

James O'Reilly

32D, Mildmay Grove South, London N1 4RL

☎ (+353) 861210196 | ✉ jamesdanielloreilly1@gmail.com | 🌐 [jamesdanielloreilly](#) | 📺 [jamesdanielloreilly](#)

Mathematics & computer science graduate completing a two-year bioinformatics MSc. Seeking machine learning and genomics research roles in industry. Significant experience in implementing pipelines for statistical analysis of NGS data, machine learning applications in genomics, and commercial software engineering. Available for employment September 2021. As an Irish citizen I have full employment and residency rights in the UK.

Education

KU Leuven

Leuven, Belgium

M.Sc. IN BIOINFORMATICS

Sept 2019 - Present

- Grade: H1.1 (First Class Honours, Cum Laude)
- Thesis: Disentangling deterministic and stochastic epigenetic sources of cellular heterogeneity in cancer. Supervisor: Prof. Bernard Thienpont at the VIB Lab for Functional Epigenetics, Leuven.
- Bioinformatics: Experience with regulatory and comparative genomics (RNA-seq, scRNA-seq, ATAC-seq, ChIP-seq), evolutionary and quantitative genetics, and structural bioinformatics. Specialised in statistical analysis of single-cell omics data and integration of multi-omic data.
- Mathematics: Deep learning, Bayesian analysis, applied multivariate statistical analysis, machine learning, dynamical systems, SVMs.

University of Bristol

Bristol, England

B.Sc. IN MATHEMATICS AND COMPUTER SCIENCE (JOINT HONOURS)

Sept 2016 - June 2019

- Final grade: H2.1 (67%)
- Maths: foundational courses in calculus, probability, analysis, and linear algebra. Specialised in discrete mathematics, graph theory and information theory; focusing on codes, communication, and cryptographic schemes.
- Computer Science: machine learning (77%) and computational neuroscience (84%). Extensive programming and development work in Python.

Work Experience

VIB Lab for Functional Epigenetics

Leuven, Belgium

BIOINFORMATICS RESEARCH INTERN

Sept 2020 - Aug 2021

- Completed a year-long project investigating sources of tumour heterogeneity in lung cancer under the supervision of Prof. Bernard Thienpont.
- Pre-processing, quality control and integration of 10X Genomics CROP-seq data. Evaluated heterogeneity metrics on CROP-seq data and identified epigenetic knockouts contributing to tumour heterogeneity. Used latent variable models (MOFA+) to disentangle sources of heterogeneity in knockout populations.

DataCamp

Leuven, Belgium

SOFTWARE ENGINEER INTERN

Jun 2020 - Aug 2020

- Contributed to a variety of projects as part of the engineering team. Implemented automated testing of code correctness in Python, built and optimised Docker images used across all of DataCamp's online courses, and maintained the python backend for automated code correction.
- Gained experience with DevOps infrastructure: shipping production-level code in an agile environment using Git, Jira, and CI.

University of Bristol - Dept. of Computer Science

Bristol

COMPUTATIONAL NEUROSCIENCE RESEARCH INTERN

Jun 2019 - Aug 2019

- NLP project investigating the application of deep neural networks in determining the statistical structure of language.

Significant Projects

- **Bristol Final Year Project:** Conceptualised, designed, and evaluated a virtual reality learning environment (VRLE) for calculus education. Demonstrated that student understanding of abstract mathematical concepts could be improved using VR technology, while facilitating both distance learning and learning for disabled students. [GitHub](#).
- **Prediction of heavy and light-chain antibody pairing** Project for the 2021 Copenhagen Bioinformatics Hackathon. Used random forests to predict heavy and light-chain antibody pairing from amino acid sequences. Identified critical regions for improved design of antibodies for targeted immunotherapy.
- **Integration of single-cell multi-omics data:** Used Multi-Omics Factor Analysis to integrate PBMC data from the novel 10X-Genomics single-cell multiome assay. Identified the combined contributions of the transcriptome and epigenome to cellular variation. Performed cell-type annotation using the Human Blood Atlas. Predicted single-cell gene expression from chromatin accessibility using ArchR. In collaboration with Prof. Stein Aerts at the VIB Laboratory for Computational Biology. [GitHub](#).
- **Chromatin accessibility analysis in NSCLC:** Implemented a full ATAC-seq pipeline for analysis of chromatin accessibility in non-small lung cancer cells. Investigated epigenetic state shifts between active and dormant states across treatments. Identified TF binding motifs and upregulated promoters of epithelial-mesenchymal transition.
- **Factorial single cell latent variable modelling of PBMC data:** Used the Slalom package to perform f-sLVM on PBMC data, using predefined genesets to identify specific biological factors contributing to transcriptional variability. Identified and regressed out technical and biological factors for downstream analyses. [GitHub](#).
- **Image classification with deep neural networks:** Implemented deep feature learning with a deep convolutional NN for image classification in Matlab. Worked to optimise network topology and hyperparameters. Critically evaluated different learning algorithms for NNs. Implemented different generative models, including deep Boltzmann machines and generative adversarial networks (GANs). Investigated convergence and stability of DCGANs, with a focus on Wasserstein GANs. [GitHub](#).

Skills

Programming Advanced: Python, R, Matlab, Bash, LaTeX. Experience with: Julia, C, C#, Java, JavaScript, haskell.

DevOps Git, Docker, CircleCI, Jira

Bioinformatics NGS tools (see below), BEAST, Pymol, GWAS, plink, model organism and disease database knowledge, and more.

NGS Essential NGS tools: Bowtie2, MACS2, GATK, CellRanger, Deeptools, etc

Statistical Analysis: Seurat, MOFA+, DeSeq, ArchR, ScanPy and more