Developing Artificial Intelligence Approaches to Investigate Epigenomic Regulatory Logic in Healthy and Disease States

We are seeking a highly motivated PhD student to join The Computational Biology and Biomedical AI group led by Dr Shamith Samarajiwa (Associate Professor in Genomics), Imperial College London. The lab (https://www.samarajiwa-lab.org/) is based at the Imperial Centre for Experimental and Translational Medicine, Department of Metabolism, Digestion and Reproduction, at the Hammersmith Hospital Campus in London.

Gene expression regulation, the process by which genes are turned on or off is governed by an intricate regulatory logic involving both DNA and the epigenome, which is shaped by environmental cues and the surrounding tissue environment. Key players in this system include DNA regulatory elements: promoters that function as on-off switches, and enhancers that act as volume controls to dictate the timing, location, and magnitude of gene activation assisted by reversible epigenomic changes. Transcription Factors (TFs) are proteins that bind to these molecular switches to direct expression. Disruptions in these complex gene and epigenomic regulatory mechanisms frequently cause disease. However, uncovering the rules that dictate target-gene selection by enhancers remains an unsolved challenge in modern biomedicine. Uncovering this regulatory logic is crucial for clarifying disease mechanisms, enhancing computational predictions, and designing targeted therapeutics. This challenge has become increasingly tractable due to breakthroughs in AI, alongside the massive growth of omic technologies and data in repositories, biobanks, bioinformatic databases, and single-cell atlases. Consequently, computational methods are vital to isolate predictive regulatory features involved in this process and advanced multimodal sequence to function AI foundation models (such as Enformer, Basenji2, Borzoi, and AlphaGenome) are proving highly effective at predicting gene expression.

To address these questions, we are offering a fully funded PhD project focused on decoding genomic variation, TF-epigenomic dynamics and enhancer-promoter interactions that drive gene expression and phenotypes across diverse biomedical models, including ageing, cancer, metabolic disease, inflammation and infection. Leveraging the Samarajiwa Lab's extensive expertise in applying machine and deep learning (Newsham *et al*, *Biol. Meth. Prot.*, 2024; Rossi *et al*, *Sci Adv.* 2022; Martens *et al., bioRxiv* 2020) and in regulatory epigenomics (Patel *et al.*, *Nat.* 2022; Tomimatsu *et al., Nat. Aging* 2021; Rodriguez *et a.l, Cancer Disc.* 2018; Kirschner *et al., PLoS Genet.* 2015). The successful candidate will develop innovative and efficient predictive AI models, and Neuro-Symbolic AI approaches incorporating knowledge graphs, logic and graph neural networks that will assist in identifying TF utilization patterns, consequences of genetic variants, characterize enhancer classes, predict and map functional enhancer targets, and ultimately help illuminate the fundamental rules governing enhancer target selection in both healthy and disease states.

Applicants must hold (or be on track to obtain) a 2:1 (or equivalent) BSc in a relevant computational subject. A master's degree in (Computer Science, Physics, Engineering, Artificial Intelligence, Deep Learning, Mathematics, Computational Biology, Bioinformatics, or Systems Biology) is preferable but not essential. Applicants must also meet Imperial College's English language requirements: further details can be found at https://www.imperial.ac.uk/study/pg/apply/requirements/english/. While prior knowledge of

biology is not required, a strong interest in applying AI based approaches to biological/ biomedical science and a willingness to learn the relevant biology required for the project is expected. Prior experience in data science, Python programming and Machine/Deep Learning model development (PyTorch or Keras) is highly desirable. An interest in biomedical applications of foundation models, LLMs, DNA language models, Knowledge Graphs, Neuro Symbolic AI approaches and use of agentic AI is also desirable.

Funding: Includes course fees at the UK-Home rate (36 months) and a tax-free stipend (42 months) based on the annual UKRI rate of £23,805. Funding for overseas fees is not provided.

Applicants should submit their CV and a cover letter, including full contact details of two referees, to Dr Shamith Samarajiwa (s.samarajiwa@imperial.ac.uk). Informal enquiries are welcome, please email Dr Shamith Samarajiwa if you wish to discuss the project prior to formal application. Closing date is 30 November 2026 at 24:00 GMT. Any applications received after this date/time or without all requested documentation will not be considered. Shortlisted applicants will be interviewed in December.