DAVID ZHANG

By bridging bioinformatics and engineering, I translate genetic and transcriptomic data into software that delivers real-world impact. With experience across the full software development lifecycle, I design, build, and deploy tools to solve bioinformatic problems — from prototyping innovative solutions to implementing and maintaining robust, production-ready pipelines.

💼 WORK EXPERIENCE

Present | 2024

Senior bioinformatics engineer
CoSyne Therapeutics
📍 London, UK (hybrid)

- Optimise and scale machine learning tools for single-cell perturb-seq data comprising millions of cells. Apply these tools to generate actionable insights and inform strategic decisions around company direction.
- Design and deploy a data pipeline to ingest, tidy and version-control data for the CoSyne knowledge graph. Automate the release of the graph to AWS using terraform and CI/CD, improving the efficiency and traceability of data updates.
- Build and maintain infrastructure tooling including docker images, terraform modules, CI/CD workflows and cruft templates to streamline bioinformatics analyses.

2024 | 2022

Senior bioinformatics software engineer
Congenica
📍 Hinxton, UK (hybrid)

- Developed scalable nextflow pipelines to process solid tumor DNA-sequencing data covering alignment, variant calling, driver mutation annotation, and therapy matching.
- Built python and R packages to improve the efficiency of clinical verification, reducing time taken by 2 weeks per quarterly release.

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

- Analysed bulk RNA-sequencing data with the aim of improving the diagnosis rate of rare disease patients. Focussed on detection of abberant splicing events as a strategy to prioritise pathogenic variants.
- Released R/Bioconductor packages that enable bioinformatics analyses and interpretation. Championed best practices for software development through teaching workshops and courses.

2016 | 2015

MSc, Neuroscience
University College London
📍 London, UK

- Grade: Merit (68%)

CONTACT
- ✉ dyzhang32@gmail.com
- 👤 Website
- GitHub
- in LinkedIn
- G Google Scholar

LANGUAGES
- 🐍 Python
- Ⓡ R
- Rust
- </> Bash

TECHNOLOGIES
- Git
- Nextflow
- Docker
- aws AWS
- Kubernetes

2015
|
2012

BSc, Biomedical science
University College London

📍 London, UK

- Grade: 2:1 (69%)

📈 OPEN-SOURCE SOFTWARE

Present
|
2022

Web development

- Portfolio website: Showcases my favourite open-source contributions.

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.

📖 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: Analyst

2021

Megadepth: efficient coverage quantification for BigWigs and BAMs
Bioinformatics

- Role: R package developer

2020

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

- Role: First author

A complete list of my publications is available via Google Scholar