MANKIRAT SINGH

Cork, Ireland | msmankiratsingh@gmail.com
linkedin.com/in/mankirat-singh-694440177 | github.com/bymankirat | bymankirat.github.io/myportfolio

PROFILE

Bioinformatician and data scientist (MSc Bioinformatics and Computational Biology, University College Cork; thesis defended October 2026, results November 2026, graduation March 2027). I work across the wet lab and the computer: NGS assay development, in vivo studies and histology at the bench, and reproducible Python and R pipelines for microbiome, metabolomics, transcriptomics and comparative genomics, with honest validation (held-out data, placebo tests, independent re-checks). Looking for PhD positions and research or data roles in bioinformatics, multi-omics, health data and applied machine learning.

EDUCATION

MSc, Bioinformatics and Computational Biology (NFQ Level 9) **Sep 2025 – Present**
University College Cork, Ireland
Status: all modules completed and thesis defended (October 2026); final results November 2026
Modules: Data Mining, Introduction to Probability and Statistics, Data Analysis I & II, Introduction to Relational Databases (SQL), Programming for Bioscientists I & II (Python), Open Source Infrastructure for Modelling with Data (computer science, information systems and mathematics content)
Thesis: multi-omics exploratory analysis of gut microbiota and metabolome changes after bariatric surgery, IBIMA Plataforma BIONAND, Málaga

PG Diploma, Bioinformatics **Jul 2024 – Aug 2025**
St. Xavier's College, Mumbai, India
Grade: CGPA 8.50 / 10
Final project: Glypican-3 as a liver-cancer biomarker (TCGA, GEPIA2, protein structure)

BSc, Life Sciences **Jul 2020 – Jul 2023**
University of Delhi, New Delhi, India
Grade: CGPA 6.97 / 10, First Division

RESEARCH EXPERIENCE

Master Thesis Researcher **May 2026 – Oct 2026**
IBIMA Plataforma BIONAND | Málaga, Spain
- Built a reproducible multi-omics pipeline for a longitudinal bariatric-surgery cohort (sleeve gastrectomy and gastric bypass patients before and after surgery, with healthy lean controls): raw Ion Torrent 16S reads processed in QIIME2 (DADA2, SILVA) and LC-MS/MS faecal metabolomics analysed in R (phyloseq, ANCOM-BC2, mixOmics sPLS and DIABLO)
- Applied paired, within-patient statistics and cross-validated multi-omics integration; the results are not yet published

Intern-Researcher, R&D (diagnostics product development) **Sep 2024 – Jan 2025**
Premas Life Sciences Pvt Ltd | New Delhi, India
- Developed NGS diagnostic kits for *Mycobacterium tuberculosis* and 3 other infectious diseases
- Improved the Illumina sequencing workflow, making it 20% more efficient, and built bioinformatics pipelines to validate each kit's sensitivity

Intern-Researcher **Jul 2024 – Nov 2024**
National Agri-Food Biotechnology Institute (NABI) | Mohali, India
- Analysed *Cyp19a1* expression across human and mouse models; ran in vivo studies on 40+ subjects and quantified behaviour from ANY-maze video-tracking data

Intern-Researcher **Jul 2023 – Nov 2023**
National Agri-Food Biotechnology Institute (NABI) | Mohali, India
- Studied the banana DNA-damage gene *Ma-SOG1* and designed expression vectors for *Agrobacterium*-mediated transformation; later followed the gene across 17 plant genomes in an automated comparative-genomics pipeline

RESEARCH & DATA PROJECTS

What Makes H7N9 Different? Flu Host-Response Transcriptomics (report · code) **R, limma, WGCNA**
- Analysed 78 public microarrays of human airway cells infected with H7N9, H3N2, H7N7 or H5N1 at 3–24 hours (limma with array weights, fgsea, WGCNA, UpSet)
- Separated a shared antiviral core of 2,803 genes from 1,279 genes specific to H7N9, confirmed against every other strain; the H7N9-down genes are enriched for fatty-acid metabolism and mTORC1 signalling

SOG1 Across 17 Plant Genomes (follow-up to my NABI work) (report · code) **Python, HMMER, IQ-TREE**
- Searched 17 genomes from green algae to banana with the Pfam NAC model, built a maximum-likelihood tree (MAFFT, trimAl, IQ-TREE, 1,000 bootstraps) and tested selection, phosphorylation sites and AlphaFold disorder
- Identified the gene I cloned as the one-to-one ortholog of Arabidopsis SOG1; the DNA-binding domain is twice as constrained as the tail (dN/dS 0.19 vs 0.40)

Glypican-3 Across 33 Cancers (report · code) **Python, TCGA, survival**
- Tested a liver-cancer marker in every TCGA cancer against TCGA and GTEx normal tissue, with false-discovery control, Cox models adjusted for age, sex and stage, and a permutation test against the "best cutpoint" trap

- In liver cancer GPC3 separates tumour from normal (AUC 0.94); adjusted for stage, higher GPC3 goes with shorter survival (HR 1.22 per SD, p = 0.04); it is switched on by loss of methylation

Clinical Trial Safety Analysis with Independent QC (report · code) **R, SQL (DuckDB)**

- Took a 254-patient medical-product trial from raw CDISC data to analysis datasets, safety tables and survival analysis, following a written specification for every derived variable
- Re-built the key results in a second, independent program: matched production on all 306 subjects and 354 adverse-event cells, with unit tests and structural checks

Clinical-Trial Risk Screening Tool (report · code) **Python, scikit-learn**

- Analysed 40,124 Phase 2/3 trials from the ClinicalTrials.gov API to show sponsors why 1 in 6 stop early: mostly low enrolment, not the science
- Built a design-stage risk score tested on later, unseen years (AUC 0.67 vs 0.57 for a rule of thumb; its riskiest 10% stopped twice as often) and an interactive calculator that explains each trial's score; removed data leakage that would have inflated results

All-Island Wind Turn-Down from Public EirGrid Data (report · code) **Python, statistics**

- Estimated how much wind the all-island grid turns down when demand is low, from four years of public 15-minute EirGrid data (2022–2025): among hours with the same forecast wind, the quietest quarter by demand produced up to 23% less wind; the estimate tracks EirGrid's published curtailment (r = 0.86, mean difference 0.9 points)
- Checked it with a placebo on low-wind hours (about zero) and 95% intervals from resampling days, and replaced my first method after it failed its own placebo test; hours at the SNSP limit rose from 1.8% to 4.4% and peak at midday, while 61% of the turned-down wind is lost at night

TECHNICAL SKILLS

- **Bioinformatics & omics:** QIIME2 (DADA2, SILVA), mixOmics (sPLS, DIABLO), MetaboAnalyst, limma, fgsea/GSEA, WGCNA, clusterProfiler, HMMER, MAFFT, IQ-TREE, AlphaFold, PyMOL, Chimera, TCGA, GTEx, GEO, Ensembl
- **Statistics & machine learning:** paired and mixed designs, survival analysis (Kaplan–Meier, Cox), multiple-testing control, bootstrap and permutation tests, regression, gradient boosting, random forests, SVM, PCA, held-out and later-year validation
- **Programming & reproducibility:** Python (pandas, NumPy, scikit-learn), R (tidyverse, admiral), SQL (DuckDB, SQLite), Git and GitHub, automated pipelines with quality checks, CDISC SDTM/ADaM, Power BI
- **Wet lab:** NGS library preparation and Illumina sequencing, PCR and qPCR, Western blotting, ovariectomy and animal handling, behavioural phenotyping (AnyMaze), vector design and Agrobacterium transformation, microtomy and H&E staining
- **Communication:** research poster “Microtomy and staining: a histological study of tissue” (DS Kothari Centre, Miranda House, 2023); public web reports explaining each project for non-specialists
- **Responsible research:** documented assumptions, leakage checks, independent re-checks of key results, clearly stated limitations

CERTIFICATIONS, COMMUNITY & INTERESTS

- **Certifications:** Kaggle – Intro to Machine Learning, Data Visualization, Python, Pandas, Intro to Programming; Decode Life – AI and Machine Learning for Drug Discovery
- **Community & sports:** charity volunteer with EHSAAS Foundation; hosted the “STROKES” art society; GRSS roller-skating champion; former NCC cadet; photography featured in Photo Vogue
- **Languages: English (fluent, medium of instruction throughout), Hindi and Punjabi (native)**

REFEREES

- **Names of referees will be supplied on request.**