Project Details	
Project Code	MRCPHS25Ex Hawkes
Title	Understanding genetic modifiers of obesity and metabolic disease
Research Theme	Population Health Sciences
Summary	A person's health is influenced by their socioeconomic status, smoking and other behavioural and environmental factors, as well as their genetic make-up. Some genetic variants directly affect protein sequence, while others act as switches that turn protein production on and off, depending on cell and life stage. But both, alone and combined, can affect a person's health. In this project, you will use recently released data on 3,000 proteins from 50,000 participants with their full genetic sequence to investigate how environmental and genetic effects combine to affect an individual's metabolic disease risk profile, in particular diabetes and obesity.
Description	In the last few years, large-scale genetic association studies, including exome and genome sequencing, of complex traits and diseases have increased our understanding of the contribution of genetic variants to an individual's health and disease risk. Despite this success, there are many examples where individuals who inherit the same disease-causing genetic variant are unaffected. In rare monogenic disease, this phenomenon is referred to as incomplete penetrance. There may be significant phenotypic heterogeneity even within a family of individuals carrying the same damaging variant. Understanding the mechanisms of how this occurs is important as it may help identify those most at risk of disease who can be prioritised for preventative interventions or targeted treatment. Current hypotheses that could explain incomplete penetrance include additional compensatory coding variants, non-coding variants affecting gene regulation, environmental exposures, epigenetic alterations or epistatic interactions. However, detecting and classifying the effect of interaction and modifier variables in the context of heritable human phenotypes is extremely difficult, largely due to poor statistical power and non-additive effects. The recent release of data on circulating protein levels, in combination with whole-genome sequencing (WGS) in UK Biobank (UKB), provides an unprecedented opportunity to test the impact of modifying and interaction variables. Rare variants can have huge effects on circulating protein levels, making them an exemplar trait for detecting interaction effects. Further, WGS allows us to examine the role of non-coding variants, including intronic and intergenic regulatory elements. Crucially, WGS covers the entire allele frequency spectrum in a population, including variants that are observed only once or twice even in a very large sample. Our work on the UKB data proteomic data has shown that non-coding variants can have effects as large as coding variants, but how they interact with coding variation - either on the same or the opposite haplotype - or with environmental exposures has not yet been explored. In this project, the student will use the unique proteomic dataset, combined with high-resolution WGS and rich phenotypic data available in 50,000 individuals in UK Biobank, to investigate modifiers of rare large-effect variants.

	Whilst this project is initially focused on understanding obesity and metabolic conditions, it is designed such that the PhD student will have the freedom to pivot towards other diseases of interest to study, provided there is a suitably related protein in the UKB proteomic data. However, this will be strictly after initial analyses applied to metabolic and diabetes-related measures of health, matching the expertise of the supervisory team. There is also sufficient scope for the student to focus primarily either on developing novel statistical methods or understanding biological disease mechanisms. Research Question: How can modifying factors in an individual's genome and environment alter their risk of disease? Aims: 1. Assess the impact of compound-heterozygous variants effects, including both coding and regulatory variants, on common and rare disease The student will use the phased UKB WGS data to investigate the impact of compound heterozygote effects. Phasing allows us to infer whether variants are on the same physical DNA strand as each other or not, and thus whether they affect one or both copies of a gene. Variants may have compensatory or epistatic effects, so different models will be explored to evaluate whether variants act additively or non-additively, or via standard modes of inheritance (e.g. dominant, recessive). 2. Assess the impact of environmental-interaction effects on rare damaging variants in genes associated with obesity or metabolic disease The environment a person grows up in, and lives in as an adult, is predicted to alter disease risk interactively with genetic variants. For example, body-mass-index (BMI) and waist-hip ratio (WHR) and lipid levels (e.g. triglycerides), which are often seen as general measures of human health that are strongly influenced by the environment, can alter the threshold for metabolic disease in combination with rare disease-causing variants. However, detecting these multiple, sometimes small interaction effects is extremely difficult, because of non-linear, non-additive, or threshold effects. The student will use the rich phenotypic data, combined with high-quality WGS and proteomic data, to begin quantifying the magnitude of environment-interaction effects on rare damaging variants. 3. Multi-ancestry analysis of interaction effects Having developed association models using the proteomic data, the student will explore interaction effects at the disease-association level (e.g. MC4R loss-of-function variants on BMI). By analysing non-proteomic data, the student will gain access to the All of Us biobank (N>250,000; increasing at time of writing). Individuals in All of Us are a considerably more diverse population than UKB: UKB is enriched for individuals of British descent with above-average health and income. The student will therefore identify ancestry-specific interaction effects, in particular ancestry-specific environmental-interaction effects.

	In the final stage of the PhD, the student will leverage the models they have developed in the two previous analysis stages to jointly model genetic and environmental interaction and modifying effects.
Supervisory Team	
Lead Supervisor	
Name	Dr Gareth Hawkes
Affiliation	Exeter
College/Faculty	Faculty of Health and Life Sciences
Department/School	Clinical and Biomedical Sciences
Email Address	g.hawkes2@exeter.ac.uk
Co-Supervisor 1	
Name	Professor Caroline Wright
Affiliation	Exeter
College/Faculty	Faculty of Health and Life Sciences
Department/School	Clinical and Biomedical Sciences
Co-Supervisor 2	
Name	Dr Aimee Hanson
Affiliation	Bristol
College/Faculty	Faculty of Health Sciences
Department/School	Bristol Medical School
Co-Supervisor 3	
Name	Professor Inês Barroso
Affiliation	Exeter
College/Faculty	Faculty of Health and Life Sciences
Department/School	Clinical and Biomedical Sciences