DAVID ZHANG

TBA

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.
- Ultimately, the goal of my PhD is to develop and apply statisical 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 Xerodoma pigmentosum
- **Grade**: Merit (68%)
- Awarded post-graduate support scheme bursary (£10,000)

2015 | 2012

BSc, Biomedical science

University College London — London, UK

- **Thesis**:
- **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-Torres[1], we used the recount3[2] dataset and LIBD samples to study the effect of complex splicing in individuals with neurological disease.

CONTACT

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

LANGUAGE SKILLS

- R
- Bash
- Git/GitHub
- Python

*Made with the R package **pagedown**.*

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

Last updated on 2021-02-26.

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

Chess analysis

- Applying python and data science principles through the analysis[6] of chess data.

2021 | 2020

Advanced R

- Notes and answers to the advanced R[7] book in the form of a bookdown[8].

2020

Kaggle town

- Organised club to study python and machine learning through kaggle[9] 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 workshop[10] on best practices for developing Bionconductor using biocthis[11]

2020

Unit testing using testthat edition 3

rstats club — Virtual Event

- Talk[12] 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[13] 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

2020

Megadepth: efficient coverage quantification for BigWigs and BAMs

Bioinformatics

- Authored with Richard Single, Vanja Paunic, Mark Albrecht, and Martin Maiers.

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.

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://github.com/dzhang32/chess
7. https://adv-r.hadley.nz
8. https://dzhang32.github.io/advanced_R/
9. https://github.com/dzhang32/kaggling
10. https://dzhang32.github.io/biocthis_workshop/
11. https://bioconductor.org/packages/release/bioc/html/biocthis.html
12. https://youtu.be/ClAin7vTwq0
13. https://github.com/ClinicianCoders/ClinicianCoders