Accelerating Genomic Research with Artificial Intelligence: From Data to Discovery

*Co-led by NUS and A*STAR Research Entities (ARES)*

The volume of human genetic data being generated worldwide is growing faster than scientists can meaningfully analyse. Large-scale initiatives such as Singapore’s own SG100K programme, which is building a genomic database of 100,000 Singaporeans, are producing rich datasets, but existing AI tools tend to examine DNA, RNA and gene activity in isolation, missing the connections between them that are often most medically significant.

This project develops MultiOmicsFM, a unified AI foundation model (a large, general-purpose AI system trained on vast datasets and adaptable to many tasks) designed to read and interpret DNA sequences, RNA sequences and gene expression profiles together. Think of it as moving from separate diagnostic tests to an integrated health report that captures how genes are structured, regulated and expressed in concert.

The project builds out in four stages: (i) curating high-quality genomic datasets from Singapore and global sources; (ii) designing a scalable AI architecture optimised for biological data; (iii) developing AI agent tools that help biologists work alongside AI to accelerate discovery; and (iv) validating the model in real-world applications such as predicting disease risk, interpreting genetic variants associated with illness, and optimising mRNA-based therapies.

By harnessing Singapore’s uniquely multi-ethnic genomic datasets, the project aims to position Singapore as a global leader in AI-driven precision medicine, i.e. the practice of tailoring medical treatment to an individual’s genetic profile.