

Project Details	
Project Code	MRCPHS25Ex Hannon
Title	Combining technologies to elucidate the epigenome's impact on health and disease
Research Theme	Population Health Sciences
Summary	The epigenome's dynamic nature and its responsiveness to environmental changes make it particularly attractive in the study of health and disease. This interest has resulted in various technologies for mapping it in individual samples, such as microarrays and Nanopore sequencing. This data science project aims to optimize and combine the benefits of these technologies to generate detailed maps of the epigenome across multiple individuals. By doing so, we can better understand how to make best use of the tools available to advance our understanding of how the epigenome influences the development of disease.
Description	<p>Background: Genetic studies have concluded that gene regulation, which controls when and in which cells a gene is active, is a key mechanism for understanding the development of diseases such as Alzheimer's disease, diabetes, and heart disease. The epigenome is a key component of gene regulation. It consists of a diverse range of modifications that attach to DNA and manipulates when and where genes are active. Unlike your DNA which stays the same throughout life, the epigenome is dynamic. It changes in response to both genetics and the environment. Therefore, it is of significant interest to researchers aiming to identify which genes play a role in the development of disease.</p> <p>This interest has been noticed by biotechnology companies, and there are multiple tools available to profile the epigenome. This includes Illumina microarrays which are affordable for large numbers of samples but only profile a small fraction (3%) of the epigenome and Oxford Nanopore sequencing which is very expensive but data rich giving an almost complete map of the epigenome in a single sample. The overarching objective of this data science PhD project is to determine the optimal approach for generating the most comprehensive map of the epigenome across multiple individuals while minimising experimental cost. This will look to combine the technologies available, harnessing their individual strengths and propose a strategy that could be applied to identify novel differences in the epigenome associated with disease and the genes affected by these.</p> <p>Specific objectives:</p> <ol style="list-style-type: none"> 1. Develop a framework for using Nanopore sequencing to identify positions in the epigenome associated with disease. This will involve using simulations to model the effect of sample size and magnitude of effect on experimental parameters such as statistical power, true positive rate, false positive rate and financial and computational cost. 2. Evaluate the use of imputation to bridge between microarray and sequencing technologies to provide comprehensive maps of the epigenome in multiple individuals. Characterise how the amount of data from each technology influences accuracy and statistical robustness to determine the optimum study design.

	<p>3. Assess the performance of commonly used epigenetic biomarkers (e.g. epigenetic clocks and algorithms to predict smoking status, protein abundance, cellular composition) in Nanopore sequencing data. Propose strategies for translating these tools typically developed using data from microarrays to sequencing based technologies.</p> <p>4. Apply these findings to an epigenetic epidemiology question of the students choosing. This could be to identify positions in the epigenome associated with a specific disease such as diabetes or Alzheimer’s disease or that change in response to an environmental exposure such as cigarette smoking or air pollution.</p> <p>While objective 4 is completely open for the student to customise to suit their own specific interests, it should be noted that objectives 1-3 are not sequential. Therefore, the student will have opportunity to prioritise these, depending upon not only their interest but the skills and techniques they want to develop. The project will involve a combination of data simulations and analysis of existing data. These data are available not only through the supervisory team and their networks, but there is additionally a lot of appropriate data in the public domain. This resource again provides the student with an opportunity to tailor the direction of the project depending upon the data they find.</p> <p>What we are looking for:</p> <p>This project would suit someone interested in developing advanced bioinformatics and data science skills. The student will be immersed in the dynamic and rapidly evolving fields of epigenetics and epidemiology while having the opportunity to work with cutting-edge technologies like Nanopore sequencing. At the end of the project the student will have acquired a host of highly desirable transferable skills. We can offer you the opportunity to work with large datasets on high-performance computing clusters, to sharpen up your coding skills in UNIX, R and/or Python and to master statistical analysis. There will be a strong emphasis on open and reproducible practices throughout, providing insight into collaborative software development using version control.</p>
--	---

Supervisory Team	
Lead Supervisor	
Name	Dr Eilis Hannon
Affiliation	Exeter
College/Faculty	Health and Life Sciences
Department/School	Clinical and Biomedical Sciences
Email Address	e.j.hannon@exeter.ac.uk
Co-Supervisor 1	
Name	Professor Jonathan Mill
Affiliation	Exeter
College/Faculty	Health and Life Sciences
Department/School	Clinical and Biomedical Sciences
Co-Supervisor 2	
Name	Dr Amy Webster
Affiliation	Exeter
College/Faculty	Health and Life Sciences
Department/School	Clinical and Biomedical Sciences

Co-Supervisor 3	
Name	Dr Josine Min
Affiliation	Bristol
College/Faculty	Faculty of Health Sciences
Department/School	Population Health Sciences, Bristol Medical School