Predicting who is at high risk for a particular disease based on their genetics is an attractive idea, but early attempts at finding genetic variants that explain a substantial proportion of disease risk have been disappointing. Additionally, these investigations are of limited use, as the genetic code cannot (currently!) be changed. However, if a genetic variant predicts increased levels of a modifiable risk factor, and the same variant is also associated with increased disease risk, then this provides a potential causal pathway by which disease risk could be decreased – namely by reducing levels of the risk factor. If several genetic variants that have statistically robust and plausibly specific associations with the risk factor are all concordantly associated with the outcome, then the likelihood of the risk factor being a good target for intervention (a causal risk factor) increases. But cases in which such genetic evidence is unequivocal are rare – in the majority of cases, there is substantial uncertainty in inferring the causal role of risk factors.

Several approaches have been proposed for making causal inferences using genetic variants (an approach known as Mendelian randomization) more robust. This project will investigate the potential role of polygenic risk scores comprising large numbers of variants in assessing causal relationships. Such scores will typically explain a greater proportion of the variance in risk factors, and hence increase the power of an investigation. However, including large numbers of variants may impact the specificity of such scores. The project will consist of empirical analyses in UK Biobank, a large publicly-assessible resource, to investigate the properties of these scores.

This project is likely to involve a substantial amount of programming, and would suit someone interested in bioinformatics and exploiting “big data”.

Title: How useful are genetic risk scores for understanding disease processes?

Supervisor: Stephen Burgess