

DAVID ZHANG

TBA



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.
- Ultimately, the goal of my PhD is to develop and apply statistical methods and software that improve the genetic diagnosis rate using RNA-sequencing. This involves detection of aberrant RNA-level events when complemented with, DNA sequencing help to resolve variants of unknown significance.

2016
|
2015

MSc, Neuroscience

University College London

📍 London, UK

- **Thesis:** The role of mitochondrial dysfunction in Xeroderma 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

2020

Honorary Researcher (2 months)

Johns Hopkins Bloomberg School of Public Health

📍 Remote

- In collaboration with Leonardo Collado-Torres¹, we used the recount3² dataset and LIBD samples to study the effect of complex splicing in individuals with neurological disease.

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

CONTACT

✉ dyzhang32@gmail.com

🐦 [dyzhang32](https://twitter.com/dyzhang32)

🐙 [dzhang32](https://github.com/dzhang32)

in [david-zhang32](https://www.linkedin.com/in/david-zhang32)

LANGUAGE SKILLS

R

Bash

Git/GitHub

Python

Made with the R package [pagedown](https://github.com/jgm/pagedown).

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

Last updated on 2021-03-09.

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.



INDUSTRY EXPERIENCE

2020



Bioinformatician internship (3 months)

Verge Genomics

📍 Remote

- Detection of aberrant splicing events in complex disease patients.
- Using AWS infrastructure



SOFTWARE & PROGRAMMING

Present
|
2020



Bioconductor packages

- `dasper`³: detection of aberrant splicing events in RNA-sequencing. Author and maintainer. XXX downloads.
- `megadePTH`⁴: BigWig and BAM related utilities. An R wrapper for the `megadePTH`⁵ software developed. Co-author and maintainer. XXXX downloads.

2021



Data science blog posts

- Published chess-related blogposts on Medium⁶. Posts were curated by Towards Data Science and selected to for the hands-on-tutorials column, which displays the best pieces for highlight data science principles.
- Applying python through the analysis⁷ of chess.com data.

2021
|
2020



Advanced R

- Notes and answers to the advanced R⁸ book in the form of a bookdown⁹.

2020



Kaggle town

- Organised club to study python and machine learning through kaggle¹⁰ problems.

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 workshop⁷⁷ on best practices for developing Bioconductor using `biothis`⁷²
- 2020

● **Unit testing using testthat edition 3**
rstats club 📍 Virtual Event

 - Talk⁷³ 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 materials⁷⁴ and lead facilitator for teaching R to clinicians.
- 2020
|
2017

● **RNA-sequencing for diagnostics**
Kings College London 📍 London, UK

 - Invited lecturer to graduate level students on 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>

- 2021 ● **Human-lineage-specific genomic elements: relevance to neurodegenerative disease and APOE transcript usage.**
Nature Communications
- Chen Z, **Zhang D**, Reynolds RH, Gustavsson EK, Garcia-Ruiz S, D'Sa K, Fairbrother-Brown A, Vandrovcova J, International Parkinson's Disease Genomics Consortium (IPDGC), Hardy J, Houlden H, Gagliano SA, Botiá J, Ryten M. Human-lineage-specific genomic elements: relevance to neurodegenerative disease and APOE transcript usage. Nature Communications.
 - Role: Analyst.
 - DOI: TBA
- 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, Botiá 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 A β 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 A β 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

- 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 likely 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 (WSCl)** 📍 Jerusalem, Israel
- 1 of 11 UK participants chosen to attend.



REFERENCES AVAILABLE UPON REQUEST



LINKS

- 1: <http://lcolladotor.github.io/cv/>
- 2: <http://bioconductor.org/packages/release/bioc/html/recount3.html>
- 3: <https://bioconductor.org/packages/release/bioc/html/dasper.html>
- 4: <https://bioconductor.org/packages/release/bioc/html/megadepth.html>
- 5: <https://github.com/ChristopherWilks/megadepth>
- 6: <https://medium.com/@dzhang32>
- 7: <https://github.com/dzhang32/chess>
- 8: <https://adv-r.hadley.nz>
- 9: https://dzhang32.github.io/advanced_R/
- 10: <https://github.com/dzhang32/kagglings>
- 11: https://dzhang32.github.io/biocthis_workshop/
- 12: <https://bioconductor.org/packages/release/bioc/html/biocthis.html>
- 13: <https://youtu.be/CIain7vTwq0>
- 14: <https://github.com/ClinicianCoders/ClinicianCoders>