


DAVID ZHANG

Bioinformatics software engineer with experience operating across the entire software development lifecycle. Skilled in prototyping and benchmarking innovative solutions, as well as implementing, testing, and integrating software into production-ready pipelines.


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

WORK EXPERIENCE


Present
|
2024

- **Senior bioinformatics engineer**
[CoSyne Therapeutics](#)  London, UK (hybrid)
 - Optimised and scaled machine learning tools to extract actionable insights from single-cell Perturb-seq datasets comprising millions of cells. Directed the project end-to-end, integrating findings to inform strategic decisions and guide company direction.
 - Designed and deployed a robust data pipeline that ingested, tidied and version-controlled data for the Neo4j knowledge graph. Automated the deployment of this graph via CI/CD using Terraform, enabling automated releases to AWS enhancing reproducibility and operational efficiency.

2024
|
2022


- **Senior bioinformatics software engineer**
[Congenica](#)  Hinxton, UK (hybrid)
 - Designed, benchmarked, and productionized scalable bioinformatics pipelines in Nextflow to process solid tumor sequencing data. Pipelines included alignment, variant calling, driver mutation annotation, and therapy matching, supporting clinical and translational applications.
 - Built a suite of Python and R packages to automate the clinical verification process, enabling earlier detection and resolution of issues. This automation reduced verification time by from 1 month per quarterly release, significantly accelerating the development cycle.

2021






- **Bioinformatician internship (2 months)**
[Verge Genomics](#)  London, UK (remote)
 - Created a reproducible aberrant splicing detection pipeline using docker for drug target discovery in C9orf72 ALS 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 ggtranscript¹, an open-source R package for visualizing transcript structures, which has recieved 150+ stars on GitHub and 250+ citations.






CONTACT

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

LANGUAGES

 Python
 Rust
 R
 Bash

TECHNOLOGIES

 Git/GitHub
 nextflow
 docker
 kubernetes
 SQL

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

Last updated on 2025-06-23.

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

SOFTWARE & PROGRAMMING

- Present
|
2022
 Portfolio website
 - My website² is built using Django/Bootstrap 5, deployed with Heroku and showcases the five projects I'm most fond of.
- 2024
 Rust packages
 - tuni³: Unify transcripts across different samples. Author.
- 2023
|
2021
 Python packages
 - codino⁴ converts a codon design to the expected amino acid frequencies, and vice versa. Author.
 - autogroceries⁵: Use Selenium to automate your grocery shop. Author.
 - stravaboard⁶: A dashboard for flexibly displaying and tracking Strava runs built using Streamlit. Author.
- 2022
|
2020
 R packages
 - ggtranscript⁷: Visualising transcript structure and annotation using ggplot2. Author.
 - megadepth⁸: BigWig and BAM related utilities. An R wrapper for the megadepth software developed by Chris Wilks. Co-author.
 - dasper⁹: Detection of aberrant splicing events in RNA-sequencing. Author,

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/10.1056/NEJMoa1915722>
- 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: Co-first author
 - DOI: <https://doi.org/10.1093/bioinformatics/btac409>
- 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>
- 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/ggtranscript>
- 2: <https://dzhang32-d2d1fbf5630e.herokuapp.com/>
- 3: <https://github.com/dzhang32/tuni>
- 4: <https://github.com/dzhang32/codino>
- 5: <https://github.com/dzhang32/autogroceries>
- 6: <https://github.com/dzhang32/stravaboard>
- 7: <https://github.com/dzhang32/ggtranscript>
- 8: <https://bioconductor.org/packages/release/bioc/html/megadepth.html>
- 9: <https://github.com/dzhang32/dasper>
- 10: <https://scholar.google.com/citations?user=PAcm4wUAAAAJ&hl=en>