# 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.

## 💼 WORK EXPERIENCE

**Present | 2024**

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

- Scaled machine learning packages to derive insights from single-cell perturb-seq data containing million of cells. Directed the project and to pull togther insights to determine company direction.
- Created a data pipeline to injest, version-control and deploy a neo4j knowledge graph. Automated the deployment and release of the graph to AWS via the CI using terraform.

**2024 | 2022**

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

- Developing, benchmarking and productionising bioinformatic pipelines in nextflow for the precision oncology product.
- Engineering nextflow and snakemake pipelines that perform alignment, variant calling, driver annotation and therapy matching using solid tumour sequencing data.

**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 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 Xerodoma pigmentosum
- Grade: Merit (68%)
- Awarded post-graduate support scheme bursary (£10,000)

## CONTACT

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

## LANGUAGES

- 🐍 Python
- ® Rust
- ® R
- </> Bash

## TECHNOLOGIES

- ⭘ Git/GitHub
- ⌁ nextflow
- 🐳 docker
- ⬡ kubernetes
- 🛢 SQL

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 website[1] is built using Django/Bootstrap 5, deployed with Heroku and showcases the five projects I'm most fond of.

2024

### Rust packages

- tuni[2]: Unify transcripts across different samples. Author.

2023
|
2021

### Python packages

- codino[3] converts a codon design to the expected amino acid frequencies, and vice versa. Author.
- autogroceries[4]: Use Selenium to automate your grocery shop. Author.
- stravaboard[5]: A dashboard for flexibly displaying and tracking Strava runs built using Streamlit. Author.

2022
|
2020

### R packages

- ggtranscript[6]: Visualising transcript structure and annotation using ggplot2. Author.
- megadepth[7]: BigWig and BAM related utilities. An R wrapper for the megadepth software developed by Chris Wilks. Co-author.
- dasper[8]: 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

A complete list of publications is available via Google Scholar[9]

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://dzhang32-d2d1fbf5630e.herokuapp.com/
2: https://github.com/dzhang32/tuni
3: https://github.com/dzhang32/codino
4: https://github.com/dzhang32/autogroceries
5: https://github.com/dzhang32/stravaboard
6: https://github.com/dzhang32/ggtranscript
7: https://bioconductor.org/packages/release/bioc/html/megadepth.html
8: https://github.com/dzhang32/dasper
9: https://scholar.google.com/citations?user=PAcm4wUAAAAJ&hl=en