

Adina Nadeem

Berlin, Germany ✉ adinanadeem@gmail.com 🌐 adinanadeem.com

Research Interests

Machine learning for biology: representation learning, protein language models, sequence-function relationships and generative modelling.

Education

M.Sc. Bioinformatics

Oct 2024 – Present

Freie Universität Berlin, Germany

Specialization in Data Science. Expected completion: 2027.

B.S. Bioinformatics

2019 – 2023

COMSATS University Islamabad, Pakistan

Experience

Bioinformatics Student Assistant

Sep 2025 – Present

Max Planck Institute for Infection Biology, Berlin | Portugal Group

- Support computational work for longitudinal malaria studies, working closely with PhD researcher Manuela Carrasquilla on *P. falciparum* persistence and within-host parasite dynamics in Mali.
- Contribute to data analysis, quality control and visualization for longitudinal malaria datasets.
- Develop and troubleshoot reproducible workflows for recurring analyses used by researchers in the lab.

Student Researcher – X-Student Research Group

Oct 2025 – Jan 2026

Charité – Universitätsmedizin Berlin

Dr. David Hillus group: Exploring Individual Immune Responses through Machine Learning

- Selected for an interdisciplinary research group using machine learning to study individual immune responses across immunological, clinical and genetic data.
- Led the computational analysis for HLA-specific prioritization of HPV16 T-cell epitopes using IEDB data, BLOSUM62 encodings, and physicochemical and immunological features.
- Trained and evaluated LightGBM and random forest models to predict immunogenicity and rank candidate epitopes; co-authored the resulting research poster.

Bioinformatics Junior Scientist

Oct – Dec 2023

Bio-code

- Developed analysis workflows using Python, Bash and R through projects in whole-genome sequencing, metagenomics, transcriptomics and gene expression.

Research Assistant

Aug 2022 – Feb 2023

Department of Biosciences, COMSATS University Islamabad

- Designed and implemented a Nextflow workflow using Python and Bash to identify polyadenylated reads in RNA-seq data, integrating adapter trimming, strand determination and alignment.
- Analysed sequencing data, produced quality-control visualisations and packaged the workflow in Docker for reproducible execution.

Selected Computational Projects

Exploring protein language model representations and sequence landscapes

Ongoing

ESM2, Python, evolutionary sequence search

Started as a deep learning module project on a topic proposed by Luca Ferretti and Vitaly Belik; now continuing as a research collaboration with a PhD student in Vitaly Belik's lab.

- Explore protein sequence landscapes and epistasis using pretrained ESM2 models, evolutionary search and ESMFold structure comparisons.
- Added adaptive mutation refinement to an existing search algorithm, adjusting mutation proposals according to progress towards a target protein.

Multi-omics machine learning for breast cancer subtypes

GitHub

RNA-seq, DNA methylation, machine learning, SHAP

- Analysed TCGA-BRCA RNA-seq expression, DNA methylation, clinical metadata and estimated immune-cell fractions to study Basal, Luminal A and Luminal B tumours.
- Compared supervised classifiers and feature sets for subtype prediction, including gene-expression, methylation, immune and age features, using holdout evaluation.
- Used SHAP and pathway enrichment to interpret predictive features; examined subtype differences in epigenetic age acceleration and immune-cell estimates.

malaria-amplicon-nf

Desktop | HPC

Nextflow, Slurm/HPC, Apptainer, DADA2, DINEMITES, Dcifer

- Built a desktop application for analysing data from the six-locus SIMPLseq amplicon panel, with FASTQ scanning, sample-sheet preparation, Nextflow execution and quality-control reports.
- Added Slurm/HPC execution with Apptainer packaging, resource configuration, resumable runs and automated collection of output tables and summaries.
- Integrated DINEMITES and Dcifer using CIGAR allele tables, participant metadata and replicate checks to analyse new versus persistent infections, infection complexity and parasite relatedness.

Single-cell var-gene expression analysis

Source figure

Python, R, single-cell RNA-seq, hierarchical clustering, data visualisation

- Built a reusable workflow to reproduce the analyses and visualisations behind Figure 6 of the Portugal group's published study of *var*-gene expression in asymptomatic and clinical malaria infections.
- Combined expression matrices and cell metadata to generate clustering, annotated heatmaps and domain-coverage plots, with configurable inputs and exported plot data for reuse.

Technical Skills

Programming: Python, R, Bash; working knowledge of SQL.

Machine learning & data analysis: supervised learning, model evaluation, feature interpretation, dimensionality reduction, clustering.

Bioinformatics & scientific computing: bulk and single-cell RNA-seq, amplicon sequencing, multi-omics, genome assembly and annotation; Nextflow, Snakemake, Docker, Linux/Slurm HPC.

Certification: AWS Certified Cloud Practitioner.