- Hosted workshop[9] on best practices for developing Bionconductor packages using biocthis[10].

2020

Unit testing using testthat edition 3

rstats club — Virtual Event

- Talk[11] regarding unit testing fundamentals, the importance of testing and new features released in the R package testthat edition 3.

2020 | 2018

R fundamentals

Clinician Coders — London, UK

- Developed materials[12] 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

recount3: summaries and queries for large-scale RNA-seq expression and splicing

Genome Biology

- Wilks C, Zheng SC, Chen FY, Charles R, Solomon B, Ling JP, Imada EL, **Zhang D**, Joseph L, Leek JT, Jaffe AE, Nellore A, Collado-Torres L, Hansen KD, Langmead B
- **Role**: Adviser.
- **DOI**: https://doi.org/10.1186/s13059-021-02533-6

2021

Developmental Consequences of Defective ATG7-Mediated Autophagy in Humans

The New England Journal of Medicine

- Collier J, Guissart C, Oláhová M, Sasorith S, Piron-Prunier F, Suom Fi, **Zhang D**, Martinez-Lopez N, Leboucq N, Bahr A, Azzarello-Burri S, Reich S, Schöls L, Polvikoski TM, Meyer P, Larrieu L, Schaefer AM, Alsaif HS, Alyamani S, Zuchner S, Barbosa IA, Deshpande C, Pyle A, Rauch A, Synofzik M, Alkuraya FS, Rivier F, Ryten M, McFarland R, Delahodde A, McWilliams TG, Koenig M, and Taylor RW.
- **Role**: Analyst
- **DOI**: https://doi.org/10.1056/NEJMoa1915722

2021

Megadepth: efficient coverage quantification for BigWigs and BAMs

Bioinformatics

- Wilks C, Ahmed O, Baker DN, **Zhang D**, Collado-Torres L, Langmead B.
- **Role**: R package developer.
- **DOI**: https://doi.org/10.1093/bioinformatics/btab152

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.
- **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.
- **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.
- **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.
- **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.
- **Role**: Analyst.
- **DOI:** https://doi.org/10.1038/s41431-018-0326-9

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.
- **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://github.com/dzhang32/auto_splice
2. https://bioconductor.org/packages/release/bioc/html/dasper.html
3. https://bioconductor.org/packages/release/bioc/html/megadepth.html
4. https://github.com/eolagbaju/ODER
5. https://github.com/dzhang32/codino
6. https://github.com/dzhang32/biotech_web_scrape
7. https://towardsdatascience.com/how-has-the-queens-gambit-impacted-the-popularity-of-online-chess-43594efe5a98
8. https://github.com/dzhang32/chess
9. https://dzhang32.github.io/biocthis_workshop/
10. https://bioconductor.org/packages/release/bioc/html/biocthis.html
11. https://youtu.be/ClAin7vTwq0
12. https://github.com/ClinicianCoders/ClinicianCoders