Project Details	
Project Code	MRC23PHSBr Haycock
Title	Using human genetics to identify novel biomarkers for enhanced prediction and early detection of cancer
Research Theme	Population Health Sciences
Summary	The student will use a novel human genetics approach ("Reverse Mendelian randomization") to identify plasma protein markers for early detection of cancer. Findings will be validated using direct protein measures in the UK Biobank study of 500,000 adults. The clinical application of these findings will be explored by integrating biomarkers into cancer prediction models with NHS data. These findings will guide and strengthen development of cancer prediction models.
Description	Background Cancer is the 2nd leading cause of death globally. Despite important advances in treatment in recent decades, the prognosis for many cancers remains poor because individuals typically present with advanced disease at which point curative treatment is no longer an option. The NHS Long Term Plan aims to increase the percentage of cancers caught at an early stage, thus potentially curable, from 50-75% by 2028. Improvements in early detection and prediction of cancer is an important vehicle to achieve that goal. Proteomic Reverse Mendelian randomization (RMR) is a novel human genetics approach that uses germline genetic variants to identify circulating proteins associated with a genetic tendency to develop disease. The approach can be used to identify circulating proteins implicated in disease aetiology as well as non-causal "bystander" proteins that may nonetheless be informative biomarkers for disease prediction and detection. In this project, the student will use a Proteomic RMR framework to identify proteins associated with a genetic predisposition to develop cancer. These findings will be validated using direct protein measures and cancer data in UK Biobank (UKB). The potential clinical application of the identified proteins will be assessed by integrating circulating protein data into current clinical cancer risk prediction models, using linked medical records in UKB. Objectives The objectives of this project are: 1) Use Proteomic RMR to evaluate the association of genetic predisposition to cancer with 4907 circulating proteins. 2) For the most promising proteins identified in Objective 1, examine association of directly measured proteins with cancer risk in 500,000 adults in UKB. 3) Determine if the predictive ability of current clinical risk prediction models for cancer used in primary care could be improved with the addition of data on candidate proteins identified in Objectives 1 & 2. Methods Objective 1: Proteomic RMR analyses To begin with, the student will prioritise the following cancers: prostate, breast, lung, and colorectal. Time permitting, this will be expanded to include pancreatic, ovarian, and overall cancer and acute myeloid leukaemia. The student will develop a genetic instrument for cancer predisposition using genome-wide significant ($P<5x10-8$) & independent ($r2<0.001$) single-nucleotide polymorphisms associated with cancer risk. The student will then test the association of genetically-instrumented cancer predisposition with 4907 proteins using inverse-variance weighted models. Sensitivity analyses will be performed to evaluate whether associations reflect a causal effect of genetic predisposition to cancer on

	proteins (i.e. rather than non-causal shared genetics with proteins). Objective 2: Analysis of directly measured proteins For top findings, the student will examine the association of directly measured circulating proteins with cancer risk. Cases will be UKB patients with the cancer of interest, eligible for primary care linkage, and matched to controls (matching on age, sex & Townsend deprivation score). Multivariable conditional logistic regression models will be used to estimate cancer risk. Objective 3: Clinical applicability of protein markers Four case-control studies will be performed to develop risk prediction models for cancer, using UKB and linked primary care data. Predictor variables will be clinical features of cancer (symptoms plus test results) currently used in primary care, patient demographics and comorbidities, plus candidate proteins identified in the PhD. Backwards elimination will be used to determine if candidate proteins (identified in objectives 1 & 2) add significant predictive power to the model, over and above the predictors currently used in clinical practice. Model calibration and discrimination will be estimated. Internal validation will be assessed with k-fold cross validation.
Supervisory Team	
Lead Supervisor	
Name	Dr Philip Haycock
Affiliation	Bristol
College/Faculty	Bristol Medical School
Department/School	Population Health Sciences
Email Address	philip.haycock@bristol.ac.uk
Co-Supervisor 1	
Name	Dr James Yarmolinsky
Affiliation	Bristol
College/Faculty	Bristol Medical School
Department/School	Population Health Sciences
Co-Supervisor 2	
Name	Dr Sarah Bailey
Affiliation	Exeter
College/Faculty	University of Exeter Medical School
Department/School	Institute of Health Research
Co-Supervisor 3	
Name	Dr Karl Smith-Byrne
Affiliation	University of Oxford
College/Faculty	
Department/School	