


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.


[Download a PDF of this 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. Performed analysis and 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)
 - Developed 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 [ggtranscript](#), an open-source R package for visualizing transcript structures, which has recieved 150+ stars on GitHub and 250+ citations.





2016
|
2015

- **MSc, Neuroscience**
University College London  London, UK
 - Grade: Merit (68%)






CONTACT

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

LANGUAGES

 Python
 Rust
 R
 Bash

TECHNOLOGIES

 Git
 Nextflow
 Docker
 Kubernetes
 SQL

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

Last updated on 2025-07-06.

2015
|
2012



BSc, Biomedical science

University College London

📍 London, UK

- Grade: 2:1 (69%)



OPEN-SOURCE SOFTWARE

Present
|
2022



Portfolio website

- My [website](#) showcases some of my open source projects. The frontend is developed with HTML5, SCSS, and JavaScript, and the backend is powered by Django. It's deployed for free using PythonAnywhere.

2024



Rust packages

- [tuni](#): Unify transcripts across different samples.

2023
|
2021



Python packages

- [autogroceries](#): Use Selenium to automate your grocery shop.
- [stravaboard](#): A dashboard for flexibly displaying and tracking Strava runs built using Streamlit.

2022
|
2020



R packages

- [ggtranscript](#): Visualising transcript structure and annotation using ggplot2.
- [dasper](#): Detection of aberrant splicing events in RNA-sequencing.
- [dasper](#): Detection of aberrant splicing events in RNA-sequencing.



SELECTED PUBLICATIONS

2022



[ggtranscript: an R package for the visualization and interpretation of transcript isoforms using ggplot2](#)

Bioinformatics

- Role: Co-first author

2021



[Developmental Consequences of Defective ATG7-Mediated Autophagy in Humans](#)

The New England Journal of Medicine

- Role: Co-first author

2021



[Megadepth: efficient coverage quantification for BigWigs and BAMs](#)

Bioinformatics

- Role: R package developer.

A complete list of my publications is available via [Google Scholar](#)

2020

● **Incomplete annotation of disease-associated genes is limiting our understanding of Mendelian and complex neurogenetic disorders**

Science advances

• Role: First Author.