Sunny Sangar

Bioinformatics | Computational Biology | Data Analysis

Melbourne, VIC, Australia · sunnysangar29@gmail.com · sunnysangar.com · github.com/SunnyS29 · LinkedIn

Bioinformatician and data analyst with an MSc in Bioinformatics and three years of applied analytics work. Experienced in taking raw sequencing, imaging, and time-series data through to reproducible pipelines and results that hold up. Particular interest in network inference, method benchmarking, and validating computational predictions against experimental evidence. Melbourne based and open to bioinformatics and computational biology roles.

TECHNICAL SKILLS

Languages: Python (pandas, NumPy, SciPy, scikit-learn), R, SQL, MATLAB, JavaScript

Bioinformatics: HT-SELEX sequence analysis, microbial community and amplicon data, co-occurrence network inference, calcium imaging ROI detection, genomic and proteomic database management

Analysis: Statistical modelling, network analysis, transfer entropy, machine learning, benchmarking and method validation, reproducible workflows

Data and tools: PostgreSQL, DuckDB, Metabase, Tableau, Git and GitHub, Excel, workflow automation, Claude API

SELECTED PROJECTS

Sourdough Microbial Networks · Python, microbial ecology 2026

- Benchmarked seven co-occurrence network inference methods on 16S sequencing data from 500 sourdough starters, testing predicted associations against 28 published pairwise competition experiments.
- Fully reproducible analysis addressing a known weakness in microbiome work, that inferred networks are rarely checked against experimental outcomes. github.com/SunnyS29/sourdough-network-benchmark

Aptamer Identification Pipeline · Python, molecular diagnostics 2025

- Pipeline for identifying candidate aptamers from raw HT-SELEX sequencing data, scoring sequences by enrichment trajectory across selection rounds rather than single-round abundance to avoid false positives that do not persist.
- Handles read processing, sequence clustering, and ranking of candidates for downstream validation. github.com/SunnyS29/aptamer_test_repo

CalcCatch · MATLAB, neuroscience imaging 2024

- Region-of-interest detection tool for calcium imaging datasets that automatically identifies active cell regions and outputs structured results, replacing manual annotation.
- Formed the supervised cell-detection stage of the MSc thesis pipeline. github.com/SunnyS29/CalcCatch

Melbourne Cyclist Crash Explorer and Safety Dashboard · DuckDB WASM, SQL, Metabase 2025

- Browser-only analytical tool running SQL against a static Parquet file, alongside a Metabase dashboard covering 13 years and 16,056 VicRoads crash records broken down by time, location, and local area.
- Built end to end from public data, including cleaning, modelling, and delivery. sunnys29.github.io/cyclist-crash-explorer

EXPERIENCE

Data Analyst · Desire International Ltd, Hong Kong (remote) Nov 2022 - Jul 2026

- Analyse sales, pricing, and product data in SQL, Python, and Excel, delivering weekly reporting to 5 to 6 stakeholders across Sales and Marketing to support commercial decisions.
- Built dashboards and automated reporting workflows that cut report turnaround from around 3 hours to under 2, a 40 percent reduction, and made recurring analysis more reliable.
- Maintain a structured product database of around 200 SKUs, improving reporting consistency and data quality.
- Designed a self-service supplier due diligence dashboard with a custom scoring engine, replacing ad hoc manual assessment.

Data Research Assistant · Ark Global, Hong Kong Jun 2021 - Sep 2021

- Analysed market data in Python and Excel to identify demand patterns and turn raw data into usable reporting for procurement and inventory decisions.
- Built automated pipelines consolidating multiple data sources, saving around 3 hours of manual preparation per reporting cycle.

EDUCATION

MSc Bioinformatics · University of Melbourne 2023 - 2025

- Thesis: reconstructed neuronal network connectivity from calcium imaging time-series data using graph theory, transfer entropy, and statistical modelling.
- Coursework: machine learning, database systems, advanced programming, statistical analysis.

BSc Bioinformatics · University of Dundee 2018 - 2022

- Biochemistry and statistics coursework focused on modelling complex biological data. Maintained SQL databases of genomic and proteomic datasets, including clock gene expression across zebrafish, rat, and human.

CERTIFICATION

ICH Good Clinical Practice E6(R3) · Global Health Training Centre, 2026. Meets TransCelerate BioPharma minimum criteria for ICH GCP Investigator Site Personnel Training.