





DAVID ZHANG



Bioinformatics software engineer who has experience developing production-ready pipelines and 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
- **Senior bioinformatics software engineer**
[Congenica](#)  Hinxton, UK
 - **Goal:** Developing, maintaining and productionising bioinformatic pipelines for the precision oncology product.
- 2022
- **Machine learning engineer**
[Ladder therapeutics](#)  London, UK (remote)
 - **Goal:** Implemented python packages that leverage 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 workflow⁷ 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)

CONTACT

- ✉ dzhang32@gmail.com
- 🌐 [Website](#)
- 🐙 [GitHub](#)
- in [LinkedIn](#)
- 📄 [Google Scholar](#)

LANGUAGES

- 🐍 Python
- 📊 R
- 🔗 Bash

TECHNOLOGIES

- 🔗 Git/GitHub
- 🔧 snakemake
- 🐳 docker
- 🔗 kubernetes
- 🗄️ SQL

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

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

Last updated on 2023-02-08.

- 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).

SOFTWARE & PROGRAMMING

- 2022
- **Portfolio website**
 - My website² is built using Django/Bootstrap 5, deployed with Heroku and showcases the five projects I'm most fond of.
- 2022
|
2021
- **R packages**
 - ggtranscript³: Visualising transcript structure and annotation using ggplot2. **Author** and **maintainer**.
 - autorecipes⁴: Automate your recipe planning. **Author** and **maintainer**.
- 2022
|
2020
- **Bioconductor packages**
 - dasper⁵: Detection of aberrant splicing events in RNA-sequencing. **Author** and **maintainer**.
 - megadept⁶: BigWig and BAM related utilities. An R wrapper for the megadept software developed by Chris Wilks. **Co-author** and **maintainer**.
 - ODER⁷: Optimising the definition of Expressed Regions. Submitted to Bioconductor. **Co-author** and **maintainer**.
- 2022
|
2021
- **Python packages**
 - autogroceries⁸: Automate your grocery shop. **Author** and **maintainer**.
 - codino⁹ 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 scrape¹⁰ information on all UK biotechnology companies.

2021

● Data science blog

- Chess-related blog post⁷⁷ 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 analyse⁷² chess.com data.



TEACHING EXPERIENCE

2020

● Developing Bioconductor packages

University College London

📍 Virtual Event

- Hosted workshop⁷³ on best practices for developing Bioconductor packages using biocthis⁷⁴.

2020

● R package development

Rstats club

📍 Virtual Event

- Presentation⁷⁵ about unit testing fundamentals, the importance of testing and new features released in the R package testthat edition 3.
- Presentation⁷⁶ about pre-commit hooks in R.
- Presentation⁷⁶ about the best practices of developing R packages.

2020

● R fundamentals

Clinician Coders

📍 London, UK

|

2018

- Developed materials⁷⁸ and lead workshops that aimed to teach R fundamentals to clinicians.

2020

● RNA-sequencing for diagnostics

Kings College London

📍 London, UK

|

2017

- 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

Bioinformatics

- Gustavsson EK, **Zhang D**, Reynolds RH, Garcia-Ruiz S, Ryten M
- **Role:** Co-first author.
- DOI: <https://doi.org/10.1093/bioinformatics/btac409>

A complete list of publications is available via Google Scholar⁷⁹

- 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, Botia JA, Ryten M, Ferrari R, Lewis PA, Williams N, Tratzuni 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://dzhang32.herokuapp.com>
- 3: <https://github.com/dzhang32/ggtranscript>
- 4: <https://github.com/dzhang32/autorecipes>
- 5: <https://bioconductor.org/packages/release/bioc/html/dasper.html>
- 6: <https://bioconductor.org/packages/release/bioc/html/megadepth.html>
- 7: <https://github.com/eolagbaju/ODER>
- 8: <https://github.com/dzhang32/autogroceries>

- 9: <https://github.com/dzhang32/codino>
- 10: https://github.com/dzhang32/biotech_web_scrape
- 11: <https://towardsdatascience.com/how-has-the-queens-gambit-impacted-the-popularity-of-online-chess-43594efe5a98>
- 12: <https://github.com/dzhang32/chess>
- 13: https://dzhang32.github.io/biocthis_workshop/
- 14: <https://bioconductor.org/packages/release/bioc/html/biocthis.html>
- 15: <https://youtu.be/CIAn7vTwq0>
- 16: https://github.com/dzhang32/rstats_pres
- 17: https://github.com/dzhang32/rstats_pres
- 18: <https://github.com/ClinicianCoders/ClinicianCoders>
- 19: <https://scholar.google.com/citations?user=PAcm4wUAAAAJ&hl=en>