





# 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/](https://dzhang32.github.io/cv/)






## WORK EXPERIENCE

- Present  
|  
2024
- **Senior bioinformatics engineer**  
[CoSyne Therapeutics](#)  London, UK (hybrid)
    - Goal: Engineering
- 2024  
|  
2022
- **Senior bioinformatics software engineer**  
[Congenica](#)  Hinxton, UK (hybrid)
    - Goal: Developing, benchmarking and productionising bioinformatic pipelines 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)
    - Goal: Set up an aberrant splicing detection pipeline for drug target discovery in C9orf72 ALS patients.
    - Used docker to setup a reproducible workflow<sup>7</sup> 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.






## CONTACT

 [dzhang32@gmail.com](mailto:dzhang32@gmail.com)  
 [Website](#)  
 [GitHub](#)  
 [LinkedIn](#)  
 [Google Scholar](#)

## LANGUAGES

 Python  
 Rust  
 R  
 Bash

## TECHNOLOGIES

 Git/GitHub  
 nextflow  
 docker  
 kubernetes  
 SQL

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

The source code is available on  
[github.com/dzhang32/cv](https://github.com/dzhang32/cv).

Last updated on 2025-06-21.

- 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 website<sup>2</sup> is built using Django/Bootstrap 5, deployed with Heroku and showcases the five projects I'm most fond of.
- 2024
- **Rust packages**
    - tuni<sup>3</sup>: Unify transcripts across different samples. Author.
- 2023  
|  
2021
- **Python packages**
    - codino<sup>4</sup> converts a codon design to the expected amino acid frequencies, and vice versa. Author.
    - autogroceries<sup>5</sup>: Use Selenium to automate your grocery shop. Author.
    - stravaboard<sup>6</sup>: A dashboard for flexibly displaying and tracking Strava runs built using Streamlit. Author.
- 2022  
|  
2020
- **R packages**
    - ggtranscript<sup>7</sup>: Visualising transcript structure and annotation using ggplot2. Author.
    - megadePTH<sup>8</sup>: BigWig and BAM related utilities. An R wrapper for the megadePTH software developed by Chris Wilks. Co-author.
    - dasper<sup>9</sup>: Detection of aberrant splicing events in RNA-sequencing. Author,

2021

## ● Data science blog

- Chess-related blog post<sup>10</sup> 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 analyse<sup>11</sup> chess.com data.



## 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: Analyst
- 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>

A complete list of publications is available via Google Scholar<sup>12</sup>



## LINKS

1: [https://github.com/dzhang32/auto\\_splice](https://github.com/dzhang32/auto_splice)

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://towardsdatascience.com/how-has-the-queens-gambit-impacted-the-popularity-of-online-chess-43594efe5a98>
- 11: <https://github.com/dzhang32/chess>
- 12: <https://scholar.google.com/citations?user=PAcm4wUAAAAJ&hl=en>