Bioinformatics and machine learning/AI based approaches to understanding the mechanism by which long non-coding RNAs regulate the inflammatory response

Supervisory team:
Main supervisor: Prof Mark Andrew Lindsay (University of Bath)
Second supervisor: Dr Fabrizio Costa (University of Exeter)
Dr Chris Bailey (University of Bath)

Host institution: University of Bath

Project description:
High-throughput sequencing has indicated that between 10 - 67% of the human genome is transcribed into RNA. However, less than 2% of the DNA codes for protein coding genes, with the remainder classified as non-coding RNA (ncRNA) and commonly divided into short ncRNAs (less than 200 nucleotides) and long ncRNAs (greater than 200 nucleotides). The microRNA family of short ncRNAs are the best characterised and are known to regulate mRNA translation via the RNA interference pathway. By contrast, much less in known about long non-coding RNAs (lncRNAs) although there is accumulating evidence showing that these regulate of a host of physiological/pathological responses including immunity and inflammation. Our understanding of the mechanism of lncRNA action is limited although it has been speculated that this is mediated through 'domains' that interact with proteins and/or base pair with RNA/DNA. However, the identification of these 'domains' has been hindered by their poor evolutionary conservation.

Activation of innate immunity provides a crucial initial defence against infection by pathogens such as bacteria, fungi and viruses. This process involves the receptor-mediated recognition of invading pathogens, leading to induction of an inflammatory response and the subsequent elimination of the microorganism. Crucially, to prevent unwanted damage, organisms have evolved multiple mechanisms that tightly control the activation and resolution of inflammation. Significantly, we have previously shown that lncRNAs might represent a novel class of genes that regulate the inflammatory response (Nature Commun 5, 3978 (2014), Front Immunol 5, 1038 (2017), Arthritis Rheum 72, 609 (2021)).

The aim of this project is to employ bioinformatics and machine learning/AI approaches to identify the 'domains' through which lncRNAs regulate inflammation. Initially, the student will employ bioinformatics approaches to identify lncRNAs that are differentially expressed following activation of the innate immune response and associated inflammation in different human cell types and across vertebrate species ranging from zebrafish to apes. This data will then be analysed using the machine learning/AI to identify the 'domains'. This exciting computer-based project will be supervised by Professor Mark Lindsay (Bath) and Dr Fabrizio Costa (Exeter) and will provide training in broad range of areas including programming (Unix, python, R), bioinformatics, data science, machine learning/AI, immunology/inflammation and RNA biology. We are looking for somebody that is keen to learn and undertake research. Some background in mathematics/computing is desirable (but not essential), as training will be provided.

Our aim as the SWBio DTP is to support students from a range of backgrounds and circumstances. Where needed, we will work with you to take into consideration reasonable project adaptations (for example to support caring responsibilities, disabilities, other significant personal circumstances) as well as flexible working and part-time study requests, to enable greater access to a PhD. All our supervisors support us with this aim, so please feel comfortable in discussing further with the listed PhD project supervisor to see what is feasible.