




DAVID ZHANG




Experienced bioinformatician who has experience developing robust, user-friendly software in python or R to facilitate diagnostics and drug discovery.

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




WORK EXPERIENCE

- present
|
2022
- **Machine learning engineer**
[Ladder therapeutics](#)  London, UK (remote)
 - **Goal:** Implementing and engineering production-ready software leveraging RNA biology and chemistry to accelerate drug discovery.
- 2021
- **Bioinformatician internship (2 months)**
[Verge Genomics](#)  London, UK (remote)
 - **Goal:** Set up an aberrant splicing detection pipeline for drug target discovery in C9orf72 ALS patients.
 - Used **docker** to setup a reproducible workflow¹ for running aberrant splicing analyses on an **AWS** instance.
- 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.

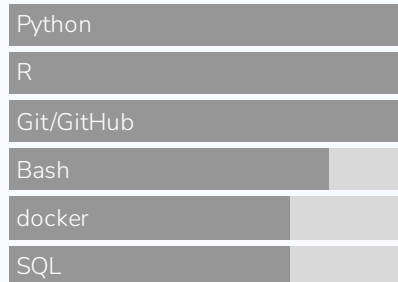
EDUCATION

- 2022
|
2017
- **PhD, Bioinformatics**
University College London  London, UK
 - **Thesis:** Using transcriptomics to improve the genetic diagnosis rate of rare disease patients.
 - Developed and released software that facilitate transcriptomic analyses with a focus on diagnostics.
- 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%)

CONTACT

✉ dyzhang32@gmail.com
 [GitHub](#)
 [LinkedIn](#)
 [Google Scholar](#)

LANGUAGE SKILLS



Made with the R packages [datadrivencv](#) and [pagedown](#).

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

Last updated on 2022-06-03.

2012
|
2007



H.S.

Queen Elizabeth's School

📍 Barnet, UK

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



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
|
2022



R packages

- **ggtranscript**⁵: Visualising transcript structure and annotation using ggplot2. **Author** and **maintainer**.
- **autorecipes**⁶: Automate your recipe planning. **Author** and **maintainer**.

Present
|
2021



Python packages

- **autogroceries**⁷: Automate your grocery shop. **Author** and **maintainer**.
- **codino**⁸ 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 scrape⁹ information on all UK biotechnology companies.

2021



Data science blog

- Chess-related blog post¹⁰ 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 analyse¹¹ chess.com data.



TEACHING EXPERIENCE

2020



Developing Bioconductor packages

University College London

📍 Virtual Event

- Hosted workshop¹² on best practices for developing Bioconductor packages using **biothis**¹³.

- 2020 ● **R package development**
Rstats club 📍 Virtual Event
- Presentation¹⁴ about unit testing fundamentals, the importance of testing and new features released in the R package testthat edition 3.
 - Presentation¹⁵ about pre-commit hooks in R.
 - Presentation¹⁵ about the best practices of developing R packages.
- 2020 |
2018 ● **R fundamentals**
Clinician Coders 📍 London, UK
- Developed materials¹⁷ 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

- 2022 ● **ggtranscript: an R package for the visualization and interpretation of transcript isoforms using ggplot2**
bioRxiv
- Gustavsson EK, Zhang D, Reynolds RH, Garcia-Ruiz S, Ryten M
 - **Role:** Co-first author.
 - DOI: <https://doi.org/10.1101/2022.03.28.486050>
- 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>

A complete list of publications is available via Google Scholar¹⁸

- 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, Botia 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>

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/ggtranscript>
- 6● <https://github.com/dzhang32/autorecipes>
- 7● <https://github.com/dzhang32/autogroceries>
- 8● <https://github.com/dzhang32/codino>
- 9● https://github.com/dzhang32/biotech_web_scrape
- 10● <https://towardsdatascience.com/how-has-the-queens-gambit-impacted-the-popularity-of-online-chess-43594efe5a98>
- 11● <https://github.com/dzhang32/chess>
- 12● https://dzhang32.github.io/biocthis_workshop/
- 13● <https://bioconductor.org/packages/release/bioc/html/biocthis.html>
- 14● <https://youtu.be/CIaIn7vTwq0>
- 15● https://github.com/dzhang32/stats_pres
- 16● https://github.com/dzhang32/stats_pres
- 17● <https://github.com/ClinicianCoders/ClinicianCoders>
- 18● <https://scholar.google.com/citations?user=PAcm4wUAAAAJ&hl=en>