

# Shon Kurian George

India | United Kingdom

[+91 9633293190](tel:+919633293190) | [LinkedIn](#) | [Portfolio](#) | [GitHub](#) | [shon.kurian.george@gmail.com](mailto:shon.kurian.george@gmail.com)

## SUMMARY

---

Bioinformatics MSc graduate with a strong foundation in biological concepts, gained through a BTech in Biotechnology. Proficient in R, Python, shell scripting (bash) and a wide array of bioinformatics tools, technologies and libraries. Demonstrated expertise in developing novel analysis pipelines, interactive applications/tools, spatial transcriptomics and genomic data analysis through hands-on projects and research. Complemented by experience in the pharmaceutical sector, vaccine services and software development.

## EDUCATION

---

**The University of Edinburgh**, United Kingdom  
*Master of Science in Bioinformatics*

Sep 2023 - Sep 2024  
Grade: Merit

Relevant Courses: Bioinformatics Programming & System Management, Statistics & Data Analysis, Information Processing in Biological Cells, Functional Genomic Technologies, Next-Generation Genomics, Machine Learning in Python, Intro to Website & Database Design for Drug Discovery.

**Dr. D. Y. Patil Biotechnology and Bioinformatics Institute**, Pune, India  
*Bachelor of Technology (Biotechnology)*

Jun 2017 - Jul 2021  
CGPA: 8.79/10

Relevant Courses: Molecular Biology, Genetics, Microbiology, Cell Biology, Immunology, Organic Chemistry, rDNA Technology, Metabolism, Genomics, Transcriptomics & Proteomics, Virology, Biostatistics, Biomembranes & Molecular Cell Signalling, Basic Pharmacology & Toxicology.

Talks and Presentations:

- Defended Bachelor of Technology (Biotechnology) thesis:  
“Industrial training report on pharmaceutical services provided in APSL, Hyderabad (Pharmacology department)”.
- Presented a critical review of the research paper by Marta N. Shahbazi et al.:  
“Developmental Potential of Aneuploid Human Embryos Cultured Beyond Implantation.” *Nature Communications*, 2020.

## BIOINFORMATICS RESEARCH EXPERIENCE

---

**Postgraduate Student Researcher** at **Abreu RNAlab**, University of Edinburgh, United Kingdom

Mar 2024 - Sep 2024

**“Using spatial transcriptomics to evaluate if microRNAs regulate expression profiles at cell-type boundaries”**

**Project Overview:**

- Developed a novel R analysis pipeline using spatial transcriptomics and statistical testing to detect and infer the effects of tissue-specific microRNAs (miRNAs) across diverse mouse tissue datasets.
- Utilised spatial transcriptomics libraries (Seurat and SPATA2) to identify neighbouring regions with differentially expressed miRNA targets, detecting signals in heart and brain datasets and analysing miRNA expression patterns in detail.
- Built an interactive R Shiny application to enhance accessibility; outlined project limitations and proposed future pipeline enhancements. Initially hosted on the university server; now deployed on AWS EC2. (**Launch App**)

**Research Skills Gained:**

- Collaborated closely with project supervisor Dr. Cei Abreu-Goodger through regular one-on-one meetings to discuss project progress and resolve research challenges.
- Developed specialised expertise in miRNAs and spatial transcriptomics, focusing on the 10x Visium Spatial Gene Expression technique and utilising the SPATA2 analysis library.
- Systematically documented research activities using Notion, strengthening essential research skills such as data management, coding practices, academic writing, and statistical analysis.

## WORK EXPERIENCE

---

**R Package Developer** working with **Jan Kückelhaus**, University Hospital Erlangen, Germany

Nov 2024 - Present

- Currently developing the *cypro* R package, aimed at streamlining image-based cell profiling workflows by integrating data from diverse platforms like CellProfiler and CellTracker, hosted on GitHub for version control.
- Enhancing the package's usability by implementing intuitive Shiny-based interfaces, allowing detailed specification down to individual well-level configurations, significantly simplifying user interaction.
- Refining the S4 class objects within cypro to ensure accurate data integration from imaging platforms, enabling comprehensive cell movement analyses. Incorporated proactive user warnings and notifications within the Shiny application, minimising user errors, safeguarding data integrity, and enhancing the user experience.
- Integrating knowledge of AWS cloud computing to build more engaging and interactive tutorials, effectively shortening the learning curve and promoting usability to a wider audience.

## SELECTED PROJECTS ([Complete List](#))

---

### [Small Molecule Database and Visualisation Application](#) ([Launch App](#))

Full-stack small-molecule compound search app built with PHP and MySQL, integrating the [NIH CACTUS](#) service for on-the-fly SMILES visualisation. The production database contains 6,600+ compounds with associated SMILES from five manufacturers, supporting fast filtering by molecular properties and supplier, selection-level statistics, and session-persistent selections. Deployed on AWS EC2 with robust security measures, reliable error handling, and automated temp-file cleanup for consistent rendering.

### [Protein Conservation and Motif Analysis Suite](#) ([Launch App](#)) ([View User Manual](#))

Terminal-based Python workflow that queries NCBI databases via EDirect for a user-specified taxonomic group and protein, assembles datasets of up to 1,000 sequences, and automates multiple sequence alignment and conservation plotting. Integrates PROSITE and EMBOSS to perform motif discovery and back-translation, outputting species counts, alignments, customisable conservation plots, and motif-frequency tables. User-centric features include input validation and logging of intermediate outputs for quality control and reproducibility.

### [Detailed Collaborative Machine Learning Project on Modifiable Risk Factors Linked to Dementia](#)

Leveraged the SHARE dataset to identify key modifiable dementia risk factors, using polynomial regression with GridSearchCV to determine a third-degree polynomial as optimal based on lowest MSE and highest R<sup>2</sup>; highlighted physical activity and social engagement as primary predictors. Contributed contextual framing and final interpretation, translating model outputs into actionable lifestyle and policy interventions as informed by the [2017](#) and [2020](#) Lancet Commissions (Livingston et al.), while critically addressing limitations.

### [Badminton Shuttle Detection Application](#) (RF-DETR & Flask)

Custom RF-DETR (Roboflow-Detection Transformer) object detection model trained for 200 epochs on 195 manually labelled images, achieving 85.7% precision, 75.0% recall, and an 80.0% F1-Score. A Flask web app presents 6 curated rally clips with a side-by-side toggle between raw footage and annotated detections, accompanied by per-rally filtering and detection statistics, deployed on an AWS EC2 instance.

### [Critique of \*Automeris io\* moth de novo Genome Assembly and Annotation](#) (incl. assembly using wtdbg2)

Executed a critical evaluation of the genome assembly workflow performed by [Skojec et al.](#), analysing 1.9 million HiFi reads and comparing assembly quality between the Hifiasm (N50: 15.78 Mb) and wtdbg2 assemblers (N50: 1.1 Mb). Demonstrated the superior performance of Hifiasm in generating a high-quality assembly with 98.4% completeness and achieving a total assembly size of 490 Mb with only 600 contigs in the first draft compared to wtdbg2, which had 3,362 contigs.

## BIOTECHNOLOGY INTERNSHIPS

---

**Industrial Training Trainee at Aurigene Pharmaceutical Services Limited (APSL), India.** Feb 2021 - Jun 2021

- Contributed to the In-Vivo Pharmacology Department, which managed 600+ rodents across 1,000+ efficacy studies annually, supporting drug discovery aligned with US FDA and DCGI standards through efficient time and project management.
- Gained exposure to multiple workflows including RNA isolation for 100+ tissue samples, ELISA with <5% CV across 96-well plates, and RT-qPCR used in regulatory submissions, demonstrating collaboration & innovative problem-solving.

**R&D Intern at Serum Institute of India Private Limited, India.** Jun 2019 - Jul 2019

- Gained exposure to state-of-the-art vaccine manufacturing processes in one of the world's largest vaccine producers, observing large-scale operations utilising fermenters and bioreactors with capacities up to 20,000 L.
- Learned about Good Manufacturing Practices (GMP) and Good Laboratory Practice (GLP), critical to Serum producing 1.5 billion vaccine doses annually in global distribution.

## CORE COMPETENCIES

---

**Programming Skills/Languages:** R, Python, Shell Scripting (bash), SQL (MySQL).

**Platforms/DevOps:** Linux/Unix, Docker, Git/GitHub, WSL, AWS EC2, Nextflow (learning).

### **Bioinformatics Frameworks and Libraries:**

**Technologies:** Spatial Transcriptomics, RNA Sequencing (RNA-Seq), Next-Generation Sequencing (NGS) Analysis.

**R/Bioconductor:** Tidyverse, ggplot2, Shiny, Seurat, DESeq2, SPATA2. **Python:** Pandas, NumPy, Matplotlib, Scikit-learn.

**Software:** RStudio, VS Code, Microsoft Office, Adobe Suite (Photoshop, Premiere Pro), Ollama, Open WebUI, Lucidchart.

**CLI Tools:** NCBI EDirect, IGV, SRA Toolkit, FastQC, MultiQC, EMBOSS Suite, HiCanu, wtdbg2, QUAST, BlobTools, etc.

**Statistical and ML:** Experimental Design, Hypothesis Testing, Multivariate Analysis, Model Training, Validation, Optimisation.

## REFERENCES

---

- Dr Alasdair Ivens** ([aivens2@ed.ac.uk](mailto:aivens2@ed.ac.uk)) University of Edinburgh, Programme Director & MSc Course Coordinator.
- Dr Ceil Abreu-Goodger** ([cei.abreu@ed.ac.uk](mailto:cei.abreu@ed.ac.uk)) University of Edinburgh, MSc Dissertation Supervisor & Lecturer.