Investigating the Role of Inflammatory Biomarkers and Incretins in the Aetiology of Type 2 Diabetes and Coronary Heart Disease using Human Genetics

Nicholas Grant Bowker

MRC Epidemiology Unit

Hughes Hall, University of Cambridge

November 2020

This dissertation is submitted for the degree of

Doctor of Philosophy

Declarations

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

It does not exceed the word limit of 60,000 words prescribed by Degree Committee of the Faculties of Clinical Medicine and Veterinary Medicine.

Summary of the thesis

Title: Investigating the Role of Inflammatory Biomarkers and Incretins in the Aetiology of Type 2 Diabetes and Coronary Heart Disease using Human Genetics.

Name: Nicholas Grant Bowker

Background: The relevance of inflammatory and incretin-related signalling pathways in the aetiology of cardiometabolic diseases is of considerable pharmacological interest but remains uncertain. Evidence from animal models and epidemiological studies point to a role for chronic inflammation for the pathophysiology of type 2 diabetes (T2D), with interleukin-6 (IL-6) proposed as a key player. Incretins such as glucagon-like peptide-1 (GLP-1) and gastric inhibitory polypeptide (GIP) are hormones that stimulate a decrease in blood glucose. Their receptors are existing T2D drug targets, yet the efficacy and safety of GIP mono- and dual agonists e.g. tirzepatide are still unknown. Genetic approaches can help to address limitations of earlier studies and systematically assess the roles of IL-6 and GIP mediated signalling for cardiometabolic diseases.

Aims: To investigate the potential causal roles of IL-6 and GIP receptor signalling for the risk of cardiometabolic diseases, specifically T2D and coronary heart disease (CHD), through analysis of large-scale genetic data from patient and population-based studies.

Methods: 1) Systematic literature searches were conducted for each topic and the existing evidence summarised in the introduction. 2) A partial loss-of-function missense variant (Asp358Ala) in the IL-6 receptor gene (*IL6R*) was used to estimate the effect of IL-6R inhibition on T2D risk in 260,614 cases and 1,350,640 controls. 3) An observational meta-analysis of new unpublished and published studies (5,421 T2D cases, 31,562 non-cases) was conducted to compare observational and genetically predicted effects of IL6 levels on T2D. 4) The specificity of the *IL6R* variant was tested by including genetic and observational associations from a range of other inflammatory markers. 5) Large-scale genomic data from 23 cardiometabolic diseases as well as anthropometric, lipid, glycaemic and ~6,000 'omic (metabolomic and proteomic) traits were brought together to systematically assess associations of a known missense variant in *GIPR*, E354, and infer specificity and potential beneficial or harmful effects of GIPR mono agonism. Bayesian multi-trait colocalisation was used to distinguish trait clusters driven by shared causal variants to identify independent causal variants driving specific trait associations.

Findings: IL-6 levels (both measured and genetically predicted) were associated with T2D risk, with a small but significant effect (odds ratio (OR) 95% confidence interval (CI) for the Asp358Ala partial loss of function variant 0.98 (0.97, 0.99); $p=2x10^{-7}$). Genetic mediation analyses estimated that IL-6 levels mediated up to 5% of the association between BMI and T2D. Colocalisation results at the *GIPR* locus identify E354 as the driver of a shared signal for GIP (higher), T2D and related traits (lower risk), and adiposity (higher) with high posterior

probability (>0.97), and demonstrate that this is distinct from cardiovascular and lipid associations nearby in *APOE* (rs7412; PP >0.99).

Interpretation: Large-scale genetic and prospective observational data provide evidence that IL-6 mediated inflammation is implicated in T2D aetiology but suggest that the impact of this pathway on disease risk in the general population is likely to be small. At the *GIPR* locus, distinct genetic signals were shown to drive associations of glycaemic and adiposity traits versus CHD and lipid traits. This study provides evidence that inclusion of GIPR agonists in dual agonists could potentiate the protective effect of GLP1 agonists on diabetes without undue cardiovascular risk, while the effects of GIP on weight gain are counteracted by GLP-1.

Acknowledgements

I would like to begin this thesis by thanking everyone who has supported me in so many ways throughout my PhD. Firstly, I would like to thank my two PhD supervisors, Dr. Claudia Langenberg and Dr. Luca Lotta, whose patient guidance and expertise helped shape this work and the researcher that I have become. I cannot thank you both of you enough for the opportunity to work with you and for your patience and guidance throughout this process. Claudia, thank you for always challenging me and pushing me to focus on the bigger picture, allowing me the opportunity to lead a collaborative research effort and for giving me the freedom and confidence to pursue my research interests. Luca, from the very start of my PhD your guidance, patience with my writing style and expertise have helped shape the person that I am today. You have given me the self-belief and advice to pursue my goals and contributed immensely to my academic and personal development. You've always been a friend and your advice and guidance have shaped me in so many ways. After all, it's not often that your supervisor renames you midway through your PhD…

Secondly, I'd like to thank Dr. Maik Pietzner and Dr. Eleanor Wheeler who have been immensely helpful at so many stages throughout this PhD. You've both taught me so much throughout this process and I can't tell you how grateful I am for your willingness and availability to provide help and advice, no matter how daft the question might be. It's been a pleasure working with you both.

Thirdly, I'd like to thank Prof Nick Wareham, the director of the MRC-Epidemiology Unit, for his direction and the research support that he helped secure for me. My funding situation was certainly unclear before I joined the Unit, but his willingness to help me will be something I will always be grateful for. I would also like to thank all the members of the aetiology group, for sharing their knowledge and expertise and for being such a supportive group. It has been a true pleasure working with you all.

I'd like to thank the other PhD students who have shared this journey with me, particularly Hannah, Ignacio, Lina and Vicky for their friendship and support.

I am very grateful for my family, whose incessant support, belief and encouragement have got me to where I am today. Finally, to Kiera, thank you for your unwavering love and support, and for always believing in me. You've helped me in more ways than I can mention. Your support and motivation mean everything to me, and I will be forever grateful for you.

Publications and presentations

Publications

Accepted or submitted publications related to the work presented in this thesis:

Bowker N, Shah RL, Sharp SJ, Luan J, Stewart ID, Wheeler E *et al.* Meta-analysis investigating the role of interleukin-6 mediated inflammation in type 2 diabetes. ***EBioMedicine*** 2020; **61**: 103062.

Bowker N, Hansford H, Burgess S, Foley CN, Au Yeung V, Erzurumluoglu AM *et al.* Genetically predicted glucose-dependent insulinotropic polypeptide (GIP) levels and cardiovascular disease risk are driven by independent mechanisms at the GIPR. *Submitted as an original research article to* ***Science Translational Medicine***.

Publications co-authored during this thesis:

Lotta LA, Stewart ID, Sharp SJ, Day FR, Burgess S, Luan J, **Bowker N** *et al.* Association of Genetically Enhanced Lipoprotein Lipase-Mediated Lipolysis and Low-Density Lipoprotein Cholesterol-Lowering Alleles with Risk of Coronary Disease and Type 2 Diabetes. ***JAMA Cardiol.*** 2018; 3: 957.

Lotta LA, Mokrosiński J, Mendes de Oliveira E, Li C, Sharp SJ, Luan J, Brouwers B, Ayinampudi V, **Bowker N** *et al.* Human Gain-of-Function MC4R Variants Show Signaling Bias and Protect against Obesity. ***Cell*** 2019; 177: 597-607.e9.

Lotta LA, Wittemans LBL, Zuber V, Stewart ID, Sharp SJ, Luan J, Day FR, Li C, **Bowker N** *et al.* Association of Genetic Variants Related to Gluteofemoral vs Abdominal Fat Distribution With Type 2 Diabetes, Coronary Disease, and Cardiovascular Risk Factors. ***JAMA*** 2018; 320: 2553.

Presentations

Oral presentations

The Eli Lilly & Co. GIP Symposium – Virtual (October 2020)

“Genomic Approaches to Studying the Cardiometabolic Profile of GIP”

Poster presentations

The Danish Diabetes Academy (DDA) Summer school – Ebberup, Denmark (August 2019)

“IL-6-mediated Inflammation and Risk of Type 2 Diabetes”.

Commonly used abbreviations

adj.: Adjusted for

BMI: Body mass index

CHD: Coronary heart disease

Chr: Chromosome

CI: Confidence interval

DEXA: Dual-energy X-ray absorptiometry

EA: Effect allele

EAF: Effect allele frequency

GIP: Glucose-dependent insulinotropic polypeptide

GLP-1: Glucagon-like peptide 1

GWAS: Genome-wide association study

HbA1c: Glycated haemoglobin

HDL: High-density lipoprotein

IFNγ: Interferon-γ

IL-6: Interleukin-6

IL-8: Interleukin-8

LD: Linkage disequilibrium

LDL: Low-density lipoprotein

MAF: Minor allele frequency

MR: Mendelian randomisation

N: Number of participants

OA: Other allele

OR: Odds ratio

PC: Principal component

Pos: Base-pair position

QC: Quality control

SD: Standard deviation

SE: Standard error

SNP: Single nucleotide polymorphism

T2D: Type 2 diabetes

TNFα: Tumour necrosis factor α

WHR: Waist-to-hip ratio

WHRadjBMI: Waist-to-hip ratio adjusted for BMI

Index of Figures

Figure 1.1: The age adjusted prevalence of diabetes in adults worldwide in 2019. **P9**

Figure 1.2: The glucose homeostatic model in healthy individuals. **P11**

Figure 1.3: The initial stages of atherosclerosis. **P12**

Figure 1.4: Progression of atherosclerotic lesions to plaques. **P13**

Figure 1.6: The healthy acute inflammatory response. **P18**

Figure 1.7: Schematic of the systematic literature review detailing the exclusion criteria and the number of articles excluded at each stage. **P28**

Figure 2.1: Heatmap of the correlations between the four cytokines and cardiometabolic risk factors. **P43**

Figure 2.2: Associations of the four cytokines with cardiometabolic risk factors per 1 SD higher cytokine levels. **P45**

Figure 2.3: Association of inverse-rank normal transformed cytokine levels with incident cardiometabolic diseases in EPIC-Norfolk. **P46**

Figure 3.1: Manhattan plot for the meta-analysis of each cytokine. **P61**

Figure 3.2: Stacked bar plot illustrating the incremental variance explained by each variant associated with cytokine levels. **P66**

Figure 3.3: Heatmap of the pairwise colocalisation between cytokine levels and cardiometabolic diseases. **P67**

Figure 4.1: Study design. **P77**

Figure 4.2: Association of IL-6 levels with type 2 diabetes and glycaemic traits. **P87**

Figure 4.3: Association of *IL6R* Asp358Ala with type 2 diabetes, glycaemic and anthropometric traits. **P89**

Figure 5.1: Correlation heatmap illustrating the pairwise genetic correlation of autoimmune diseases and cardiometabolic traits with inflammatory protein levels. **P107**

Figure 5.2: Scatterplot illustrating the effects of rs7412 and rs429358 on the set of inflammatory proteins which colocalised with T2D and CHD in *APOE*. **P111**

Figure 5.3: Volcano plot of the association between the *IL6R* missense variant (Asp358Ala; rs2228145) and all 4,979 human proteins from the SOMAscan v4 platform. **P113**

Figure 5.4: Volcano plots illustrating the association of polygenic scores for 13 cardiometabolic traits with levels of the 335 inflammatory proteins and four cytokines. **P116**

Figure 5.5: Volcano plots illustrating the association of polygenic scores for 13 cardiometabolic traits with levels of the 335 inflammatory proteins and four cytokines. **P119**

Figure 5.6: Bar plot showing the gene-sets that were significantly enriched in proteins that colocalised with either T2D or CHD. **P121**

Figure 6.1: Overview of insulin potentiation by GIP and GLP-1. **P129**

Figure 6.2: Schematic of the systematic literature search. **P137**

Figure 7.1: Associations between E354 (rs1800437) and cardiometabolic diseases endpoints, glycaemic traits, cardiovascular risk factors and lipids, anthropometric traits and biomarkers. **P154**
Figure 7.2: Similarity heatmap for each cluster at the *GIPR* locus across prior and threshold permutations. **P159**
Figure 7.3. Regional association plots depicting CHD lead variants in the *GIPR* region. **P162**
Figure S2.1: Associations of IFNγ, IL-6, IL-8 and TNFα with cardiometabolic risk factors per 1 SD higher cytokine levels, adjusting for fasting insulin levels. **P265**
Figure S2.2: Association of natural log transformed cytokine levels with incident cardiometabolic diseases in EPIC-Norfolk. **P266**
Figure S2.3: Association of deciles of inverse-rank normal transformed cytokine levels with incident T2D in EPIC-Norfolk. **P267**
Figure S2.4: Association of deciles of inverse-rank normal transformed cytokine levels with incident CHD in EPIC-Norfolk. **P268**
Figure S2.5: Association of cytokine levels with incident T2D, defined using an established definition of incident T2D. **P269**
Figure S2.6: Association of inverse-rank normal transformed cytokine levels with incident cardiometabolic diseases in EPIC-Norfolk. **P270**
Figure S2.7: Association of BMI, WHR and WHRadjBMI with incident cardiometabolic diseases in EPIC-Norfolk. **P271**
Figure S2.8: Interaction between cytokine levels and anthropometric traits on incident cardiometabolic diseases. **P272**
Figure S3.1: Regional association plot depicting the 10Mb region of chromosome 12 associated with TNFα. **P273**
Figure S3.2: Quantile-quantile plots for the respective cytokine meta-analyses. **P274**
Figure S4.1: Design of the observational and genetic mediation analyses. **P275**
Figure S4.2: Conditional quantile regression of CRP levels in EPIC-Norfolk. **P276**
Figure S4.3: Association of IL-6 levels in deciles with incident type 2 diabetes in EPIC-Norfolk. **P277**
Figure S4.4: Association of BMI deciles with incident type 2 diabetes in EPIC-Norfolk. **P278**
Figure S4.5: Association of Asp358Ala with type 2 diabetes and coronary heart disease by genotype in UK Biobank. **P279**
Figure S4.6: Association of Asp358Ala with type 1 diabetes. **P280**
Figure S4.7: Excluding type 1 diabetes cases using a 29-variant polygenic risk score. **P281**
Figure S4.8: Association of Asp358Ala with continuous metabolic traits. **P282**
Figure S4.9: Interaction of Asp358Ala with type 2 diabetes genetic risk factors on type 2 diabetes. **P283**

Figure S4.10: Projection of the range of possible effects of *IL6R* blocking therapy on the risk of type 2 diabetes in a primary prevention setting. **P284**
Figure S4.11: Funnel plot for publication bias assessment. **P285**
Figure S5.1: Correlation heatmap illustrating the pairwise genetic correlation between autoimmune diseases and cardiometabolic traits. **P286**
Figure S5.2: Volcano plot of the association between a polygenic score for higher triglyceride levels with levels of the inflammatory proteins which colocalised with T2D at non-pleiotropic loci. **P287**
Figure S5.3: Volcano plot of the association between a polygenic score for higher triglyceride levels with levels of the inflammatory proteins which colocalised with CHD at non-pleiotropic loci. **P288**
Figure S7.1: Association of E354 and cardiovascular disease sub-types in UK Biobank. **P289**
Figure S7.2: Associations between E354 and regional adiposity compartments in 435,387 participants measured by bio-impedance. **P290**
Figure S7.3: Associations between E354 and human protein levels. **P291**
Figure S7.4: Associations between E354 and human metabolite levels. **P292**
Figure S7.5. Gaussian graphical model illustrating the partial correlation network in 11,966 participants between X-12283 and first and second order connections most correlated with X-12283. **P293**
Figure S7.6: Stacked regional association plot showing the cluster of cardiovascular-related traits which colocalise near the *GIPR* locus. **P294**
Figure S7.7: Regional association plot illustrating the cluster of traits which colocalise with the GIP measures at the *GIPR* locus. **P295**
Figure S7.8: Heatmap matrix depicting the pairwise colocalisation between cardiometabolic traits at the *GIPR* locus. **P296**
Figure S7.9. Matrix illustrating the LD between each of the independent CHD variants and rs1800437 (E354) estimated using 5 European populations in LDlink. **P297**
Figure S7.10: Volcano plot showing the associations between rs1964272 and 4,979 human protein levels. **P298**

Index of Tables

Table 1.1: WHO diagnostic criteria for diabetes. **P8**

Table 1.2: Study characteristics of the GWAS studies estimating the association of genetic variants with IFNγ, IL-6, IL-8 and TNFα levels. **P29**

Table 1.3: Summary of the GWAS results for the four cytokines of interest. **P31**

Table 2.1: Participant characteristics for the Fenland and EPIC-Norfolk studies. **P41**

Table 2.2: Distribution parameters and assay performance across the four cytokines in Fenland and EPIC-Norfolk. **P42**

Table 3.1: Lead variants used for independent signal selection. **P60**

Table 3.2: Conditionally independent association signals for each cytokine. **P63**

Table 3.3: Bayesian fine mapping results showing the lead variant of each 99% credible set and its corresponding posterior probability. **P64**

Table 3.4: Correlation estimates between the four cytokines. **P65**

Table 3.5: Narrow-sense heritability estimates from LDSC. **P65**

Table 4.1: Summary of the study design. **P76**

Table 4.2: Study participants. **P85**

Table 5.1: Summary of the included traits and their data sources. **P101**

Table 5.2: Loci where T2D colocalised with inflammatory proteins. **P109**

Table 5.3: Loci where CHD colocalised with inflammatory proteins. **P110**

Table 5.4: Number of proteins which colocalise with disease at the 3 pleiotropy classes. **P112**

Table 6.1: Details of the studies included in the systematic review. **P138**

Table 6.2: Genetic variants associated with levels of GIP and GLP-1 in published studies. **P139**

Table 7.1: Study participants. **P146**

Table 7.2: Clusters of colocalised traits identified by the main and sensitivity analyses at default settings. **P158**

Table 7.3: Independent CHD variants identified using approximate conditional analysis. **P160**

Table S1.1: The search strategy used for the systematic literature review. **P217**

Table S3.1: Number of variants removed from each cytokine meta-analysis at each quality control step. **P218**

Table S4.1: Summary of the studies participating in the different analyses of the manuscript. **P219**

Table S4.2: Study characteristics and adjusted risk ratios from studies of incident type 2 diabetes associated with levels of IL-6. **P221**

Table S4.3: Systematic literature review search terms. **P223**

Table S4.4: Quality ratings of the studies included in the meta-analysis of association between IL-6 levels and incident type 2 diabetes. **P224**
Table S4.5: Fenland participant DEXA and glycemic measures. **P225**
Table S4.6: Standard deviation values used to convert estimates for continuous metabolic traits between clinical and standardized units and their source. **P226**
Table S4.7: Polygenic risk scores used in Stage 2 and their sources. **P227**
Table S4.8: Correlation of IL-6 levels with DEXA adiposity measures in Fenland. **P235**
Table S4.9: Variance and heritability in type 2 diabetes and coronary disease occurrence explained by *IL6R* Asp358Ala. **P236**
Table S4.10: Association of 358Ala with type 2 diabetes conditioning on rs2481065. **P237**
Table S5.1: Inflammatory proteins with evidence of an association with T2D or CHD. **P238**
Table S5.2: Sentinel variants for the 85 loci used in multi-trait colocalisation. **P246**
Table S5.3: Gene sets in which inflammatory proteins colocalising with T2D are significantly enriched. **P248**
Table S5.4: Gene sets in which inflammatory proteins colocalising with CHD are significantly enriched. **P251**
Table S6.1: Search strategy used for the systematic literature search on incretin GWAS studies. **P254**
Table S7.1: Summary of the participating studies. **P255**
Table S7.2: Clusters of colocalised traits identified by the main analysis across the permutations of prior 2 and the regional and alignment thresholds. **P257**
Table S7.3: Clusters of colocalised traits identified by the sensitivity analysis across the permutations of prior 2 and the regional and alignment thresholds. **P260**
Table S7.4: Association of E354 with the traits it was associated with at nominal significance from the 2SMR analysis, after conditioning on independent SNPs for each trait. **P263**
Table S7.5: Association of other previously identified fasting GIP variants with CHD. **P264**

Table of Contents

Declarations i

Summary of the thesis ii

Acknowledgements iv

Publications and presentations v

Publications v

Presentations v

Commonly used abbreviations vi

Index of Figures vii

Index of Tables x

Introduction 1

Aims 1

Outline of the thesis 4

Rationale 6

Chapter 1: Overview of the literature 7

1.1: Background 7

Definition of type 2 diabetes and coronary heart disease 7

Worldwide impact of cardiometabolic diseases 8

Hyperglycaemia and the initiation of insulin resistance in T2D pathogenesis 9

The role of cholesterol in the initiation of atherosclerosis in CHD pathogenesis 11

1.2: Genetics as a framework for causal inference and risk factor prioritisation 13

1.3: Inflammation as a link between obesity and cardiometabolic diseases 16

Inflammation: the body's natural defence to insult 16

Inflammatory mechanisms linking obesity to T2D and CHD 17

Observational evidence establishing cytokines as therapeutic targets of interest 19

Clinical trials targeting cytokines in cardiometabolic diseases 22

Overview of the genetic architecture of IFNγ, IL-6, IL-8 and TNFα levels in population-based studies 26

Chapter 2: Observational Analyses of Cytokine Levels with Cardiometabolic Diseases 32

Contributions and collaborations 32
Abstract 33
2.1: Introduction 34
2.2: Methods 34
Study participants 34
Cytokine measurement and phenotype generation 35
Anthropometric measurement and quantification of blood biochemical markers 36
Correlations between cytokine levels and cardiometabolic risk factors 37
Associations between cytokine levels and cardiometabolic risk factors 37
Associations between cytokine levels and risk of incident CHD and T2D 38
2.3: Results 40
Study participants' characteristics 40
Cytokine measurement using the MSD platform 40
Correlations and observational association between cytokine levels and cardiometabolic risk factors 42
Associations of cytokine levels with incident cardiometabolic diseases 44
2.4: Discussion 47
Chapter 3: Discovery, Refinement and Characterisation of Loci Associated with Cytokine Levels 51
Contributions and collaborations 51
Abstract 52
3.1: Introduction 53
3.2: Methods 53
Phenotype preparation 53
GWAS for IFNγ, IL-6, IL-8 and TNFα levels 54
Meta-analysis and quality control of GWAS analyses 54
Regional association plots and independent signal selection 54
Bayesian fine mapping 55
Correlation between cytokines and variant-based heritability estimation 56
Proportion of variance explained by variants associated with cytokine levels 57

Colocalisation of variants associated with cytokine levels and cardiometabolic diseases 57
3.3: Results 59
GWAS and meta-analysis for loci associated with cytokine levels 59
Independent signal selection using formal conditional analysis 61
Bayesian fine mapping of independent association signals 61
Correlation and variant-based heritability across the four cytokines 64
Incremental variance explained by variants associated with cytokine levels 65
Colocalisation of cytokine variants with cardiometabolic diseases 66
3.4: Discussion 67
Chapter 4: Meta-analysis Investigating the Role of Interleukin-6 Mediated Inflammation in Type 2 Diabetes 72
Contributions and collaborations 72
Publications related to this chapter 72
Abstract 73
4.1: Introduction 74
4.2: Methods 75
Study design 75
Studies and participants 77
Exposure and outcome variables 78
Statistical analysis 80
4.3: Results 84
IL-6 levels, anthropometric and glycaemic measures and risk of T2D 84
Genetically impaired IL6R signalling and T2D risk 87
Mediation by the IL-6 pathway of cardio-metabolic disease risk associated with higher adiposity 89
4.4: Discussion 90
Chapter 5: Estimating the Inflammatory Proteins and Pathways Implicated in Cardiometabolic diseases: A Multi ‘Omics Approach 93
Contributions and collaborations 93

Abstract 94
5.1: Introduction 95
5.2: Methods 97
Profiling of the plasma proteome 97
Inflammatory protein identification 98
Genetic correlation between inflammatory proteins and cardiometabolic traits 98
Reciprocal lookup of independent variants in inflammatory protein targets and cardiometabolic disease GWAS summary statistics 101
Multi-trait colocalisation between inflammatory proteins and cardiometabolic traits 101
Association of candidate variants with all human proteins 103
Association between polygenic scores for cardiometabolic traits and inflammatory protein levels 103
Gene-set enrichment analysis 104
5.3: Results 105
Reciprocal lookup of cardiometabolic and inflammatory GWAS sentinels 107
Multi-trait colocalisation between inflammatory proteins and cardiometabolic traits 107
Association of polygenic scores for cardiometabolic traits with inflammatory protein levels 113
Gene-set enrichment analysis 118
5.4: Discussion 121
Chapter 6: Incretins and the incretin effect 126
Contributions and collaborations 126
6.1: The physiological roles of incretins and their alterations in obesity and T2D 127
The physiological roles of incretins 127
The incretin effect in health and T2D 129
The role of GLP-1 in satiety, food consumption and weight loss 129
The role of GIP in obesity, dyslipidaemia and risk of T2D 130
6.2: The role of incretin levels in cardiovascular disease 131
6.3: Incretin-based therapies for cardiometabolic diseases 132
6.4: Overview of the genetic architecture underpinning incretin levels 135

Introduction 135
Methods 135
Results and discussion 136
Chapter 7: Genetically predicted glucose-dependent insulinotropic polypeptide (GIP) levels and cardiovascular disease risk are driven by independent mechanisms at the GIPR 140
Contributions and collaborations 140
Abstract 141
7.1: Introduction 142
7.2: Methods 144
Study design 144
Study participants 144
Genotyping and imputation 145
Profiling of the plasma proteome 145
Plasma metabolomic profiling 146
Statistical analysis 146
7.3: Results 152
Characterisation of a missense variant E354Q (rs1800437) in the *GIPR* 152
Associations of E354 with plasma protein and metabolite levels 154
Multi-trait colocalisation across cardiometabolic traits 154
Conditional analysis at the *GIPR* locus 159
7.4: Discussion 162
Chapter 8: Concluding Discussion 165
8.1: Summary of the findings 165
8.2: Strengths 168
8.3: Limitations 170
Approximation of pharmacological agonism or antagonism using genetic variants 170
Limitations to causal inference 171
Colocalisation analyses assume a single causal variant 172
Associations between genetic variants and phenotypes may be false positives, biased or study-specific 172

Statistical uncertainty 175
8.4: Directions for future research and clinical implications 176
The need for larger GWAS to facilitate the discovery of further variants associated with biomarkers 176
Expanding the scope to include different tissues 177
Consideration of other diseases 178
Investigation of incretins within the context of metabolic disease as a whole 178
A focus on the in-depth characterisation of genetic loci 179
Clinical implications 180
8.5: Conclusion 181
References 182
Supplementary Materials 210
Supplementary Methods 210
Supplementary Tables 216
Supplementary Figures 264

Introduction

Aims

The central aim of this PhD thesis is to identify the aetiological pathways underpinning the roles of inflammatory biomarkers and incretins, two areas of considerable pharmacological interest, in cardiometabolic diseases. This will be based on the integration of large-scale genomic data in deeply phenotyped cohorts with multiple 'omics datasets to further our understanding of the molecular underpinnings of type 2 diabetes (T2D) and coronary heart disease (CHD). The primary focus for sections investigating inflammatory biomarkers is on levels of four pro-inflammatory cytokines: interferon-γ (IFNγ), interleukin-6 (IL-6), interleukin-8 (IL-8) and tumour necrosis factor α (TNFα) but is later expanded to all inflammatory proteins.

I will address the following sub-aims and objectives:

- **Sub-aim 1:** To review the current observational, genetic and clinical literature on the aetiological role of inflammation in T2D and CHD.
    - Objective 1.1: Review and critically evaluate the literature to identify knowledge gaps which can be addressed.
    - Objective 1.2: Perform a systematic literature search to gain an understanding of the current knowledge of the genetic determinants of IFNγ, IL-6, IL-8 and TNFα levels.
- **Sub-aim 2:** To investigate the observational associations of IFNγ, IL-6, IL-8 and TNFα levels with cardiometabolic diseases and related risk factors.
    - Objective 2.1: Estimate the association between cytokine levels and risk factors for T2D and CHD.
    - Objective 2.2: Estimate the association between cytokine levels and incident T2D and CHD.
- **Sub-aim 3:** To conduct genome-wide association studies (GWAS) and meta-analysis to estimate the association between genetic variants and levels of IFNγ, IL-6, IL-8 and TNFα.
    - Objective 3.1: Perform conditional analyses and Bayesian fine mapping to identify independent secondary signals in each locus and the putative causal variants underlying each association signal.
    - Objective 3.2: Estimate the observational and SNP-based correlation between IFNγ, IL-6, IL-8 and TNFα as well as their SNP-based heritability

- Objective 3.3: Estimate the colocalisation between variants associated with cytokine levels and T2D and CHD.

- **Sub-aim 4**: To investigate the role of IL-6 and IL-6 receptor (IL-6R) mediated inflammation in the pathophysiology of T2D.
  - Objective 4.1: Conduct a systematic review of prospective studies investigating the association between IL-6 levels and incident T2D and meta-analyse estimates from these studies with estimates from this thesis.
  - Objective 4.2: Estimate the association of a loss-of-function missense variant (Asp358Ala) in *IL6R*, previously shown to mimic the effects of IL-6R inhibition, in a large trans-ethnic meta-analysis of T2D case-control studies.
  - Objective 4.3: Estimate the proportion of the association between body mass index (BMI) and risk of cardiometabolic diseases that is mediated by IL-6 levels on both an observational and genetic level.
  - Objective 4.4: Project the putative therapeutic efficacy of IL-6R antagonism on T2D risk by comparing the effects of Asp358Ala to evidence from clinical trials of established IL-6R antagonists.
- **Sub-aim 5**: To investigate the inflammatory proteins and the biological pathways involved in risk of cardiometabolic diseases by integrating large-scale proteomic and genetic data.
  - Objective 5.1: Identify inflammatory proteins and perform a reciprocal lookup of the sentinels for inflammatory proteins and cardiometabolic diseases.
  - Objective 5.2: Identify loci that are shared between inflammatory proteins and cardiometabolic diseases.
  - Objective 5.3: Perform a gene set enrichment analysis to estimate whether inflammatory proteins which colocalised with cardiometabolic diseases were enriched in particular pathways.
  - Objective 5.4: Estimate the association between polygenic scores for established cardiometabolic risk factors and inflammatory protein levels to determine whether inflammation is partly driven by established risk factors.
- **Sub-aim 6**: To review the current clinical and genetic literature on the aetiological role of incretins in T2D and CHD.
  - Objective 6.1: Review and critically evaluate the literature to identify knowledge gaps which can be addressed.
  - Objective 6.2: Perform a systematic literature search to review the current understanding of the genetic determinants underpinning incretin levels.

- **Sub-aim 7**: To estimate whether glucose-dependent insulinotropic polypeptide (GIP) receptor (GIPR) mediated signalling is associated with raised CHD risk.
    - Objective 7.1: Characterise the clinical consequences of the *GIPR* missense variant rs1800437 (E354).
    - Objective 7.2: Estimate distinct clusters of colocalised cardiometabolic traits driven by shared causal variants in the *GIPR* region.
    - Objective 7.3: Perform conditional analyses to identify variants contributing to associations between the *GIPR* locus and cardiometabolic traits of interest.

Outline of the thesis

Chapter 1 provides an overview of the current literature. I begin with a brief outline of both T2D and CHD, including their definition, global impact and aetiology. Next, I provide an overview and critical evaluation of recent genetic and integrative 'omics approaches to the identification of causal risk factors for cardiometabolic diseases as well as their uses in drug target discovery and validation. Following this, I provide an overview of chronic inflammation in cardiometabolic diseases and critically review existing observational and clinical evidence linking inflammation to cardiometabolic diseases. I finish with an overview of studies investigating the genetic architecture of four proinflammatory cytokine levels: IFNγ, IL-6, IL-8 and TNFα.

Chapter 2 investigates the observational association of levels of the four cytokines with incident cardiometabolic diseases and related risk factors. Firstly, the association between cytokine levels and anthropometric, glycaemic and lipid traits is estimated using multivariable linear regression in up to 10,335 participants from the Fenland study. Next, the association between cytokine levels and incident T2D and CHD is estimated using Cox regression in up to 7,514 participants from the EPIC-Norfolk cohort.

Chapter 3 describes a GWAS and meta-analysis of the four cytokine levels in approximately 17,000 participants from the Fenland and EPIC-Norfolk studies. I start by outlining the phenotype preparation and move to describe how the GWAS and meta-analysis were conducted, and the quality control measures used at each step. Next, I describe the independent signal selection to identify secondary signals at each locus and the Bayesian fine mapping to refine each association signal to a likely causal variant. Following this, I outline the SNP-based heritability and correlation estimation across the four cytokines using LD score regression. Finally, I describe a colocalisation analysis to estimate whether association signals are shared between any of the four cytokines and either T2D or CHD.

Chapter 4 details an in-depth investigation into the role of IL-6 mediated signalling risk of T2D. I begin with a description of a systematic review of 15 prospective studies investigating the association between IL-6 levels and T2D risk and the meta-analysis of estimates from these studies with estimates from this thesis in up to 5,421 cases and 31,562 non-cases. Next, I describe a trans-ethnic meta-analysis estimating the association of a loss-of-function missense variant in *IL6R*, Asp358Ala, with T2D in 260,614 cases and 1,350,640 controls. Following this, I describe an analysis estimating the proportion of the association between BMI and risk of cardiometabolic diseases that is mediated by IL-6 levels on both an observational and genetic level. Finally, I outline an analysis where I project the putative therapeutic efficacy of IL-6R antagonism on T2D risk by comparing the effects of Asp358Ala to evidence from clinical trials of established IL-6R antagonists.

Chapter 5 focusses on identifying the inflammatory proteins and biological pathways involved in the aetiology of cardiometabolic diseases by integrating large-scale genomic data with aptamer-based proteomic data. I begin by describing how inflammatory proteins were identified using GO terms and how a reciprocal lookup of sentinel variants was performed between inflammatory proteins and cardiometabolic diseases. I then describe how Bayesian multi-trait colocalisation was applied to identify candidate variants where inflammatory proteins colocalised with cardiometabolic diseases. I then describe a gene-set enrichment analysis that estimated whether the colocalised inflammatory proteins were enriched in particular pathways. Finally, I describe an analysis to determine whether inflammation is partly driven by cardiometabolic risk factors wherein I estimate the association between polygenic scores for established cardiometabolic risk factors and inflammatory protein levels.

Chapter 6 provides a critical review of the current literature surrounding the role of incretins in cardiometabolic diseases. I begin by describing the physiological role of incretins in cardiometabolic diseases and how this is perturbed by obesity and T2D. I then review recent epidemiological evidence linking incretin levels to raised cardiovascular disease risk. Following this, I review clinical studies and data from randomised-controlled trials targeting incretins as potential therapies for use in obesity and T2D. I end with an overview of studies investigating the genetic architecture underpinning incretin levels.

Chapter 7 describes genetic analyses to estimate whether higher GIPR-mediated fasting GIP levels are associated with raised CHD risk. I begin by describing two-sample Mendelian randomisation (2SMR) analyses integrating large-scale genomic data for cardiometabolic diseases, anthropometric, glycaemic, lipid and ~6,000 metabolomic and proteomic biomarkers to characterise the clinical consequences of the *GIPR* missense variant E354. Next, I describe how a Bayesian multi-trait colocalisation framework was used to estimate distinct clusters of cardiometabolic traits driven by shared causal variants in the *GIPR* region. Finally, I outline conditional analyses to identify independent variants driving associations between the *GIPR* locus and cardiometabolic traits of interest.

Chapter 8 is a concluding discussion which summarises the findings of this thesis, highlights the general limitations of research, discusses the clinical implications and outlines possible directions for future research.

Rationale

The establishment of large population-based biobanks such as UK Biobank together with reduced costs for genotyping and sequencing has facilitated an unprecedented increase in the scale and availability of genomic data. In addition, biobanks collect detailed clinical information such as electronic health records, lifestyle and anthropometric measures on all participants making them an invaluable resource for the “hypothesis free” investigation of the mechanisms underpinning complex disease risk. As a result, human genetic data is now being applied at scale for the purposes of drug development. Genetics provides a framework which is robust to non-genetic forms of confounding and reverse causality that may be applied efficiently to in drug development efforts to facilitate the discovery, characterisation, prioritisation and prediction of candidate drug target efficacy. Recent technological advances in ‘omics methods have facilitated the simultaneous measurement of thousands of molecular traits such as gene expression, protein and metabolite levels in large epidemiological cohorts. This has facilitated the integration of genomic data with ‘omics data at scale in deeply phenotyped cohorts. The availability of this large-scale ‘omics data in deeply phenotyped epidemiological cohorts at scale offers a timely opportunity to investigate the molecular underpinnings and pathways involved in cardiometabolic diseases such as T2D and CHD. While much of the aetiology of cardiometabolic diseases is largely understood, there are areas such as the roles of inflammatory biomarkers and incretins, which are unclear and of considerable pharmacological interest.

Chronic inflammation in cardiometabolic diseases is hypothesised to be one of the links between obesity and risk of cardiometabolic diseases, partly mediating this relationship. The causal relationships underpinning the role of inflammation in cardiometabolic diseases are still unclear, however, a role for inflammation in insulin resistance has been posited. It is unknown whether inflammation is an independent causal risk factor for cardiometabolic diseases or whether it is resultant of established risk factors such as obesity. In addition, the inflammatory proteins and pathways involved in cardiometabolic diseases are unknown.

Until recently, our knowledge of glucose homeostasis was largely limited to insulin and glucagon, however, the discovery of the insulin potentiating action of incretin hormones has sparked considerable interest in the pharmacological modulation of these hormones in T2D. However, recent epidemiological evidence has implicated higher GIPR-mediated fasting GIP levels in raising cardiovascular risk, jeopardising the development of therapies targeting this pathway. Investigation of the relationship between GIPR agonism and cardiovascular risk using genetics will provide an efficient, systematic, cost-effective method of assessing the cardiovascular safety of these therapeutics.

Chapter 1: Overview of the literature

1.1: Background

Definition of type 2 diabetes and coronary heart disease

Type 2 diabetes (T2D) is a chronic condition characterised by hyperglycaemia resulting from an inability to mount an adequate insulin response, defined as insulin resistance[1,2]. To cope with the hyperglycaemia, the body attempts to compensate by increasing insulin secretion relative to the degree of insulin resistance[1]. Diagnostic criteria set out by the WHO[2] state that diabetes may be diagnosed if one of the criteria for hyperglycaemia are met (**Table 1.1**).

Table 1.1: WHO diagnostic criteria for diabetes [a]

Risk factor	Threshold
Fasting plasma glucose	≥7.0 mmol/L (126 mg/dL)
Two-hour plasma glucose	≥ 11.1 mmol/L (200 mg/dL)
HbA1c	≥ 6.5%

Abbreviations: HbA1c; glycated haemoglobin

a. Adapted from the WHO Global Report on Diabetes 2016[2]

Coronary heart disease (CHD) is characterised by a narrowing or blockage of the coronary arteries over time via a process known as atherosclerosis[3]. In atherosclerosis, lipids and cholesterol within a fibrous matrix are ectopically deposited within the arterial wall, eventually forming atherosclerotic plaques[4]. Over time, this leads to an irregular arterial surface, causing the arteries to narrow, limiting the flow of blood and increasing the risk of myocardial infarction or stroke[3,4]. Diagnosis consists of blood tests to measure cholesterol levels, a major risk factor for CHD[4,5], which are followed by functional assessments such as electrocardiogram (ECG) or imaging techniques such as computed tomography (CT)[6].

Cardiometabolic disease is a collective term referring to the co-occurrence of several metabolic perturbations, chief among which are: insulin resistance or glucose intolerance, obesity (central adiposity), dyslipidaemia and hypertension[7,8]. Cardiometabolic disease represents a cluster of correlated risk factors that identify individuals at high risk for both T2D and CHD.

Worldwide impact of cardiometabolic diseases

Global increases in the prevalence of obesity preceded subsequent increases in the prevalence of both T2D and CHD, supporting the notion that obesity leads to a series of metabolic perturbations that have the capacity to mediate or accelerate processes underlying both diseases[9,10]. In 2019, the global prevalence of diabetes was estimated to be 9.3%, translating to 463 million people living with diabetes[1]. Of the three most common forms of diabetes, T2D is the most prevalent, accounting for approximately 90% of all cases[1]. By 2045, this is projected to increase to 700 million people[1] (**Figure 1.1**). This increase is not limited to countries of lower socioeconomic status, however, the predicted rate of growth is higher in low to middle income countries[2]. This places a huge burden on healthcare infrastructure amounting to an estimated $760 billion per year in 2019, a 4.5% increase on the estimate for 2017, and is likely to increase to $845 billion in 2045[1,11,12].

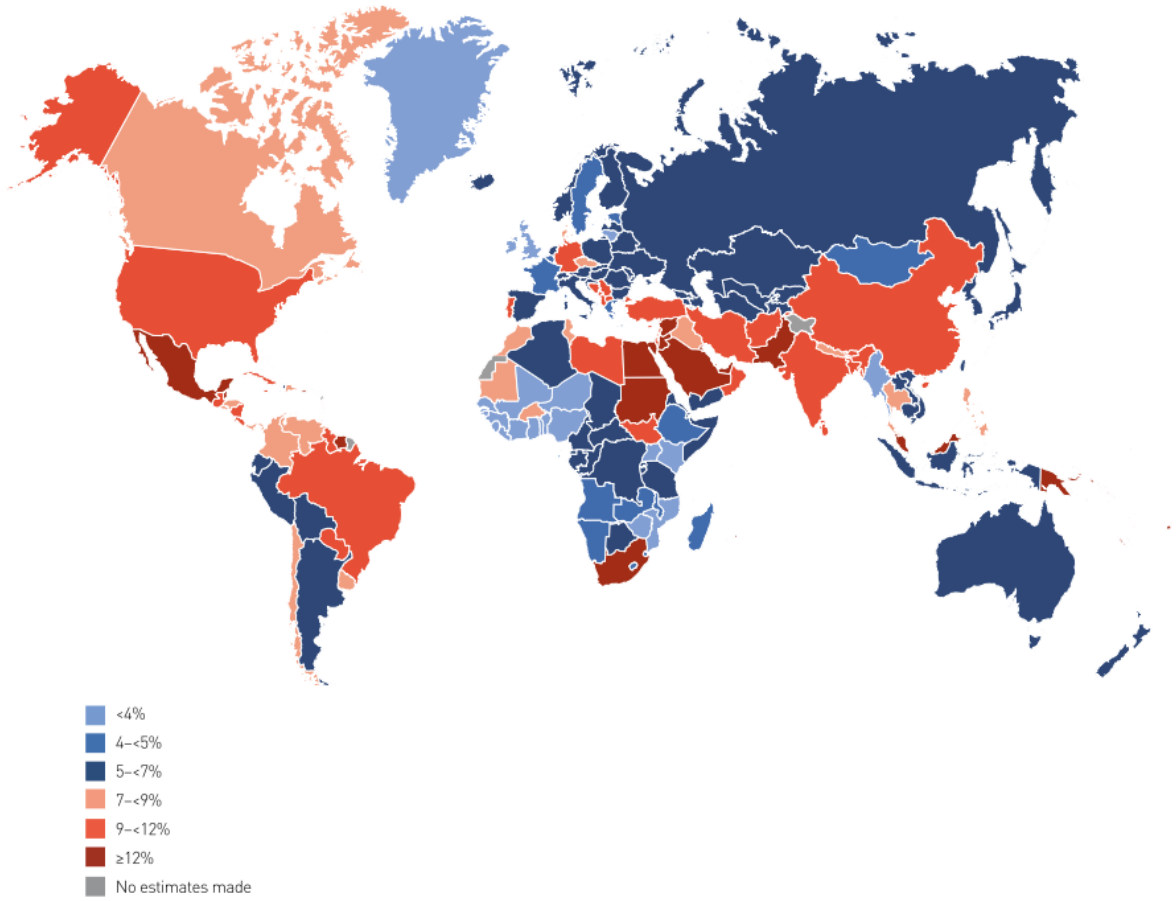


Figure 1.1: The age adjusted prevalence of diabetes in adults worldwide in 2019. Figure taken from the IDF Diabetes Atlas, 9th edition (2019)[1].

In line with the notion of obesity being a major driver for T2D and CHD, the prevalence of cardiovascular disease (CVD) has increased. CVD is the leading cause of all-cause mortality and places a huge burden on economies worldwide[13]. Globally, more than 75% of deaths resulting from CVD occur in low to middle income countries[13]. In 2017, 92 million people in the

USA were estimated to have at least 1 form of cardiovascular disease[14]. This is projected to increase to 44% of the adult population of the USA by the year 2030[14]. Empirical evidence of the correlation between T2D and CVD shows that T2D patients have between 1.5 to 3 times higher risk of developing a form of CVD compared to the general population[15]. T2D is also associated with earlier onset[16] and greater severity[17] of CVD events.

These global trends are driven by the current socioeconomic climate where greater consumption of energy-dense, convenience foods leads to chronic positive energy balance, a risk factor for obesity[1]. This effect is exacerbated by aging populations and a lack of physical activity, which leads to a more sedentary lifestyle, all of which are risk factors for obesity. Taking this into consideration, a better understanding of the mechanisms linking obesity to T2D and CHD is critical for the design of prevention and treatment interventions aimed at reducing the burden of these conditions on global health.

Hyperglycaemia and the initiation of insulin resistance in T2D pathogenesis

Glucose homeostasis in healthy individuals is governed by the key peptide hormone insulin and is maintained by a balance between insulin secretion and sensitivity[18]. In the healthy fasted state, blood glucose concentrations are maintained by balancing endogenous glucose production via glycogenolysis and gluconeogenesis in the liver and its subsequent use by tissues such as the brain (**Figure 1.2, lower half**)[19]. Endogenous glucose production provides a buffer against hypoglycaemia supplying glucose-dependent tissues such as the brain with glucose[10]. Tissues such as skeletal muscle that are able to metabolise other carbon sources, are provided with non-glucose substrates such as non-esterified fatty acids from adipose tissue lipolysis to preserve fuel for the brain[10,19]. Following a carbohydrate-rich meal, blood glucose levels are raised, providing the stimulus for pancreatic insulin secretion. Insulin-mediated glucose uptake into tissues such as skeletal muscle and adipose tissue follows, along with the inhibition of hepatic gluconeogenesis and adipose tissue lipolysis (**Figure 1.2, top half**)[10,18].

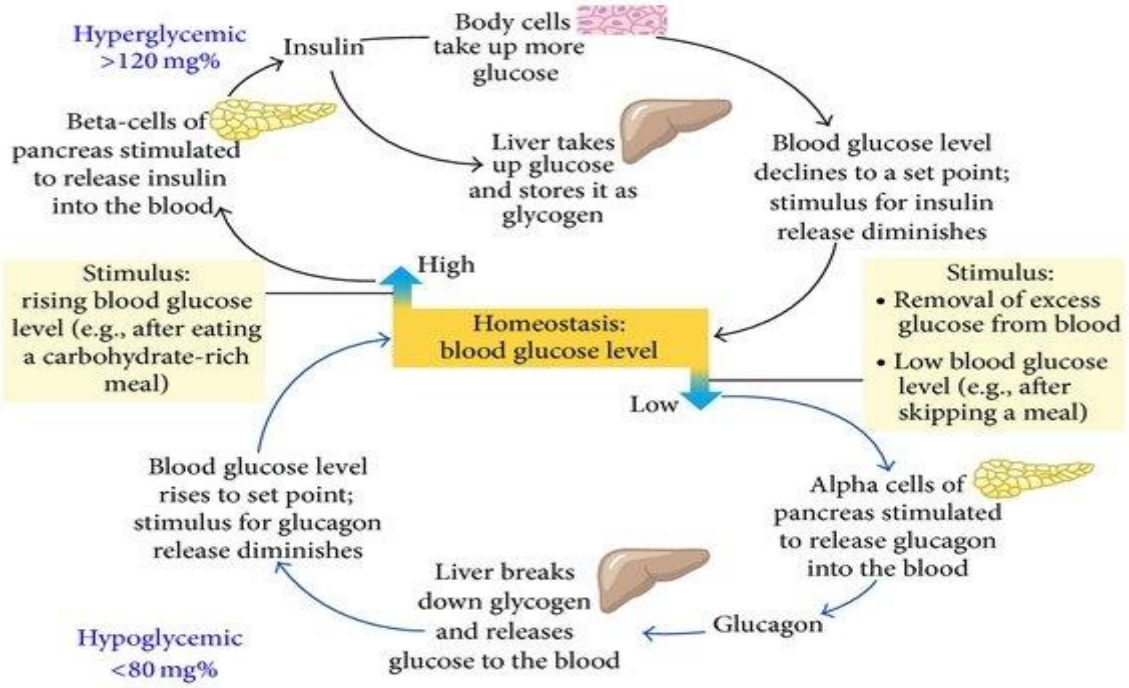


Figure 1.2: The glucose homeostatic model in healthy individuals. The lower half depicts glucose homeostasis in the fasting state. The top half shows the fed state. Taken from: Principles of Molecular Medicine, 16th Ed.[20]

In healthy individuals, pancreatic β cells are adaptable to variations in insulin action and are able to compensate accordingly[10,18,19]. When this compensatory capacity is exceeded, the expected effects of insulin on tissue glucose uptake from blood are insufficient[18,21]. Obesity and physical inactivity are established risk factors for insulin resistance and often precede its development[10,19,21]. Insulin resistance is characterised by several metabolic effects: insulin-mediated glucose uptake into skeletal muscle and adipose tissue is reduced, incomplete suppression of hepatic glucose production and incomplete suppression of adipose tissue lipolysis[10,18,21]. The ensuing hyperglycaemic state impairs β cell function through glucose toxicity, leading to impaired insulin secretion[18,19]. Additionally, due to the limited availability of intracellular glucose, fatty acids from increased adipose tissue lipolysis are metabolised, leading to higher blood plasma lipid levels[10]. Together with an inability of the β cells to compensate and maintain euglycemia, these metabolic disturbances further the onset of insulin resistance and together contribute to T2D pathogenesis. The compensatory capacity of β cells to counteract the effects of insulin resistance varies between individuals and is thought to have a substantial genetic component[22].

Risk factors influencing T2D incidence can be broadly categorised into modifiable and unmodifiable factors. Differences in T2D risk have been shown to vary with ethnicity[23–25],

previous family history of T2D[26], age[27], sex[27] and genetic predisposition all of which are unmodifiable. Despite this, a larger proportion of T2D risk factors are behavioural lifestyle or environmental factors which can be modulated via interventions to alter T2D risk. These include: being overweight or obese[28,29], high central adiposity[30], socioeconomic inequalities regarding treatment access[31], unbalanced diet[32,33], lack of physical activity[28,32], sedentary lifestyle[34,35] and smoking[32].

The role of cholesterol in the initiation of atherosclerosis in CHD pathogenesis

Atherosclerosis is the process underlying many cardiovascular diseases, including CHD, stroke and peripheral artery disease, and refers to the formation of fibrous, lipid-rich lesions embedded in the arterial wall[4]. LDL cholesterol particles, which are lipid-rich packets coated in phospholipids with apolipoprotein B running through their centre, are designed to transport cholesterol in blood[4]. LDL levels in excess of the required physiological range of between 0.5–0.8 mmol/L[36] have been causally linked with atherosclerosis in both observational[37] and genetic studies[38]. This chronic excess of LDL leads to the accumulation of LDL in the arterial intima, the innermost layer of the arterial wall[4]. The exact mechanism whereby arterial LDL accumulation leads to localised inflammation still remains to be elucidated[4]. However, this process leads to the recruitment of monocytes, immature circulating phagocytic cells of the innate immune system that later mature into macrophages, to the arterial wall by chemokines secreted by activated endothelial cells[4,39]. Monocytes then bind to adhesion molecules on the endothelial cell surface, facilitating their entry to the intima (**Figure 1.3**)[4,40].

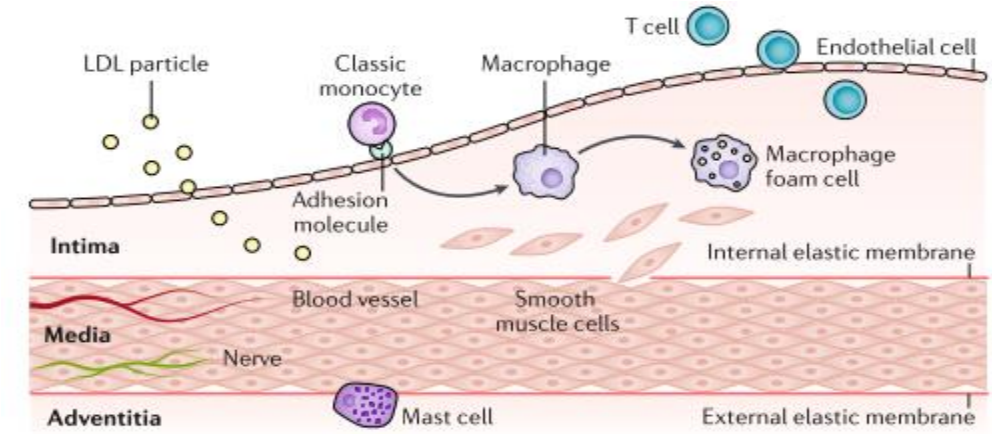


Figure 1.3: The initial stages of atherosclerosis. The layers of the arterial wall are shown, along with the immune cell recruitment as a result of LDL deposition. Figure taken from[4].

Arterial macrophages sequester LDL via scavenger receptors with the aim of maintaining lipid homeostasis to the point of overload, thus forming foam cells[41] a hallmark of atherosclerotic lesions[40]. Along with monocytes, T cells of the adaptive immune system and smooth muscle

cells are also recruited to the lesion via chemoattraction and play roles in the growth and maintenance of the lesion to eventually form a plaque[4].

The growth and maintenance of the plaque is characterised by a number of processes, chief among which are a thickening of the intima layer, the formation of a fibrous cap and the establishment of a necrotic core in the centre of the plaque (**Figure 1.4**)[4,40]. Smooth muscle cells secrete extracellular matrix molecules such as collagen and elastin which accelerate the accumulation of lipids, thickening of the intima and the formation of the fibrous cap[4,42]. Foam cells become trapped in the intima as their capacity to migrate is compromised leading to their eventual death, forming part of the necrotic core along with cholesterol crystals and other cellular material[4,40]. Over time, impaired clearance of dead smooth muscle cells and macrophages contributes to the formation and maintenance of the necrotic core. Finally, T cell secreted cytokines such as interferon-γ (IFNγ) can impair collagen formation by smooth muscle cells, affecting repair of the fibrous cap[4]. Along with this, macrophage secretion of matrix metalloproteinases (MMPs) that degrade collagen, weakens the cap further thus increasing the likelihood plaque rupture, often leading to acute myocardial infarction[40,43].

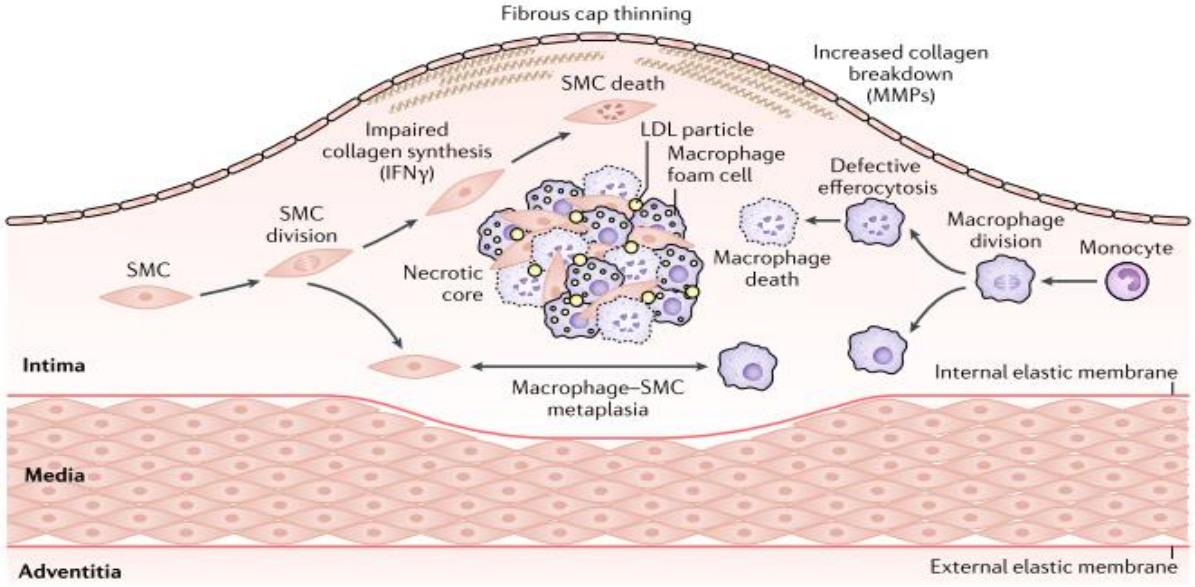


Figure 1.4: Progression of atherosclerotic lesions to plaques. Intima thickening, cellular apoptosis and the formation of the fibrous cap and necrotic core are shown. Figure taken from[4].

Much like T2D, risk factors for CHD can be sub-divided into modifiable and non-modifiable risk factors. The non-modifiable risk factors are similar to those of T2D, with the inclusion of T2D as an additional risk factor[15,44–48]. The modifiable lifestyle factors include: being overweight or obese[49], high central adiposity[30], smoking[50], lack of physical activity[51] but with the addition of: alcohol use[52], hypertension[53], prevalent T2D[15] and hyperlipidaemia[54].

1.2: Genetics as a framework for causal inference and risk factor prioritisation

Human genomic data are now being employed at scale for the purposes of drug discovery and validation efforts[55–58]. This is partly because only 10% of candidate therapies are estimated to proceed through development from Phase 1 trials to release[59]. Approximately 79% of drug candidates fail due to a lack of efficacy or safety concerns[59]. Evidence from human observational studies is useful for identifying putative drug targets, however, is liable to confounding and reverse causality[55]. Comparatively, genetic variants are robust to reverse causality[60] and non-genetic forms of confounding[60] while also segregating at random[61], analogous to a randomised-controlled trial (RCT)[62]. Genetics therefore has the potential to efficiently prioritise candidate targets but also lower drug attrition rates[57,58]. This is exemplified by successes such as *HMGCR*, a target of statins, which was prioritised on the basis of its association with cholesterol synthesis and therefore its inhibition confers cardio-protective benefits[63]. Retrospective studies have shown that the proportion of drug mechanisms supported by genetic evidence increases throughout the development process making targets twice as likely to be approved[57,64]. These estimates assume that the causal mechanism is clear[64] and are dependent on the strength of the genetic evidence, as targets associated with Mendelian disorders are five times more likely to reach approval[65].

The establishment of large population-based biobanks such as UK Biobank[66] facilitates the unbiased prioritisation and characterisation of putative drug targets at scale. Large biobanks have tripled the effective sample size for T2D GWAS studies, dramatically improving the statistical power of these analyses[22,67]. Recent technological advances in 'omics methods have facilitated the simultaneous measurement of thousands of molecular traits in large epidemiological cohorts. Combining this with lifestyle and anthropometric measures allows for greater characterisation of the molecular mechanisms underpinning complex disease[68]. The vast majority of therapeutics perturb protein function[58]. Therefore, integrating genomic information with gene expression and proteomic data enables the characterisation of the molecular consequences and mediators of this perturbation[67].

Due to the properties of genetic variants[60–62], they may be used in a Mendelian randomisation (MR) framework to empirically estimate the effects of molecular traits on disease risk[69–72]. MR methods are well established, efficient methods of prioritising evidence from observational studies, that treat genetic variants as instrumental variables (IVs)[69,71–73]. The assumptions surrounding the use of IVs in MR have been reviewed elsewhere[62], briefly, 1) the IV must be reliably associated with the exposure, 2) the IV must not be associated with confounder of the

association between the exposure and outcome, 3) there is no independent pathway from variants to outcome that is not mediated via the exposure[71]. Methodological advancements have facilitated multivariable analyses accounting for the effects of genetic variants on correlated traits[74,75] as well as methods accounting for the effects of pleiotropy[76] or outliers[77–80] in a model. An example of this is the causal association of LDL cholesterol with CHD risk, an established risk factor for CHD[81]. In univariable analyses, both LDL cholesterol and apolipoprotein B (ApoB) were associated with CHD risk[81]. However, in multivariable MR, only ApoB was significantly associated with CHD risk, suggesting that ApoB and not LDL is causally associated with CHD risk[81]. This work would not have been possible without the scale and availability of publically available genotypic data which facilitates the estimation of variant–exposure and variant–outcome associations from independent studies of unrelated individuals in a framework termed two-sample MR (2SMR)[71,79]. This approach has several advantages, firstly, 'winners' curse', which underestimates causal effects in one-sample MR, is unlikely in two-sample MR[82], secondly, weak IVs bias estimates towards the null[83].

Genetic variants may further support the drug development process by predicting any off-target or adverse effects, estimating the target specificity[56,58]. Returning to the example of *HMGCR*[84], while statins have been shown to lower LDL cholesterol levels, there appears to be an dose-dependent increase in T2D risk[85,86]. Using variants in the *HMGCR* gene, the associations with glycaemic, anthropometric and lipid traits were estimated, with the authors concluding that T2D risk was partly mediated via HMGCR inhibition[84]. Later studies demonstrated that lowering triglyceride levels via lipoprotein-lipase (*LPL*) may be used in conjunction with statins to further reduce CHD risk while negating any statin-mediated increases in T2D risk[5].

Pharmacomimetic genetic variants mimic the effects of established therapies and may therefore be used to uncover new treatment indications for existing drugs[67]. This strategy reduces the risk of failure as the safety profile of existing drugs is established, it accelerates the development process and finally reduces the overall cost of development[87]. The initial evidence that this may be possible is often gleaned through systematic phenome-wide scans estimating the association of the variant with diverse well-defined phenotypes[87]. Critically, the pathway of interest needs to be causal for both disease indications and the drug mechanism must be shared[87]. An example of particular relevance to this thesis, involves antagonism of the interleukin-6 receptor (IL6R) using a monoclonal antibody, tocilizumab, an established treatment for rheumatoid arthritis[88]. Observational evidence has suggested that IL-6 levels play a role in CHD risk[89,90]. Using variants in the *IL6R* gene in an MR framework, the potential efficacy of IL6R antagonism on CHD risk was investigated[88]. IL6R signalling was estimated to

be causal for CHD, suggesting that antagonism of IL6R signalling in CHD may confer protective effects[88].

There are several challenges and considerations of this work. Firstly, the translation of GWAS findings into meaningful biological insights with a clear mechanism of action, is a challenge related to the interpretation of any GWAS result. This is largely because the majority of GWAS findings lie in non-coding regions with comparatively small effect sizes[91]. Secondly, linkage disequilibrium (LD) between variants in the genome often makes discrimination of a causal variant underlying a GWAS association signal in a genomic region challenging[92]. Thirdly, pleiotropic effects whereby a variant may affect two or more unrelated traits via diverse pathways (horizontal pleiotropy) are challenging to use in an MR framework as they invalidate the assumption that there is no independent pathway from variants to outcome that is not mediated via the exposure[71]. Therefore, elucidating the mechanism of action of a pleiotropic variant on a particular trait is challenging, as the variant may influence the trait via many different mechanisms. Finally, often quantification of the target of interest is not possible, therefore, a proxy which is directly impacted by the target of interest may need to be quantified instead. In the case of HMGCR and LDL cholesterol, the quantification of HMGCR activity is challenging compared to LDL. Therefore, quantification of the reduction in LDL levels may be used as a proxy for HMGCR activity[84,93].

In summary, genetics is an efficient and robust framework to the types of confounding that affects observational studies[60–62]. Genetics has the potential to accelerate and improve the success rate of the drug development process as well as prioritise drug potential drug targets by empirically estimating their efficacy, safety and potential for use in other indications[56,58,67,69–72]. This capacity for large-scale genetic studies will only increase with the establishment of biobanks that integrate genetic data with high throughout ‘omics methods for deep molecular characterisation.

As eluded to in previous sections, much of the aetiology of cardiometabolic diseases is largely understood. However, there is still uncertainty in areas which are of considerable scientific, clinical and pharmaceutical interest. Two such examples, which are the topics of this thesis, are the role of inflammatory proteins and incretins in the aetiology of cardiometabolic diseases. This thesis will integrate large-scale genomic data with multiple ‘omics datasets to further our understanding of the molecular underpinnings of inflammatory proteins and incretins to estimate how levels of these traits relate to risk of T2D and CHD.

1.3: Inflammation as a link between obesity and cardiometabolic diseases

Inflammation: the body's natural defence to insult

Inflammation is an evolutionarily conserved, coordinated response by the immune system to disturbances in cellular or systemic physiology such as tissue injury or infection[94]. The immune system is composed of both the innate and adaptive immune responses. Innate immunity serves as a first-line, immediate and non-specific defence to immune insult[95]. In contrast to this, adaptive immunity is a delayed and specific response that centres on priming highly specialised T and B lymphocytes using antigens from the insult itself[95]. Innate immune cells such as phagocytic macrophages or neutrophils sense pathogen- or endogenous cell damage-associated molecular patterns (PAMPs or DAMPs) using pattern recognition receptors on their surfaces[96]. This stimulus leads to the secretion of soluble effector molecules called cytokines and chemokines which facilitate signalling between diverse cell types to coordinate and regulate the immune response on both a local and systemic level[96–98].

Cytokines can be divided into two broad groups, pro- or anti-inflammatory, depending on whether they contribute to the propagation or resolution of inflammation[96]. Under normal conditions, stimulation of the immune system leads to the production of proinflammatory cytokines and chemokines which recruit leukocytes and other immune cells to the insulted tissue (**Figure 1.6**)[98]. Once the pathogen or cellular debris has been removed, the inflammatory stimulus diminishes, and a period of resolution begins. The secretion of anti-inflammatory cytokines provides the signal for infiltrating immune cells to return to the circulation and allow for tissue repair[94,98]. Thereafter, a post-resolution phase is thought to occur where adaptive immunity is established in the insulted tissue through the infiltration of T and B lymphocytes. These are subsequently primed against future reoccurrence of the threat[98]. This delicate balance between pro- and anti-inflammatory processes allows for a temporally restricted, acute inflammatory response while balancing resultant tissue damage[98].

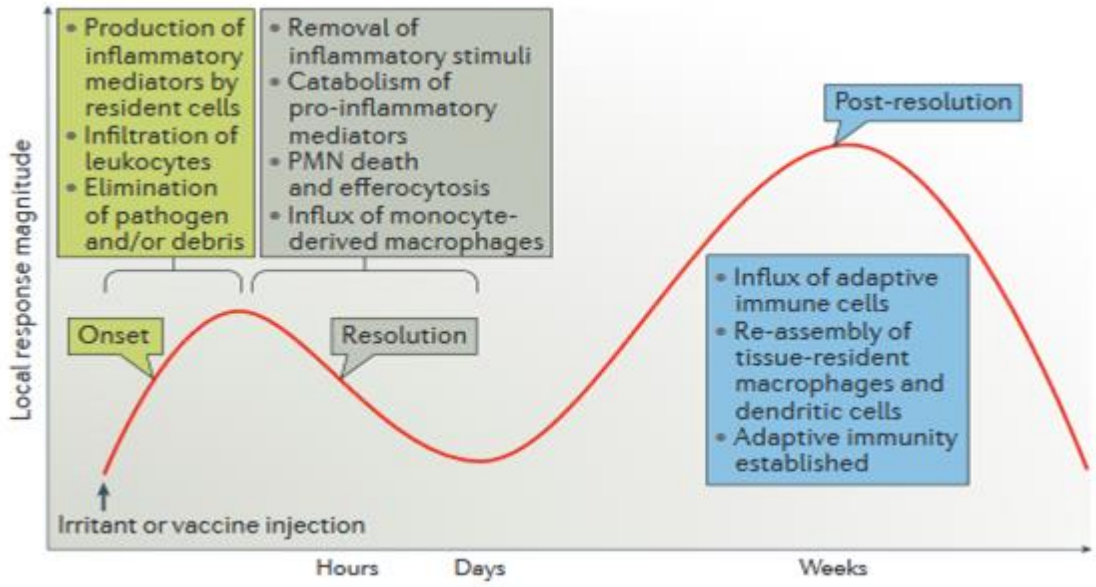


Figure 1.6: The healthy acute inflammatory response. This is characterised by a rapid initialisation phase to eliminate insults, followed by a resolution phase to enable tissue recovery. Finally, the persistent post-resolution phase is initiated which primes the adaptive immune system to future insults. Figure taken from[98].

In cardiometabolic diseases, resolution of the inflammatory response can be hindered, resulting in a state of chronic low-grade inflammation[94,98]. This is characterised by persistent activation of proinflammatory mediators such as cytokines, leading to significant tissue damage due to constant cytokine upregulation[94]. The triggers for this are unknown, however, tissue damage arising from obesity-related metabolic dysfunction is thought to disrupt immune homeostasis and play a role in its establishment[99]. Several correlated risk factors for chronic inflammation have been identified such as: diet-induced microbiome dysbiosis[100–103] and lifestyle factors such as chronic stress[104], insomnia[105,106] and physical inactivity[107,108], many of which are shared risk factors for obesity[94]. The following section will expand upon some of the inflammatory mechanisms thought to link obesity to cardiometabolic diseases.

Inflammatory mechanisms linking obesity to T2D and CHD

Historically, the sole function of adipose tissue was thought to be long-term energy storage, however, recently this notion has been challenged and it is now appreciated that it plays a crucial role in the regulation of metabolic homeostasis[109]. Several key immune and metabolic signalling pathways converge upon adipose tissue, facilitating cross-talk between the two systems to maintain homeostasis of both systems[110–113]. Two key features of adipose tissue allow it to perform this integrated role, firstly, adipose tissue sequesters blood lipids for storage as triglycerides[114]. Secondly, adipose tissue is capable of secreting adipokines, molecules which regulate systemic metabolism by relaying information about energy stores to

neuroendocrine systems and maintain the adipose cellular environment[115,116]. Adipose tissue is therefore a key regulator of metabolic homeostasis that is adaptable in response to dynamic metabolic stimuli.

A hallmark of obesity is adipose dysfunction, one of the consequences of which, is the onset of chronic inflammation[111,113]. Obesity induces systemic low-grade inflammation in many metabolically active organs such as the pancreas[117], liver[118] and skeletal muscle[119]. However, dysfunctional white adipose tissue is the most studied of these given its roles in energy storage and metabolic homeostasis[120]. Adipocytes respond to positive energy balance by expanding rapidly in both size and number to accommodate further lipid storage[121]. As a result, a mildly hypoxic environment develops, inducing stress signals that facilitate extracellular matrix remodelling and angiogenesis to restore adequate oxygen supply[122]. Adipose tissue expansion is constrained by pre-defined limits of expansion[121] which, when exceeded, stresses the adipose compartment leading to dysfunction and the propagation and maintenance of chronic inflammation[110–113,121,123–126]. Evidence from several different forms of partial lipodystrophy6,7 and functional studies of peripheral adipose storage compartments8-10 suggests that a primary inability to expand gluteofemoral or hip fat can also underpin subsequent risk of cardiometabolic diseases.

In obesity, adipose tissue expands beyond the capability to sustain adequate angiogenesis, leading to a state of chronic hypoxia[127,128] and therefore an unresolved stress signal, triggering inflammation[113,123,126,129]. Coupled to this, efficient adipose expansion requires the remodelling of the surrounding extracellular matrix to ensure proper angiogenesis and functionality of the tissue[113,126]. The rate at which adipocytes expand in obesity leads to dysfunctional extracellular matrix remodelling and fibrotic deposition of collagen, placing physical stress upon adipocytes[113,126,130,131]. Impeding further expansion of adipose tissue results in raised circulating lipid levels[132,133], a risk factor for atherosclerosis[38,134]. This, in turn, may result in lipid deposition within organs, namely ectopic deposition[121]. The combined effects of the above-mentioned adipose dysfunctions often result in adipocyte death and necrosis[135,136], further propagating the inflammatory response and recruiting phagocytic macrophages to the inflamed adipose tissue[137,138].

Over 25 years ago, the seminal finding that tumour necrosis factor alpha (TNFα) is over-expressed in obese murine adipose tissue established the first link between obesity and insulin resistance, through inflammation[139]. This finding was later replicated in humans[140,141]. This led to a proliferation in studies identifying other pro-inflammatory adipokines that are expressed in adipose tissue such as monocyte chemoattractant protein 1 (MCP-1)[142], interleukins 6[143,144]

and 8[145] (IL-6 and IL-8) and interferon-γ (IFNγ)[146,147], the expression of which becomes dysregulated in obesity[109]. Furthermore, higher TNFα and IL-6 levels have been associated with adipose lipolysis through the inhibition of lipoprotein lipase[111,148], leading to dysfunctional triglyceride storage and higher circulating fatty acids which may have implications for CHD risk. Two later studies revolutionised inflammation research in obesity, firstly, adipose-resident macrophages were discovered[137] which regulate adipose function and provide immune surveillance[149]. Secondly, macrophage infiltration into obese adipose tissue was demonstrated[138], dramatically increasing the proportion of macrophages to an estimated 41% of cells in obese adipose tissue compartments[137]. Furthermore, the predominate source of the observed TNFα overexpression in obese adipose tissue was attributed to these adipose-resident macrophages rather than adipocytes[137].

Based on this new evidence, focus shifted to further characterising the role of macrophages in dysfunctional obese adipose tissue with reference to how they propagate and maintain chronic inflammation. Macrophage infiltration was found to be driven, in part, by MCP-1 mediated chemotaxis[150], the levels of which are raised during obesity[151]. In addition to increasing in number, adipose macrophages alter their inflammatory state in obesity from an anti-inflammatory to a proinflammatory phenotype, resulting in higher pro-inflammatory cytokine secretion[152]. The trigger for this phenotypic switch is unclear, however, fatty acids as a result of metabolic dysregulation have been posited as a possible trigger[153], along with adipose dysfunctions mentioned previously. Related to this, macrophages have been observed to cluster around dying adipocytes in necrotic clusters called crown-like structures which are characterised by high levels of proinflammatory cytokine secretion[135]. Given this evidence, obesity is a likely driver for chronic inflammation, however, further research into the obesity-mediated inflammatory pathways in cardiometabolic diseases is needed.

Observational evidence establishing cytokines as therapeutic targets of interest

Cytokines have been posited as a possible link between obesity and cardiometabolic diseases due to their function as key signalling and effector molecules[109]. Observational evidence has established an association between obesity and higher proinflammatory cytokine levels, which are predictive of incident T2D[154] and CHD[155] independent of BMI and relevant risk factors. This is evidenced by strict weight loss interventions in obese individuals at risk for T2D which demonstrate a reduction in cytokine levels such as C-reactive protein (CRP) and others[156–158], during weight loss[159]. CRP is an acute-phase protein produced in the liver and routinely measured in clinical trials[160], upon which many inflammatory pathways converge, making it a sensitive marker of generalised inflammation[161].

In this study, the levels of four cytokines: IFNγ, IL-6, IL-8 and TNFα are of interest because of their putative roles in the aetiology of cardiometabolic diseases. IFNγ is secreted by T lymphocytes and adipocytes and is thought to function as a potent activator of macrophages, promoting the switch to a proinflammatory phenotype and subsequent secretion of IL-1 and IL-6[162,163]. IL-8 has been shown to recruit macrophages and neutrophils to atherosclerotic plaques through a process known as chemotaxis[164–166]. IL-6 and TNFα are two of the most studied cytokines in cardiometabolic diseases[94,109,148,154,167,168]. While both have been associated with insulin resistance in murine models[139,140,169–171], the translation of these findings to human systems remains uncertain. The finding that IL-6 levels are acutely raised by skeletal muscle secretion during exercise hampers interpretation of the role of IL-6 in insulin resistance, as skeletal muscle is the main source of insulin-mediated glucose disposal[172]. Furthermore, differences in cardiometabolic risk may exist with regards to acute vs. chronic exposure to raised cytokine levels, thus further complicating the interpretation of their role in disease pathophysiology.

Since the initial description of an association between TNFα and insulin resistance[139], the number of epidemiological studies investigating cytokines and risk of cardiometabolic diseases has increased. A recent cross sectional study of 1,558 multi-ethnic participants reported that TNFα levels were higher in participants with impaired glucose tolerance and T2D, and were correlated with insulin resistance metrics[173]. However, a smaller European study of 519 German adolescents failed to replicate this association, adding uncertainty to the literature[174]. In addition to TNFα, IL-8 and IL-6 levels and their association with fasting insulin and homeostatic model of insulin resistance (HOMA-IR) were investigated. Like TNFα, contradictory results have been shown for the association between IL-8 levels and fasting insulin or HOMA-IR[174,175]. In contrast, IL-6 levels were significantly associated with both after adjustment for age, sex and lipids[174]. However, both associations were attenuated after adjustment for BMI, suggesting that these effects may be mediated by obesity[174]. Despite BMI being a simple and widely available proxy for obesity, there have been questions surrounding its validity as a proxy for body fat distribution[30,176–179], thought to be one of the underlying causes for cardiometabolic diseases. Similar questions surrounding the validity of HOMA-IR as a proxy for insulin resistance have been raised[180]. This has implications for research using these measures as proxies for obesity and insulin resistance, respectively, as they may incompletely capture the phenotype of interest or fail to capture the truly causal aspect. This may lead to inaccurate inferences from results while also impacting the statistical adjustment for these traits in models.

In the largest prospective study of up to 7,683 European participants, IL-6 levels were associated with both HOMA-IR and fasting but not 2-hour insulin levels[181]. A doubling in IL-6 levels from baseline corresponded to a 2.2% increase in fasting insulin levels from baseline, identical to that of HOMA-IR[181]. The same was true for HOMA-β, a metric of pancreatic β-cell function, indicating raised β-cell function as a compensatory mechanism. These associations remained significant despite extensive adjustment for age, sex, BMI, lifestyle factors and lipid traits[181]. In contrast, while IFNγ levels have been proposed to play a role in insulin resistance on a cellular level[182], no epidemiological studies have examined this effect in a suitably powered cohort.

When testing the association between IL-8 levels and incident T2D, again no evidence for association was found, when BMI and extensive lifestyle and cardiometabolic risk factors were accounted for[183,184]. The largest systematic review and meta-analysis to date of TNFα levels and their association with incident T2D is comprised of five prospective studies amounting to 509 incident T2D cases and 4,716 controls[185]. The association between TNFα and incident T2D was shown to be non-significant, with substantial heterogeneity between studies. However, in a sensitivity analysis that removed low-quality studies, TNFα levels were shown to be significantly associated with incident T2D[185]. This highlights the uncertainty surrounding the association between TNFα levels and incident T2D and presents an opportunity for suitably powered studies to provide clarification.

The number of studies investigating the association between IL-6 levels and incident T2D are comparatively more abundant. Two systematic reviews have been performed on the subject[185,186], the first consisting of 10 prospective studies in up to 4,480 cases and 19,709 non-cases. This first study showed a 33% higher risk of incident T2D per 1 log pg/mL higher IL-6 levels[186]. This was replicated in the second meta-analysis of 16 prospective studies comprising 4,751 cases and 24,929 non-cases, which showed a 32% higher risk of incident T2D per 1 log mg/L higher IL-6 levels[185]. In sensitivity analyses, neither systematic review displayed evidence of publication bias or significantly different estimates within subgroups of included studies. Again, there was a lack of epidemiological studies investigating the association of IFNγ with incident T2D.

Two previous studies have investigated the association of IL-8 levels with incident CHD[187,188]. The first, a case-cohort study consisting of 381 European cases and 1,977 non-cases, showed 38% higher risk of incident CHD for individuals in the highest tertile of the IL-8 distribution[187]. This association held despite extensive adjustment for relevant cardiometabolic risk factors, however, was attenuated when adjusted for IL-6 and CRP levels[187]. The second, a prospective

Swedish cohort of 4,011 participants, replicated this finding using similar adjustment and also showed no significant association with either myocardial infarction or ischemic stroke[188]. Similar to T2D, two systematic reviews have examined associations between IL-6 levels (an established CHD risk factor) and incident CHD[90,189]. The first of these comprised 17 studies amounting to 5,730 cases and 19,038 non-cases, and showed a 26% higher CHD risk per 1 SD higher IL-6 levels[90]. The second and larger meta-analysis of 25 studies with a total of 7,982 CHD cases found a 25% higher CHD risk per 1 SD higher IL-6 levels[189]. This second meta-analysis also examined associations of TNFα with incident CHD in 7 studies, comprising 1,656 CHD cases and 4,346 non-cases[189]. They found a 17% higher CHD risk per 1 SD higher TNFα levels[189]. Finally, in a case-control study consisting of 931 CHD cases and 974 controls, IFNγ levels were found to be significantly associated with CHD risk, corresponding to a 46% higher risk in individuals in the top tertile of the IFNγ distribution[190]. However, upon adjustment for IL-6 levels this association was attenuated and no longer significant[190]. This finding is consistent with established IL-6 biology and CHD risk[39].

In conclusion, comparatively fewer studies have investigated IL-8 and IFNγ levels in CHD and T2D patients. Typically, these studies have also been conducted in smaller sample sizes compared to those studying IL-6 or TNFα. The consequence of using a few small studies as an evidence base translates to unclear association evidence between IL-8 and IFNγ and cardiometabolic diseases. Comparatively, association estimates for IL-6 and TNFα with CHD and T2D are more established. While this highlights the potential for larger, suitably powered studies to clarify the associations of IL-8 and IFNγ with cardiometabolic diseases, the causal evidence for each cytokine of interest in cardiometabolic diseases is also yet to be established.

Clinical trials targeting cytokines in cardiometabolic diseases

Several established therapeutic and lifestyle interventions used in clinical practice have demonstrated anti-inflammatory properties alongside their primary mechanism of action[9]. Interventions such as weight management diets[191] and bariatric surgery[192] have proven anti-inflammatory effects. Similarly, glucose-lowering therapies such as metformin[193–197] and incretin modulators[198–201] are associated with small reductions in circulating levels of inflammatory markers. This appears to be a beneficial secondary feature shared by many interventions that seek to lower blood glucose levels. Anti-inflammatory therapies commonly used in rheumatic disease like salicylate[202,203] or low-dose methotrexate[204,205] have broad anti-inflammatory effects and are effective at lowering glucose levels in cardiometabolic diseases. However, this is not always the case, as glucocorticoids have been shown to raise glucose levels thereby promoting insulin resistance, despite their efficacy against inflammation[9].

As previously discussed, TNFα was the first cytokine thought to participate in insulin resistance[139]. Despite this, relatively few clinical trials using TNFα antagonists have been conducted. This is surprising given that several TNFα antagonists have been developed and are routinely used to treat conditions such as rheumatoid arthritis[206] and psoriasis[207]. Observational studies using these antagonists have shown improved glycaemia and a concurrent protective effect against incident T2D[206,207]. However, their results need to be treated with caution as they were not randomised controlled trials. Trials of TNFα antagonists in patients with cardiometabolic diseases have enrolled small patient numbers and have often involved short treatment exposures[208–210]. Their results indicate that whilst lower TNFα-mediated inflammation was observed, no significant improvements in glycaemia or insulin sensitivity were found[208–210]. This apparent lack of effect may be due to statistical power issues or short trial durations. The largest trial to-date enrolled 40 obese participants without diabetes, randomised to either placebo or 50mg doses of Etanercept™, a TNF receptor antagonist, for 6 months[211]. Contrary to observational results, Etanercept™ treatment led to 10.8% lower fasting glucose levels and 22.1% higher adiponectin levels, when compared to placebo[211]. However, no significant differences in fasting insulin levels were observed[211]. These results lend support to continued investigation using larger sample sizes with sufficient trial duration to assess the efficacy of TNFα antagonism in patients with cardiometabolic diseases.

To-date, no clinical trials of IFNγ antagonists in cardiometabolic diseases have been conducted. Comparatively, clinical trials for IL-6[212] and IL-8[213] antagonists have been initiated but only the IL-6 trial has been completed. Sarilumab™, an IL-6Rα antagonist previously trialled for use in rheumatoid arthritis[214], was tested in combination with disease-modifying anti-rheumatic drugs (DMARDs) in rheumatoid arthritis patients with and without T2D[212]. A total of 179 T2D patients were compared to 1,803 non-T2D patients from 2 randomised, placebo-controlled studies[214,215]. A greater reduction in fasting glucose was observed in the Sarilumab™-treated T2D group after 24 weeks compared to the non-T2D group. Similarly, HbA1c levels were lower in both the Sarilumab™ treated groups but not in the placebo. This effect was largest in the T2D group at week 24 where HbA1c was 0.43% lower compared to baseline levels[212]. These results are encouraging and support future trials of Sarilumab™ in T2D patients without rheumatic disease and not in combination with other therapies.

While relatively few trials have focussed on the cytokines of interest, considerably more effort has been invested in IL-1ß antagonism trials. IL-1 acts as a master cytokine, capable of inducing the expression of diverse cytokines[216]. There are 2 forms of IL-1, namely IL-1α and IL-1ß, the latter requires processing by the NLRP3 inflammasome for maturation to its active form[217]. IL-1ß signals through the type 1 IL-1 receptor (IL-1R1) to activate NF-κB, a central

transcription factor controlling IL-6, IL-8 and TNFα expression, among others[217]. Given IL-1ß's role as an upstream controller of cytokines that have been implicated in cardiometabolic diseases, considerable research on IL-1ß antagonism has been conducted. In an initial trial, 70 T2D patients were randomised to receive daily doses of Anakinra™, an IL-1R antagonist, or placebo for 13 weeks[218]. At 13 weeks, the Anakinra™ group had significantly lower HbA1c, IL-6 and CRP levels and displayed improved ß-cell secretory function, assessed using the ratio of proinsulin to insulin, compared to placebo[218]. Participants were then followed for a further 39-weeks following Anakinra™ discontinuation[219]. Improved insulin secretion persisted, as did lowered inflammation, suggesting that this may be a possible therapeutic for T2D[219]. This proof-of-concept study paved the way for the larger Canakinumab™ anti-inflammatory thrombosis outcomes study (CANTOS) trial consisting of 10,061 patients with previous myocardial infarction and CRP levels of more than 2mg/L[220]. Three doses of Canakinumab™ were compared to placebo: 50mg, 150mg and 300mg, each administered every 3 months[220]. The pre-specified primary endpoints of the trial were non-fatal myocardial infarction, non-fatal stroke or cardiovascular death[220]. Reductions in CRP level from baseline were observed in all Canakinumab™ groups, with the greatest reduction of 41% in the 300mg group[220]. The 150mg group had a 15% lower hazard for the primary cardiovascular endpoints compared to placebo[220]. This effect was also observed in the 300mg group, which had a 14% lower hazard for the primary cardiovascular endpoints compared to placebo, whereas the 50mg group showed a 7% lower hazard[220]. However, only the 150mg dose met the prespecified multiple testing threshold for both the primary and secondary end points[220]. A secondary pre-specified analysis evaluated the efficacy of Canakinumab™ on incident T2D risk among participants with pre-diabetes at baseline[221]. Additionally, the effects of Canakinumab™ on fasting glucose and HbA1c levels in patients with and without prevalent T2D were also assessed[221]. However, Canakinumab™ did not significantly reduce the risk of incident diabetes in participants without diabetes at recruitment or improve glycaemia in participants with prevalent diabetes[221]. In summary, IL-1ß antagonism substantially reduced circulating inflammatory markers and proved effective in lowering the risk of recurrent cardiovascular events, independent of lowering lipid levels[220]. However, did not reduce the risk of new-onset diabetes in participants without diabetes at recruitment, or improve glycaemia in participants with diabetes[221].

There are several considerations regarding the suitability of the CANTOS trial for the evaluation Canakinumab™ to reduce the risk of incident T2D in the general population. Firstly, participants with previous myocardial infarction and high CRP levels were recruited, limiting the generalisability of the results. The trial was not designed to estimate the effect of Canakinumab™ in patients with established T2D, therefore, the effects of Canakinumab™ on

glucose control may not be representative[221]. Secondly, the trial noted a violation of the proportional hazards assumption, observing a reduction in the risk of incident T2D in the early phases of the trial while later estimates showed an increased risk of incident T2D[221]. Similarly, HbA1c was reduced in the first 6-9 months of the trial but was attenuated over time so that the overall effect was non-significant. The authors suggest that this may have been because lifestyle interventions and alterations in other anti-diabetic therapies were allowed during the course of the trial, which may have affected glycaemic control[221].

Inflammatory signalling is complex and involves considerable pathway crosstalk between constituent proteins to be effective. Therefore, it is unclear whether antagonism of single pathways, or combinations thereof, will have the best therapeutic potential. In addition, inter-personal differences in inflammatory pathway may impact therapeutic efficacy[9]. The concern with immune antagonism is the potential for immunosuppression. However, considering that immunomodulation therapy is established for autoimmune and rheumatic disease, this may represent a possible area for drug repurposing. Finally, possible immunosuppression or adverse effects of immunomodulation can be estimated with minimal confounding using genetic data. As genetic effects represent life-long exposures, the effects thereof can be re-scaled to represent the effect over short treatment durations, mimicking the effects of a trial[222].

In summary, inflammation is thought to be a potential mediator of obesity-related risk and cardiometabolic diseases, with a role in insulin resistance being suggested[9,109,112,123,223]. Evidence from small observational studies[90,154,155,168,175,181,185,186,188–190,224,225] and clinical trials[218,219] has suggested that lowering inflammation may confer protective benefits on glycaemic control, however, results vary. In addition, recent evidence from the CANTOS trial[221,226] has challenged the findings of a protective effect in T2D. In light of this, several questions remain to establish their relevance in cardiometabolic diseases:

1. The genetic determinants for the majority of inflammatory proteins are unknown, GWAS studies of proinflammatory cytokine levels have only recently been conducted in modest sample sizes[227].
2. It is unclear whether inflammation is causal for cardiometabolic diseases or whether it is a consequence of cardiometabolic diseases or their related risk factors.
3. It is unclear which inflammatory proteins and biological pathways may be involved in the aetiology of T2D and CHD.

Overview of the genetic architecture of IFNγ, IL-6, IL-8 and TNFα levels in population-based studies

Introduction

This systematic search surveys the current genetic epidemiology knowledge of published GWAS studies of the four cytokines of interest in population-based studies.

Methods

To assess the literature surrounding the association of genetic variants with levels of the pro-inflammatory cytokines (IFNγ, IL-6, IL-8 or TNFα), I performed a systematic literature search using the PubMed (MEDLINE) and GWAS catalogue databases. My search strategy focused on previous GWAS studies that have investigated the association of levels of at least one of the four cytokines of interest in population-based cohorts. I searched the databases for all clinical trials, journal articles and meta-analyses published in English between June 2007 (the year the WTCCC GWAS[228] was published, considered as the inception of large-scale GWAS analyses) and 21 April 2020 using human participants only. The search strategy used is detailed in **Table S1.1** and led to the inclusion of 8 full text articles in the review after using the exclusion criteria detailed in **Figure 1.7**. Briefly, studies were screened for their eligibility and were removed if the study was not a GWAS or if cytokine levels were measured in diseased patients where cytokine levels may be elevated due to disease or infection.

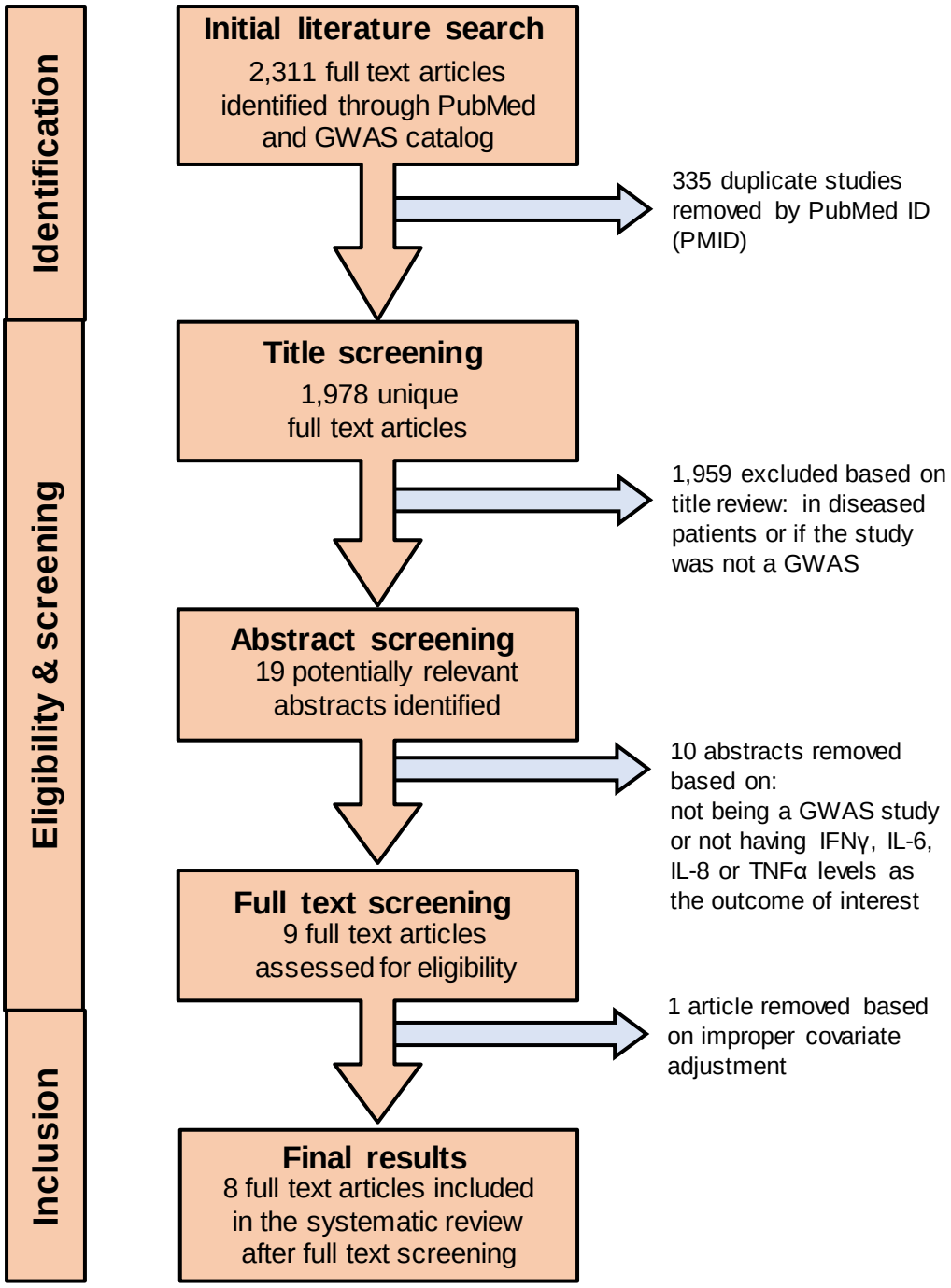


Figure 1.7: Schematic of the systematic literature review detailing the exclusion criteria and the number of articles excluded at each stage.
Abbreviations: IFNγ, interferon-γ; IL-6, interleukin-6; IL-8, interleukin-8; TNFα, tumour necrosis factor-α.

Table 1.2: Study characteristics of the GWAS studies estimating the association of genetic variants with IFNγ, IL-6, IL-8 and TNFα levels							
Study, year (reference)	**Population**	**Proportion of women (%)**	**Study Design**	**Cytokine Quantification Assay**	**Genotyping Chip**	**Covariate Adjustment**	**Transformation of Cytokine Measures**
Sliz *et al.*, 2019[229]	13,577 Finnish individuals: 5,284 from NFBC1966 + Ahola-Olli 2017	52	Prospective	Bio-Plex 200 system using fasted plasma samples	Illumina 670k HumanHap array	Age, sex, BMI, 10 PCs	Inverse-normal transformed residuals
Sun *et al*., 2018[230]	3,301 European blood donors from the INTERVAL study	49	Prospective	SOMAscan platform using EDTA plasma samples[a]	SOMAscan Assay V3.2 proteomics	Age, sex, duration between blood draw and processing, 3 PCs	log10-transformation
Ahola-Olli *et al.,*2017[227]	8,293 Finnish individuals from 3 cohorts: The Cardiovascular Risk in Young Finns Study (YFS), FINRISK1997 and FINRISK2002	Not reported	Prospective	Bio-Plex Pro Human Cytokine Assay using fasted samples of EDTA Plasma (FINRISK1997), heparin plasma (FINRISK2002), serum (YFS)	Not reported	Age, sex, BMI, 10 PCs	Inverse-normal transformed residuals
Suhre *et al*. 2017[231]	1000 German individuals from the KORA study + replication in 338 Arab and South Asian individuals from the QMDiab study	51	Prospective	SOMAscan platform using EDTA plasma samples[a]	SOMAscan Assay V3.2 proteomics	KORA: Age, sex, BMI, QMDiab: Age, sex, BMI, Diabetes state, 3 PCs	Inverse-normal transformation
Sidore *et al.*2015[232]	Genotyped 6,921 Sardinians, of which, 2,120 were whole genome sequenced	53	Prospective	ELISA using fasted serum samples	Illumina HiSeq 2000, Illumina OmniExpress, Illumina Cardio-Metabochip, Illumina Immunochip, Illumin ExomeChip	Age, sex, Age^2, BMI, smoking status	Quantile transformation

Ayele *et al.*2012[233]	707 non-diabetic African Americans in the Howard University Family Study	57	Prospective	ELISA using fasted plasma samples	Affymetrix Genome-Wide Human variant Array 6.0	Age, sex, BMI, 2 PCs	log_{10}-transformation
Naitza *et al*. 2012[234]	4,694 Sardinians and replicated in 1,392 Sardinians	57	Prospective	ELISA using fasted serum samples	Affymetrix 500K Mapping Array, Affymetrix 10K Mapping Array, Affymetrix 6.0 Array, Illumina Metabochip, Illumina ImmunoChip	Age, sex, Age^2, BMI, smoking status	Quantile transformation
Melzer *et al*. 2008[235]	1200 European individuals from the InCHIANTI study	55	Prospective	ELISA using fasted serum samples	Illumina Infinium HumanHap550 v1 + v3	Age, sex	Ladder of powers, log10-, box-cox probit transformation

Results and Discussion

Throughout the GWAS era many studies have demonstrated that the associations of common genetic variants with common disease often have relatively small effect sizes, providing further support for the polygenicity of common disease[228,235]. However, the converse is also true as lower frequency variants tend to have larger effects on the phenotype[236]. This applies to protein traits too, however, often a single variant may have a disproportionately large effect on protein levels. This is often a *cis*-acting variant in the gene encoding the protein or a *trans*-acting variant in the receptor of the respective protein.

The initial systematic search identified a total of 2,311 articles which were narrowed down to 8 full text articles (**Table 1.2)** after applying the exclusion criteria. The results of these studies identified variants that are robustly associated with circulating levels of IL-6 and TNFα. Two studies reported associations with TNFα levels and six reported associations with IL-6 levels (**Table 1.3**). Comparatively, none of the included studies found any significant associations with either IL-8 or IFNγ levels. Two studies[227,229] examined the association of genetic variants with levels of IL-8 or IFNγ, however, no significant associations were found, which may be attributable to a lack of statistical power.

Table 1.3: Summary of the GWAS results for the four cytokines of interest

Cytokine	Study (ref)	Variant	Nearest Gene	LD	EA	EAF	Beta	P-value	N, Population
IL-6	[233]	6:133397598	*RP11-314E23.1*	-	T	0.01	-0.52[a]	$8.6x10^{-9}$	707, African Americans
IL-6	[234]	rs657152	*ABO*	1	T	0.27	0.22[b]	$2.1x10^{-29}$	4,694, Sardinians
IL-6	[232]	rs643434	*ABO*		A	0.26	0.22[b]	$5.8x10^{-27}$	6,921, Sardinians
IL-6	[234]	rs4129267	*IL6R*	0.99	T	0.26	0.11[b]	$2.4x10^{-8}$	4,694, Sardinians
IL-6	[232]	rs12133641	*IL6R*		G	0.26	0.12[b]	$6.9x10^{-9}$	6,921, Sardinians
IL-6	[235]	rs4129267	*IL6R*		T	0.37	0.69[c]	$1.2x10^{-57}$	1,200, Europeans
IL-6	[231]	rs4129267	*IL6R*		T	0.38	1.2[d]	$1.6x10^{-265}$	1,000, Europeans
IL-6	[230]	rs4129267	*IL6R*		T	0.40	1.11[d]	$7.4x10^{-1101}$	3,301, Europeans
TNFα	[229]	rs17074575	*DLEU1*	-	G	0.002	2.13[d]	$2.7x10^{-9}$	13,577 Europeans
TNFα	[235]	rs505922	*ABO*	0.16	N/A[e]	N/A[e]	N/A[e]	$6.8x10^{-40}$	1,200, Europeans
TNFα	[235]	rs8176746	*ABO*		N/A[e]	N/A[e]	N/A[e]	$2.0x10^{-14}$	1,200, Europeans

Abbreviations: IL-6, Interleukin-6; TNFα, Tumour necrosis factor-α; SD, standard deviation; LD, Linkage disequilibrium; EA, Effect allele; EAF, Effect allele frequency; N, Number of participants.

a. SD units of log_{10} transformed cytokine
b. SD units of quantile-transformed cytokine
c. SD units of ln-transformed cytokine
d. SD units of inverse-normal transformed cytokine
e. Estimates were not reported in the main text or supplementary information

Three loci have been found to be associated with IL-6 levels in population-based cohorts. Using a Sardinian founder population of 4,694 participants, rs657152 at the *ABO* locus was associated with lower IL-6 levels per G-allele[234]. This finding was later replicated in the same population using an additional 1,392 participants[234] and by an independent, larger study using 6,921 participants 2,120 of whom were whole genome sequenced[232]. An association signal for the T-allele of a non-coding variant outside of the *HLA* gene-region (chr6:133397598) at *RP11-314E23.1* with IL-6 levels was found[233]. This was shown to lower IL-6 levels by 0.52 standard deviation units of log_{10}-transformed IL-6[233]. However, the sample size of 707 non-diabetic African American participants is small thus replication of this result in a larger sample size is required. Several studies have identified association signals of genetic variants at the *IL6R* locus with higher IL-6 levels[230–232,234,235]. Two studies[230,231] used the aptamer-based SOMAscan assay to quantify cytokine levels. Protein quantification is normalised to an internal standard, thus cytokine levels are relative abundances and poorly correlated with gold standard ELISA-based methods. However, studies using these assays do identify and replicate relevant hits. All variants found to be associated with IL-6 levels are in high LD (R^2 > 0.99) with a functional missense variant, rs2228145, in *IL6R* which mimics IL6R antagonism. Large-scale meta-analyses have shown that rs2228145 leads to 34.3% higher s-IL6R levels and 14.6% higher IL-6 levels[160]. The associated variants at the *IL6R* locus have also been found to be *cis*-acting eQTLs and pQTLs for IL-6 levels in large cohorts[230,231,235].

The largest available meta-analysis to-date of TNFα levels consisted of up to 13,577 Europeans[229]. This study showed that a low frequency intergenic variant, rs17074575, nearby the deleted in lymphocytic leukaemia 1 (*DLEU1*) gene was associated with 2 SD higher TNFα levels[229]. Despite this considerably large effect, not much is known about how this region leads to higher TNFα levels. In addition to this, two independent *trans*-association signals, rs505922 and rs8176746, in the pleiotropic *ABO* locus encoding the Histo-blood group ABO system transferase were found to be associated with TNFα levels in 1,200 European participants[235].

These results highlight the abundance of studies investigating the genetic determinants of IL-6 and TNFα levels compared to the very few studies investigating IL-8 and IFNγ. Overall very few loci associated with cytokine levels were identified, this highlights an opportunity for larger, more powered studies to expand on the current knowledge. These results highlight the effects of established pleiotropic loci such as *ABO* on cytokine levels as well as the lack of specificity for the majority of results, aside from the association of *IL6R* with IL-6 levels. Critically, the lack of association between variants and levels of IL-8 and IFNγ highlights a gap in our knowledge that requires further investigation in large studies.

Chapter 2: Observational Analyses of Cytokine Levels with Cardiometabolic Diseases

Contributions and collaborations

Measurement of the cytokines and other blood-based biomarkers were performed by the laboratory team at the MRC Epidemiology Unit under the supervision of Deborah Lucarelli and Vasileios Kaimakis. Anthropometric and regional adiposity measurements were performed by members of the Fenland and EPIC-Norfolk study teams under the supervision of Emmanuela De Lucia Rolfe. The data related to these measures were collated and maintained by the data management team at the MRC Epidemiology Unit under the supervision of Adam Dickinson. Stephen J. Sharp provided statistical support for the setup of the Cox regression analyses. I designed the study, analysed the data and wrote the first draft of the report which was later reviewed by Luca Lotta and Claudia Langenberg.

Abstract

Background: Previous studies have helped to establish the observational associations of IL-6 and TNFα levels with cardiometabolic diseases and related risk factors, however, there is a significant gap in the knowledge surrounding the association of IFNγ and IL-8 levels with cardiometabolic traits.

Aims: This chapter aims to characterise the observational associations between the levels of four cytokines IFNγ, IL-6, IL-8 and TNFα, and cardiometabolic traits measured in two large prospective population-based studies, EPIC-Norfolk and Fenland.

Methods: Observational associations between levels of each cytokine and anthropometric, glycemic, lipid and regional adiposity traits from DEXA were estimated using multivariable linear regression adjusting for age and sex in up to 10,335 participants from the Fenland cohort. Next, the association between levels of each respective cytokine and incident T2D and CHD were estimated using multivariable cox regression adjusting for disease-specific covariates in up to 7,514 participants from the EPIC-Norfolk cohort.

Findings: Levels of IFNγ, IL-6 and TNFα showed similar patterns of association with all measures of regional and overall adiposity, insulin and most lipid measures in the Fenland cohort, consistent with an adverse metabolic profile. However, adjustment for BMI attenuated the association estimates for the majority of traits, except for associations with lower HDL, higher triglycerides, waist-to-hip ratio (WHR) and fasting insulin levels. Comparatively, IL-8 levels were associated with only two outcomes, namely higher WHR and HbA1c, which remained associated after BMI adjustment. In EPIC-Norfolk, only IL-6 and TNFα levels were significantly associated with incident T2D after extensive adjustment for cardiometabolic risk factors (HR per SD higher cytokine levels of 1.19; 95% CI: 1.07, 1.32; P = 0.002 and 1.16; 95% CI: 1.04, 1.30; P = 0.008, respectively). The same was true for incident CHD (HR per SD higher cytokine levels of 1.16; 95% CI: 1.09, 1.24; $P = 3x10^{-6}$ and 1.16; 95% CI: 1.09, 1.24; $P = 4x10^{-6}$, respectively).

Conclusion: This study demonstrates that IL-6 and TNFα levels are observationally associated with incident T2D and CHD and a wide variety of cardiometabolic risk factors but that associations with cardiometabolic traits are likely to be mediated by higher BMI. The evidence supporting the involvement of IFNγ and particularly IL-8 levels in cardiometabolic diseases is less convincing.

2.1: Introduction

Obesity is hypothesised to be a major driver of chronic inflammation in cardiometabolic diseases, with obesity-mediated insulin resistance also thought to play a role[94,111–113,118,119,123,125,126,138]. Due to their integral role in innate immunity, cytokine levels are subject to complex regulatory networks and feedback loops[237]. In addition, regulatory networks governing the levels of multiple cytokines often overlap and interdependent[237].

The observational association of cytokine levels with cardiometabolic diseases and related risk factors has been investigated in previous studies, often with inconsistent results (reviewed in **Chapter 1**). As outlined in **Chapter 1**, the levels of four pro-inflammatory cytokines are of interest to this study, namely IFNγ, IL-6, IL-8 and TNFα. Much of the previous work in this area has been conducted using comparatively small sample sizes[155,174,175,183,186,225,238–240]. While previous observational studies have examined the associations of IL-6 and TNFα levels with cardiometabolic traits[90,155,168,175,181,183,185–187,189,190,238,239], there is a significant gap in the knowledge surrounding the relevance of IFNγ and IL-8 levels in cardiometabolic diseases.

In this study, I sought to investigate the observational association of IFNγ, IL-6, IL-8 and TNFα levels with cardiometabolic diseases and related risk factors using two large population-based cohorts, Fenland and EPIC-Norfolk.

2.2: Methods

Study participants

EPIC-Norfolk[241], a constituent cohort of the European Prospective Investigation of Cancer (EPIC)[242] study, is a prospective cohort of over 20,000 individuals aged between 40-79 years, living in Norfolk (a County of the UK) at time of baseline recruitment (1993-1997). Participants were recruited from general practice surgeries in the city of Norwich and surrounding areas[241]. A total of 77,630 participants were invited to take part in the study, 25,639 of whom attended a baseline visit[243]. Participants attended a baseline visit upon recruitment, during which standard anthropometric and dual energy X-ray absorptiometry (DEXA) measurements were taken, environmental and lifestyle factors were assessed, and blood sampling was done. Three follow-up consultations for monitoring and further sample collection were done: three, thirteen and twenty years after baseline assessment. A total of 15,786 participants attended the second follow-up visit[243]. Participants who attended a visit between 1998 and 2000 and had blood samples available for cytokine measurement were included. The date of cytokine measurement is taken as baseline for the purposes of this study. The study was approved by

the Norfolk Research Ethics Committee (ref. 05/Q0101/191) and all participants gave their written consent before entering the study.

The Fenland study[244] is a population-based study of 12,435 individuals without diabetes who were born between the years of 1950 and 1975 and recruited through general practice surgeries in Cambridge, Ely and Wisbech (in the Cambridgeshire County in the UK). An initial baseline visit in 2005-2015 for the purposes of metabolic phenotyping including DEXA measurements and DNA collection was done. Follow-up visits commenced in 2015 and are ongoing. Participants who attended the baseline visit and had available blood samples for cytokine measurement were included in this study. Ethical approval for the study was given by the Cambridge Local Ethics committee (ref. 04/Q0108/19) and all participants gave their written consent prior to entering the study.

Cytokine measurement and phenotype generation

Cytokine levels were measured in serum samples from fasted participants in Fenland[244] and participants fasted *ad libitum* in EPIC-Norfolk[241] using the Meso Scale Discovery® Human Pro-inflammatory Panel 1 multiplex kit (Rockville, MD, USA). Four pro-inflammatory cytokines (IFNγ, IL-6, IL-8 and TNFα) were measured using a 4-plex sandwich immunoassay. Assays were provided in 96-well plates where each well was coated with cytokine-specific capture antibodies. Upon the addition of sample and buffer to the well, cytokines were conjugated to the capture antibodies on the well surface. Detection antibodies were added to the reaction binding to the analyte of interest which have coupled electrochemiluminescent labels (MSD SULFO-TAG™). Applying voltage to the plate electrodes caused the labels to emit light, the intensity of which provided a quantitative measure of the cytokine level as measured by the MSD Sector Imager 6000 instrument. To ensure result consistency, standards of known concentration were included on each plate along with 3 intra-plate duplicates of the samples to be measured. A pooled sample of serum from the study population was included on each plate as was a sample of stock, normal human serum (Millipore (UK) Ltd., UK).

Quality control criteria were applied on a plate by plate basis to ensure measurement consistency. The following criteria were applied: the calculated concentration of standards, pooled and QC samples should fall within ±20% of the mean. Light intensity signals of the standards were checked for consistency with what was expected with coefficient of variation (CV) metrics of less than 20%. Calculated concentration CV of standards, QC and samples of interest should be less than 20%. Distribution parameters of plates should be within 30% of the overall value for at least two of the 25th, 50th and 75th percentiles. Plates that failed any of these criteria were not included in the final data. A standard curve was fitted using the

standards of known concentration with detection limits at 2.5 SD from the blank standard and 0 SD from the highest standard as bounds.

Any missing data points where measurement was not attempted or failed because of technical failure were set to missing. In cases where measurements were below the assay detection limit, data points were interpolated to random values between 0 and the lower limit of detection. Following this, two transformations were created: log transformed and inverse-rank normal transformed after the log-transformation.

Anthropometric measurement and quantification of blood biochemical markers

Fasting glucose, fasting insulin and HbA1c in Fenland[244] were the values of circulating glucose (in mmol/L), insulin (log-transformed and expressed in log-pmol/L) and glycated haemoglobin (in %) measured in whole blood after overnight fasting. Two-hour glucose was the value of glucose (in mmol/L) measured in plasma two-hours after a 75-gram oral glucose challenge. Glucose was measured using the Dimension RxL Integrated Chemistry System (Siemens, Germany). In Fenland[244] total cholesterol, high-density lipoprotein (HDL) cholesterol and triglyceride concentrations (in mmol/L) were measured using enzymatic assays (Siemens Healthcare). Low-density lipoprotein (LDL) cholesterol values were calculated using the Friedewald formula[245]. Weight was measured to the nearest 200 grams using a calibrated electronic scale (TANITA model BC-418 MA; Tanita, Tokyo, Japan). Height was measured to the nearest 0.1 cm with a wall-mounted stadiometer (SECA 240; Seca, Birmingham, United Kingdom). BMI (in kg/m^2) was calculated as weight divided by height squared. Waist and hip circumferences were measured to the nearest 0.1 cm with a non-stretchable fibre-glass insertion tape (D loop tape; Chasmors Ltd, London, United Kingdom). Waist-to-hip ratio (WHR) was the ratio between the waist and hip circumferences. Average physical activity per week (in kJ/kg/day) was measured over seven days using a calibrated heart rate and movement sensor (Actiheart, CamNtech, Cambridge, UK). Measures for any participant that did not wear the sensor for more than 48 hours within the seven-day period were removed from the analysis.

Overall and regional body fat mass in Fenland was quantified by dual-energy X-ray absorptiometry (DEXA) using a Lunar Prodigy advanced fan beam scanner (GE Healthcare, Bedford, UK), encore v14.10.022 and CoreScan® software (GE Healthcare, Bedford UK) a whole-body, low-intensity X-ray scan that precisely quantifies fat mass in different body regions. All body images were manually processed, and demarcations corrected following a standardized protocol. The trunk included neck, chest, abdominal and pelvic areas. The legs were defined as the region below the lower borders of the trunk. The abdomen was defined

as the portion of the trunk between ribs and pelvis, defined by the outline of the iliac crest. The gluteofemoral region included hips and upper thighs. Visceral abdominal fat mass was estimated using subcutaneous fat width and the anteroposterior thickness of the abdominal wall. These measurements were used to extrapolate abdominal subcutaneous fat mass. Visceral abdominal fat mass was calculated by subtracting the subcutaneous abdominal fat mass from the total abdominal fat mass.

In EPIC-Norfolk[241] total cholesterol and high-density lipoprotein (HDL) cholesterol concentrations (in mmol/L) were measured with the RA 1000 Tecnicon analyser (Bayer Diagnostics, Basingstoke, UK)[246]. Low-density lipoprotein (LDL) cholesterol values were calculated using the Friedewald formula[246].

Correlations between cytokine levels and cardiometabolic risk factors

Pearson's correlations were estimated between levels of the four cytokines (transformed as described above) and anthropometric, glycemic, lipid and regional adiposity traits from DEXA. Inverse-rank normal transformations were used for all variables to facilitate comparison across the four cytokines, to mitigate the effect of outliers and to minimise the effect of skewed variables.

Associations between cytokine levels and cardiometabolic risk factors

The associations between circulating levels of the four cytokines with various cardiometabolic risk-factors were estimated in the Fenland cohort[244]. Any participants with missing DEXA measures were excluded and participants where measurements for visceral fat mass were unable to be interpreted by the iDEXA software[247] were interpolated between 0 and 1 gram. The following overall adiposity and DEXA traits were considered: WHR, BMI, body fat percentage, android fat mass, gynoid fat mass, leg fat mass, android to gynoid fat mass ratio, appendicular lean mass, peripheral fat mass, visceral adipose tissue and subcutaneous adipose tissue. The following glycaemic traits were considered: fasting glucose, 2-hour glucose and HbA1c and fasting insulin. The following fasted lipid measures were considered: high-density lipoprotein cholesterol, low-density lipoprotein cholesterol and total triglycerides. Finally, a measure of physical activity energy expenditure in kJ/kg/day was used. An inverse-rank normal transformation was then applied to the outcome variables to enable comparison across the four cytokines, to mitigate the effect of outliers and to minimise the effect of skewed variables. Insulin measures were log-transformed prior to inverse-rank normal transformation. The associations between adiposity, glycaemic, lipid and physical activity traits (outcomes) and cytokine levels (exposure; transformed as described above) were estimated per standard deviation (SD) higher cytokine levels using multivariable linear regression adjusted for age

and sex. Separate analyses adjusting for age, sex and BMI as well as a model additionally adjusting for fasting insulin levels were performed. A significance threshold of P<0.003 was used, accounting for 17 tests. All analyses were performed using STATA v14 (StataCorp, Texas, USA).

Associations between cytokine levels and risk of incident CHD and T2D

The association between inverse-rank normal transformed cytokine levels (exposure) and incident T2D or CHD (outcomes) was estimated using multivariable Cox regression models in the EPIC-Norfolk[241] cohort. Inverse-rank normal transformed cytokine levels were used to facilitate comparison of effect sizes across the four cytokines regarding incident disease risk. Cytokine measurements were performed on blood samples collected at health check 2 in EPIC-Norfolk and therefore this serves as the study "baseline" for the purpose of this analysis.

Incident T2D was defined as new-onset T2D in a participant without evidence of diabetes at health check 2 (when IL-6 levels were measured), defined on the basis of (a) a hospital admission record or mortality registry record of "type 2 diabetes mellitus" (International Statistical Classification of Diseases and Related Health Problems Tenth Revision [ICD-10] code E11); or (b) self-reported physician diagnosis of T2D during a follow-up visit; or (c) a value of HbA1c above 6.5% (48 mmol/mol) at a follow-up visit. Participants with T2D at the time of cytokine measurement were excluded. Incident CHD was defined as new-onset CHD in a participant without CHD at baseline based on an electronic health record consistent with ischemic heart disease (ICD-10 codes I20-25). The date of end of follow-up for these analyses was 31st of March 2016. Censoring criteria were either an incident disease event (disease cases) or the end of follow-up time (non-cases). Participants with confirmed prevalent disease at baseline or unavailable prevalent disease status were removed from these analyses.

Associations with incident disease were estimated per 1 SD higher cytokine levels. Covariates included in a minimally adjusted model were age at baseline and sex. Extensively adjusted models for CHD included the following covariates: age at baseline, sex, BMI, WHR, education level, smoking status, average physical activity per week, systolic blood pressure, prevalent T2D, total cholesterol. Extensively adjusted models for T2D included the following covariates: age at baseline, sex, BMI, WHR, ethnicity, education level, family history of T2D, smoking status, units of alcohol per week, average physical activity per week. A significance threshold of P < 0.0125 was used, accounting for the four cytokines tested. Deciles of each cytokine were defined based on distribution of each cytokine in the non-case group. This was done to prevent skewing of the deciles by participants with incident disease, which were likely to have

higher cytokine levels. However, sensitivity analyses showed that the cut-offs between deciles were unaltered when considering the entire population. The association between deciles of each cytokine and incident disease were estimated as above, per 1 SD higher cytokine levels, relative to the lowest decile. The likelihood ratio test was used to test for departure from linearity when comparing treatments of the cytokine distribution as either a continuous or categorical measure. No departure from linearity was observed, therefore cytokine levels were treated as a continuous measure.

A number of sensitivity analyses were conducted to contextualise the associations between cytokine levels and incident disease. Firstly, an established definition for incident T2D[248] was used to validate the findings of the definition used above. This definition was based on a combination of primary and secondary care registers, self-report, medication use, hospital admissions, and mortality data[248]. The analysis was performed as described above, using a censor date of the 31st December 2007. Secondly, to estimate the effect of adjusting for correlated measures of overall (BMI) as well as regional (WHR) adiposity, a model adjusting for all covariates mentioned above was run without including either BMI or WHR. Subsequent models added BMI or WHR separately to the covariates to estimate their independent effects. Thirdly, to estimate the attenuation effect of adjusting for IL-6 and TNFα levels in the association of BMI, WHR or WHR adjusted for BMI (WHRadjBMI) with incident disease, a model adjusting for all covariates included in the main analysis was used without adjustment for BMI, WHR or WHRadjBMI. In subsequent models IL-6 and TNFα were then added iteratively to models to determine their attenuation effects. Lastly, to assess whether IL-6 and TNFα levels interact with either BMI or WHR in the association with incident disease, separate models adjusting for all covariates as in the main analysis were run within tertiles of the BMI and WHR distribution respectively, while including linear interaction terms for BMI and WHR. All analyses were conducted using STATA v14 (StataCorp, Texas, USA).

2.3: Results

Study participants' characteristics

Characteristics of the participants included in the analyses are shown in **Table 2.1**. Compared to participants who attended the baseline health check in EPIC-Norfolk[241], participants in this study were more likely to be older and female with a reduced likelihood of being a current smoker. Cytokine levels were measured at baseline in the Fenland study, therefore, participant characteristics were highly comparable to the wider study population[249].

Table 2.1: Participant characteristics for the Fenland and EPIC-Norfolk studies

Study	Fenland	EPIC-Norfolk
Participants, N	10,335	7,514
Age at baseline, mean years (SD)	48 (7)	62 (9)
Women, N (%)	5,491 (53)	4,598 (61)
Men, N (%)	4,844 (47)	2,916 (39)
Current smokers, N (%)	1,271 (12)	581 (8)
BMI in kg/m^2, mean (SD)	26.8 (4.9)	26.4 (3.8)
Waist-to-hip ratio, mean (SD)	0.88 (0.09)	0.84 (0.09)
Systolic blood pressure in mmHg, mean (SD)	123 (15)	134 (18)
Diastolic blood pressure in mmHg, mean (SD)	74 (10)	81 (11)
LDL cholesterol in mmol/L, mean (SD)	3.4 (0.9)	3.8 (1)
HDL cholesterol in mmol/L, mean (SD)	1.5 (0.4)	1.5 (0.5)
Triglycerides in mmol/L, median (IQR)	1 (0.7, 1.4)	1.6 (1.1, 2.2)
IFNγ in pg/mL, median (IQR)[a]	2.83 (2.02, 4.26)	3.37 (2.35, 5.10)
IL-6 in pg/mL, median (IQR) [a]	0.51 (0.34, 0.75)	0.59 (0.42, 0.85)
IL-8 in pg/mL, median (IQR)[b]	9.69 (7.7, 12.19)	6.7 (5.09, 9.19)
TNFα in pg/mL, median (IQR) [a]	1.56 (1.3, 1.88)	1.84 (1.53, 2.21)
IFNγ in log-pg/mL, mean (SD) [a]	1.13 (0.77)	1.28 (0.74)
IL-6 in log-pg/mL, mean (SD) [a]	-0.77 (0.94)	-0.55 (0.83)
IL-8 in log-pg/mL, mean (SD)[b]	2.28 (0.39)	1.98 (0.61)
TNFα in log-pg/mL, mean (SD) [a]	0.44 (0.32)	0.61 (0.33)

Abbreviations: N/A, not available; N, number of participants; SD, standard deviation; BMI, body mass index; IFNγ, Interferon-gamma; IL-6, Interleukin-6; IL-8, Interleukin-8; TNFα, Tumour necrosis factor alpha; IQR, Interquartile range

a. Participant numbers were the same for IFNγ, IL-6 and TNFα as they were measured from the start of the study
b. IL-8 levels were measured in a subset of participants

Cytokine measurement using the MSD platform

Pro-inflammatory cytokine levels were measured in 10,335 people from the Fenland study and 7,514 people from the EPIC-Norfolk study (**Table 2.2**). The distribution parameters and assay performance metric including the limits of detection and CVs for the MSD assay are shown in **Table 2.2**. Mean intra-plate CVs ranged from 3 to 7% in EPIC-Norfolk and from 4 to 8% for Fenland. Detection limits for the respective cytokines were largely comparable between Fenland and EPIC-Norfolk. Median levels of the four cytokines showed very little variability between the Fenland and EPIC-Norfolk studies (**Table 2.1**).

Table 2.2: Distribution parameters and assay performance across the four cytokines in Fenland and EPIC-Norfolk								
Study	**Fenland**	**EPIC-Norfolk**	**Fenland**	**EPIC-Norfolk**	**Fenland**	**EPIC-Norfolk**	**Fenland**	**EPIC-Norfolk**
Cytokine	**IFNγ**		**IL-6**		**IL-8**		**TNFα**	
Total samples	9,966	7,292	10,335	7,420	10,335	7,514	10,068	7,447
Samples in detection range (%)	97.64	99.00	78.68	89.78	87.87	99.92	99.83	99.92
Samples below detection range (%)	1.71	0.89	18.55	9.99	0.02	0.04	0.06	0.05
Samples below assay fit curve (%)	0.07	0.11	1.67	0.23	0	0.04	0.01	0.03
Samples failed due to technical error (%)[a]	0.58	0	1.09	0	0.01	0	0.10	0
Missing samples (%)[b]	0	0	0	0	12.10[c]	0	0	0
Mean LLOD (pg/mL)	0.79	0.78	0.31	0.30	0.17	0.18	0.16	0.16
Mean ULOD (pg/mL)	2,801.68	2,994.14	1,312.22	1,300.92	1,141.90	1,148.13	652.34	653.62
Mean Intra-plate CV (%)	7	7	8	7	4	3	6	5

Abbreviations: IFNγ, Interferon-gamma; IL-6, Interleukin-6; IL-8, Interleukin-8; TNFα, Tumour necrosis factor alpha; LLOD, Lower limit of detection; ULOD, Upper limit of detection; pg/mL, picograms per millilitre; CV, Cross-validation; %, percentage

a. Samples which failed due to poor duplicate (intra-plate duplicates >20% difference in measures) or technical error
b. Samples for which no measurement was available – not due to technical error
c. IL-8 measurements had more missing samples as early measurements were conducted on IL-1β and the remainder of samples were measured for IL-8.

Correlations and observational association between cytokine levels and cardiometabolic risk factors

Correlation estimates showed that IL-6 and TNFα were strongly correlated with the majority of cardiometabolic risk factors whereas IFNγ and IL-8 were correlated with these to a lesser extent (**Figure 2.1**). HbA1c was estimated to be the trait with the strongest correlation with IL-8 levels.

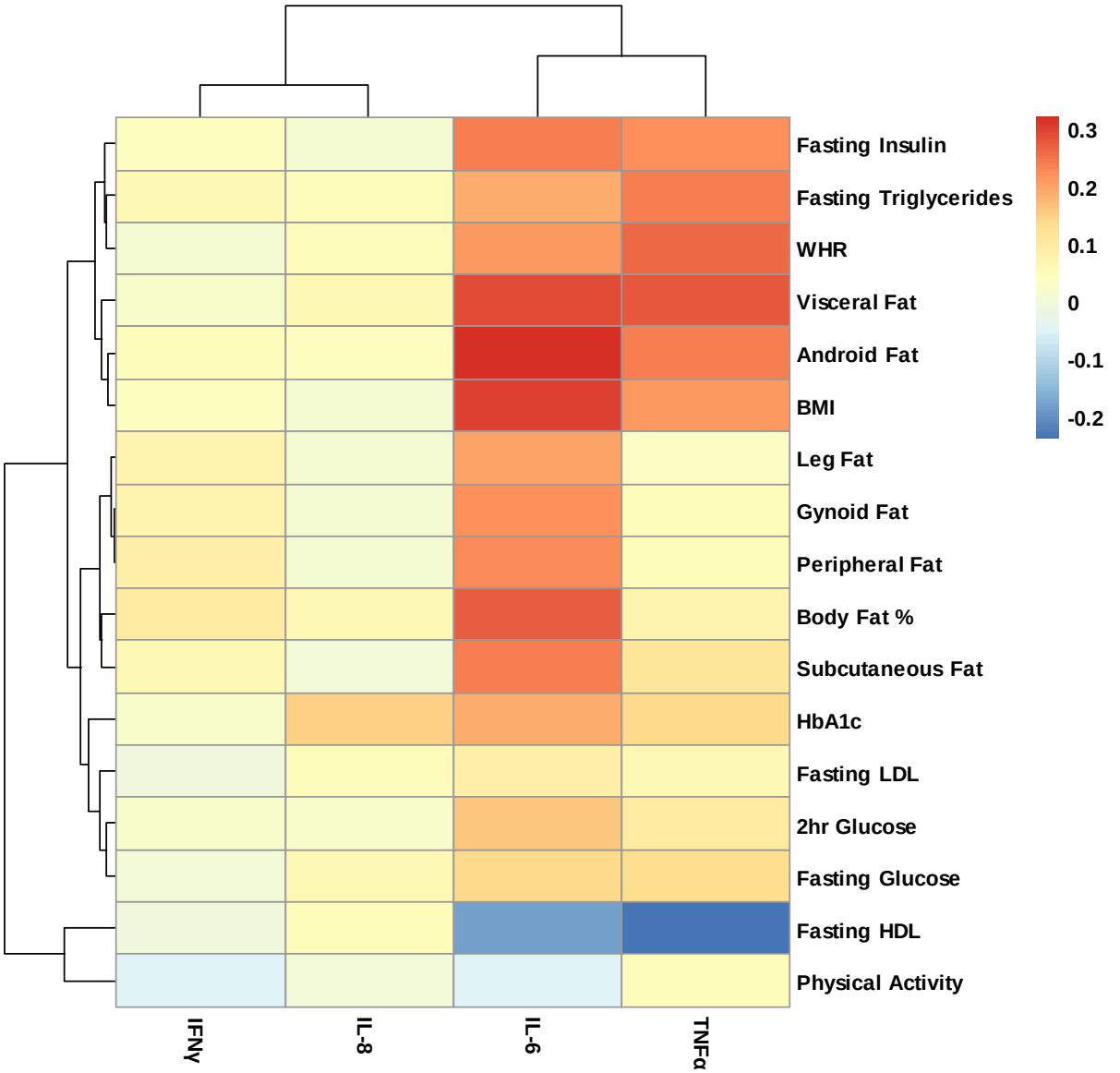


Figure 2.1: Heatmap of the correlations between the four cytokines and cardiometabolic risk factors. All traits are unadjusted for BMI and clustered by correlation estimate similarity using hierarchical clustering. Abbreviations: BMI, Body mass index; WHR, Waist-to-hip ratio; HbA1c, Glycated haemoglobin; HDL, High-density lipoprotein; LDL, Low-density lipoprotein; IFNγ, Interferon gamma, IL-6, Interleukin-6; IL-8, Interleukin-8; TNFα, Tumour necrosis factor alpha

The association between cardiometabolic traits and cytokine levels were estimated in 10,335 participants from the Fenland study (**Figure 2.2**). The association between cytokine levels and cardiometabolic traits was broadly consistent with the correlations presented above, in terms of estimate direction. Of the traits significantly associated with cytokine levels, only HDL and physical activity were shown to be significantly lowered by higher cytokine levels, all others were raised. Levels of IFNγ, IL-6 and TNFα showed broadly consistent directions of association with all measures of regional and overall adiposity, insulin and some lipid measures. However, the estimated strength of association with each trait was weaker for IFNγ levels compared to IL-6 and TNFα.

In line with the hypothesis that inflammation in cardiometabolic diseases is likely to be driven by obesity, BMI adjustment led to the attenuation of the majority of association estimates. This effect was particularly evident for measures of regional adiposity. Levels of all four cytokines remained associated with WHR after BMI adjustment, despite a considerable attenuation in effect size. IFNγ, IL-6 and TNFα levels remained associated with body fat percentage and android fat mass, whereas only IL-6 and TNFα were associated with visceral fat mass. Only TNFα levels were associated with subcutaneous fat mass after BMI adjustment.

Levels of IFNγ, IL-6 and TNFα also remained associated with lower HDL but higher WHR, fasting insulin and triglycerides following BMI adjustment, an effect consistent with an insulin resistance phenotype. Although adjusting the association between IFNγ, IL-6 and TNFα and these traits for both BMI and fasting insulin levels led to a minor attenuation in effect estimate strength, all associations remained significant (**Figure S2.1**). Compared to the other three cytokines, IL-8 levels were significantly associated with comparatively fewer traits overall. IL-8 levels were associated with significantly higher WHR, body fat percentage, android fat mass, visceral fat mass and HbA1c after adjustment for both BMI and fasting insulin levels (**Figure S2.1**).

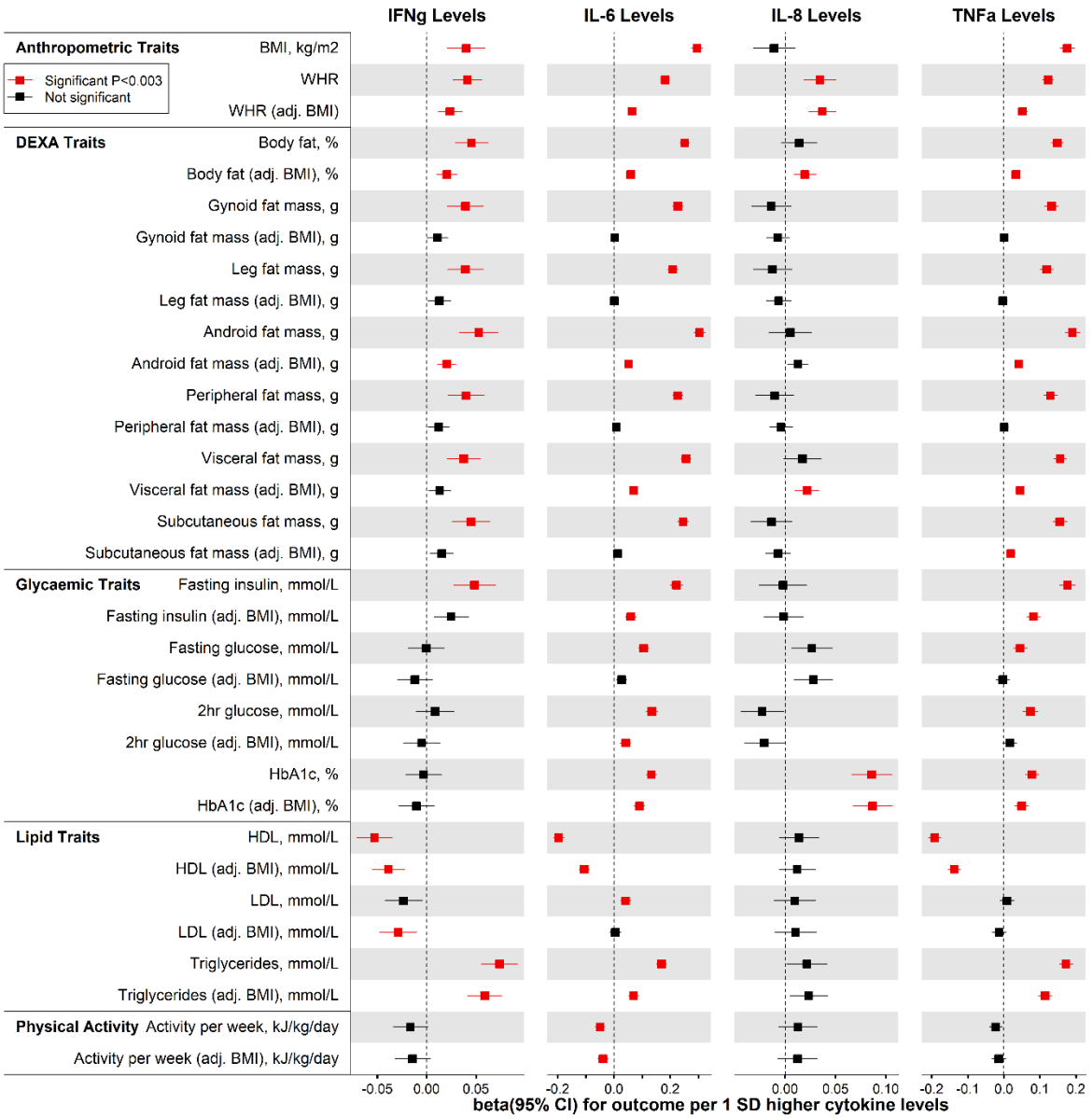


Figure 2.2: Associations of the four cytokines with cardiometabolic risk factors per 1 SD higher cytokine levels. Risk factors are grouped into anthropometric, DEXA, glycaemic, lipid and physical activity categories. Traits adjusted for BMI are marked as such in the trait name. Traits that are significantly associated (P<0.003) with a respective cytokine are coloured in red, non-significant results are in black. Effect estimates are expressed per 1 SD higher cytokine levels.

Abbreviations: BMI, Body mass index; adj., Adjusted for; WHR, Waist-to-hip ratio; mmol/L, millimoles per Litre; HbA1c, Glycated haemoglobin; HDL, High-density lipoprotein; LDL, Low-density lipoprotein; g, Grams; Avg., Average; kJ, Kilojoules; Kg, Kilograms; IFNγ, Interferon gamma, IL-6, Interleukin-6; IL-8, Interleukin-8; TNFα, Tumour necrosis factor alpha.

Associations of cytokine levels with incident cardiometabolic diseases

In EPIC-Norfolk, the incidence rates of T2D and CHD were 3.3 cases per 1000 person-years and 11.6 cases per 1000 person-years respectively. IL-6 and TNFα levels were significantly associated with higher hazard of incident T2D with hazard ratios (HR) per SD higher cytokine levels of 1.19 (95% CI: 1.07, 1.32; P = 0.002) and 1.16 (95% CI: 1.04, 1.30; P = 0.008) respectively (**Figure 2.3**). IL-6 and TNFα levels were also significantly associated with higher

hazard of incident CHD (HR = 1.16; 95% CI: 1.09, 1.24; P = $3x10^{-6}$) and (HR = 1.16; 95% CI: 1.09, 1.24; P = $4x10^{-6}$) respectively (**Figure 2.3**). The magnitude of the association between levels of these two cytokines and each cardiometabolic disease were highly comparable and directionally consistent. No significant associations between either IFNγ or IL-8 and either disease were found. Associations with log-transformed cytokine levels and incident cardiometabolic diseases are shown in **Figure S2.2**. Associations between the deciles of each cytokine and incident T2D and CHD are shown in **Figures S2.3 and S2.4**, respectively. No significant departures from linearity were observed for any of the four cytokines.

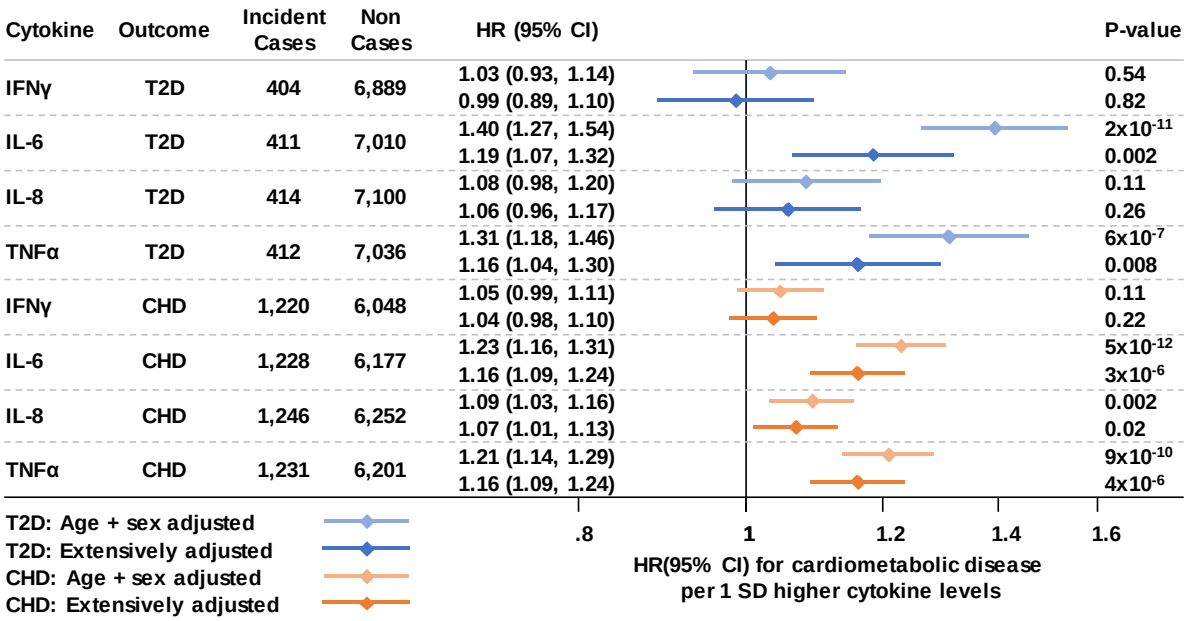


Figure 2.3: Association of inverse-rank normal transformed cytokine levels with incident cardiometabolic diseases in EPIC-Norfolk. Results for T2D and CHD adjusted for age and sex only are shown in light blue and light orange respectively. Extensively adjusted results for T2D and CHD are shown in blue and orange respectively. **T2D specific covariates:** age, sex, BMI, WHR, ethnicity, education level, family history of diabetes, smoking status, units of alcohol per week and physical activity. **CHD specific covariates:** age, sex, BMI, WHR, education level, smoking status, physical activity, systolic blood pressure, prevalent diabetes and total cholesterol.

Abbreviations: HR, Hazard ratio; SD, Standard deviation; BMI, Body mass index; WHR, Waist-to-hip ratio; T2D, Type 2 diabetes; CHD, Coronary heart disease; IFNγ, Interferon gamma, IL-6, Interleukin-6; IL-8, Interleukin-8; TNFα, Tumour necrosis factor alpha.

Sensitivity analyses to validate the definition of incident T2D replicated the significant associations of IL-6 and TNFα levels with incident T2D as well as the non-significant association between IFNγ and incident T2D (**Figure S2.5**). While fewer incident cases were included in the analysis using this definition, the effect directions were consistent, and the strengths of the effect estimates were marginally stronger. In addition, a significant association between IL-8 levels and T2D was found (HR, 1.23; 95% CI, 1.09, 1.39; P=0.001). Sensitivity analyses estimating the effect of adjusting for BMI and WHR showed considerable attenuation of association effect estimates for each cytokine with incident disease when compared to

models where BMI and WHR were not adjusted for, as expected (**Figure S2.6**). No association for IFNγ with either incident CHD or T2D was shown in unadjusted models. However, IL-8 was significantly associated with CHD in a model without either BMI or WHR adjustment. Similarly, adjustment for IL-6 and TNFα levels in the association of BMI, WHR and WHRadjBMI with incident CHD or T2D showed marginal attenuation of association estimates as a result of adjusting for cytokine levels (**Figure S2.7**). Association estimates for all three adiposity measures were smaller for CHD compared to those for T2D. All adiposity measures were significantly associated with incident CHD or T2D, except for WHRadjBMI with CHD (**Figure S2.7)**. In addition, no significant interactions were found between cytokine levels and either BMI or WHR in the association of cytokine levels with incident cardiometabolic diseases (**Figure S2.8**).

2.4: Discussion

This biomarker study supports the hypothesis that proinflammatory cytokine levels are associated with aetiology of cardiometabolic diseases. However, these results suggest that the impact of cytokine levels on disease risk in the general population is likely to be small and largely driven by higher BMI. This study is the first to investigate the associations between cytokine levels and measures of regional adiposity from DEXA scans and make use of them for extensive adjustment when estimating the association with incident cardiometabolic diseases. This study also represents an approximate 1.5 fold increase in sample size used to estimate these associations compared to previous studies[155,174,175,183,186,225,238–240].

IL-6 and TNFα levels were found to be significantly associated with incidence of both CHD and T2D, even with extensive adjustment for established cardiometabolic risk factors. The magnitude and direction of association for both cytokines with incident T2D was comparable to recent meta-analyses consisting of 4,480 incident cases and 19,709 non-cases from 10 prospective studies of IL-6 levels[186] and 2,780 incident cases and 10,078 non-cases from 5 prospective studies of TNFα levels[185]. Similarly, associations of both cytokines with incident CHD were comparable to previous, smaller European studies[189,190]. This analysis showed that higher IFNγ levels were not significantly associated with either incident disease outcome, providing the first evidence that IFNγ levels are not associated with incident T2D. in addition, the results of this study replicate previous findings in a European cohort for incident CHD[190] consisting of 931 CHD cases and 971 controls.

In comparison, in the main analysis IL-8 levels were not found to be associated with incidence of either CHD or T2D, results which are consistent with previous findings from smaller prospective studies[175,183,187,188]. However, results from a sensitivity analysis using an established definition of incident T2D[248] showed an association between IL-8 levels and incident T2D. There are several explanations for this, 1) the main analysis used a definition that included HbA1c measures where participant HbA1c measures above 6.5% at a single timepoint would have led to the inclusion of these participants as incident cases, even if they were not strictly incident cases. This may therefore have contaminated the pool of participants used as incident cases, contributing to a higher incident case number while diluting any putative association of IL-8 levels with incident T2D. 2) The censor date in the main analysis was considerably longer than the sensitivity analysis, 31st March 2016 versus 31st December 2007. Over the much longer interval, the effect of IL-8 levels on T2D incidence may have been less pronounced, which may also partly account for the reason why association estimates were stronger in the sensitivity analysis. Interpreting this within the context of findings from this study showing that IL-8 levels are associated with higher HbA1c levels, shows that HbA1c

levels may well mediate the effect of IL-8 on incident T2D risk. Together these results indicate that IFNγ levels are not associated with risk of incident cardiometabolic disease events, especially when compared to IL-6 and TNFα, despite being associated with several metabolic risk factors. The association between IL-8 levels and incident T2D will require validation in large, independent cohorts.

Levels of IFNγ, IL-6 and TNFα were all found to be associated with higher overall and regional adiposity measures, fasting insulin levels, triglycerides and lower HDL levels. These associations with measures of regional adiposity reflect previous findings which demonstrate higher circulating cytokine levels in the plasma of obese human participants, as discussed in **Chapter 1**[250–254]. In line with previous findings, this study found associations between IFNγ, IL-6 and TNFα with fasting insulin. This is thought to be a result of compensatory increases in insulin production in the early stages of obesity-related insulin resistance[111,174,181,255,256]. Higher IFNγ, IL-6 and TNFα levels were associated with lower HDL cholesterol levels, even after adjusting for BMI, as has been previously shown[190,257,258]. Lower HDL and higher LDL levels are associated with increased CHD risk[259–262]. The anti-inflammatory capacity of HDL is thought to protect against CHD and T2D by controlling cytokine production through cholesterol efflux from macrophage foam cells and the concurrent inhibition of LDL oxidation[259,263]. Alternatively, this association may be secondary to the association with insulin resistance, of which lower HDL but higher triglycerides, WHR and fasting insulin levels are biomarkers[259]. However, adjusting the association between cytokine levels and these traits for both fasting insulin and BMI led to a marginal attenuation in effect estimates while remaining significantly associated. Future work using genetics to estimate the association between cytokine levels and these traits, as genetic estimates are robust to residual confounding[61], will help to estimate whether obesity-mediated insulin resistance may be a potential driver of cytokine levels in cardiometabolic diseases. Similar association patterns were observed for IL-6 and TNFα in terms of estimated strength and direction for all associations except for LDL cholesterol and physical activity where no significant associations with TNFα were found.

While IFNγ levels showed similar patterns of association to IL-6 and TNFα regarding measures of overall and regional adiposity, comparatively fewer associations with glycaemic traits were found. This is mirrored by the comparatively fewer traits found to be associated with IL-8 levels. In fact, IL-8 levels were only found to be associated with higher WHR and HbA1c levels, which are in line with previous reports[224,264,265]. The association between IL-8 levels and higher HbA1c may be mediated via pathways independent of BMI, as BMI adjustment did not attenuate the association estimate. This may indicate that while IFNγ and IL-8 levels are raised in obesity, their effects on glycaemic control may not be as pronounced as either IL-6 or TNFα.

The results of this study therefore suggest that many of the observed associations between cardiometabolic traits and cytokine levels, particularly IFNγ, IL-6 and TNFα, are likely to be mediated by higher BMI, in line with established findings[168,174,189,224,266,267]. This is evidenced by the attenuation of many significant associations with cardiometabolic traits when BMI was accounted for. Comparatively, the effects of IL-8 levels on cardiometabolic traits may be independent of BMI.

There are limitations to this study. Firstly, in EPIC-Norfolk cytokine levels were measured at health check 2 and not at baseline. As only a subset of participants that attended the baseline assessment returned for the second heath check, this will have led to a marginal skewing of the baseline population characteristics. As a result, participants who returned for the second health check were older and more likely to be female than at baseline, therefore, it is possible that the estimates obtained in this study may be affected by selection bias. Secondly, the associations between cytokine levels and risk factors for cardiometabolic diseases were exploratory in nature and were therefore only adjusted for age, sex and BMI. This therefore does not rule out the possibility of residual confounding. Thirdly, as this work is observational in nature, definitive assessment of causality between cytokine levels and cardiometabolic diseases is not possible in this context. Finally, IL-8 levels had considerably higher missingness levels as IL-1β levels were measured initially. This may have led to decreased power for some of the analyses, particularly the Cox regression estimating the association of IL-8 levels with incident CHD, where IL-8 levels were close to the significance threshold.

Future work could involve testing interactions between cytokine pairs on their respective associations with cardiometabolic risk factors, for example, testing the interaction between IL-8 and IL-6 levels in their association with HbA1c, as IL-6 levels have been shown to be significantly correlated with IL-8 levels[224]. This would provide further insight into the biology of the inflammatory component that is present in cardiometabolic diseases and may show how cytokines act synergistically to influence phenotypes. Similarly, mediation analyses would aid our understanding of the extent to which inflammation contributes to established pathways for risk of cardiometabolic diseases. Finally, investigating whether the inflammation observed in cardiometabolic diseases predisposes patients to a worse prognosis or disease complications would be invaluable to improving the clinical treatment of patients with cardiometabolic diseases.

This study showed that IL-6 and TNFα are associated with incident cardiometabolic diseases, however, the predictive value of a biomarker is dictated by the strength of its association with

incident disease[268]. Given the comparatively small estimates of association between cytokine levels and incident disease, it is unlikely that the measurement and consideration of IL-6 and TNFα will have more than an incremental effect in predicting incident cardiometabolic diseases above that of clinical risk factors. In this case, the predictive value of each cytokine may not be of primary importance but rather whether cytokine levels are causally associated with either disease. In the case where the levels of a particular cytokine are estimated to be causally relevant for cardiometabolic diseases, even relatively small changes in cytokine levels may be therapeutically targeted and bring about clinically relevant reductions in disease risk. Approaches using genetic variants such as Mendelian randomisation (MR) can try to address this question, as associations between variants and phenotypes of interest are less susceptible to the effects of confounding[269]. Causal inference is difficult in observational studies due to this confounding and the possibility for reverse causality[72]. Thus, further study into the genetics underpinning cytokine levels is critical to aid our wider understanding of inflammation but also to assess the causality of cytokine levels with cardiometabolic diseases and thereby further our understanding of disease aetiology.

Chapter 3: Discovery, Refinement and Characterisation of Loci Associated with Cytokine Levels

Contributions and collaborations

I designed the study, prepared phenotypes for GWAS and performed all analyses described in this chapter. Jian'an Luan provided statistical support when running the GWAS. I wrote the first draft of this chapter which was later reviewed by both Luca Lotta and Claudia Langenberg.

Abstract

Background: To date, very few GWAS studies of cytokine levels have been conducted, most have been conducted in small sample sizes, identifying variants which are associated with IL-6 and TNFα levels. No variants have been found to be associated with levels of IL-8 and IFNγ at genome-wide significance.

Aims: To conduct a GWAS and meta-analysis of IFNγ, IL-6, IL-8 and TNFα levels in the Fenland and EPIC-Norfolk studies.

Methods: GWAS was performed on levels of each cytokine in Fenland and EPIC-Norfolk separately, using multivariable linear regression. The results were then combined using inverse-variance weighted fixed effects meta-analysis. Variants were clumped into 2Mb loci, 1Mb either side of the lead variant in each region. Secondary signals in each locus were then identified using a step-wise selection procedure in GCTA and conditional analysis, conditioning on the effects of the lead variant in each locus, was performed using multivariable linear regression. Bayesian fine mapping was then used to generate 99% credible sets at each locus. Next, SNP-based heritabilities and correlations between the four cytokines were estimated using LD score regression. Finally, the colocalisation between cytokine levels and T2D and CHD was estimated using a pairwise Bayesian colocalisation framework, COLOC.

Findings: In up to 17,000 participants, the largest cytokine GWAS to date, 22 variant-cytokine associations were identified within 19 loci, 16 of which are novel associations with the corresponding cytokine and not in the *HLA* gene region. Six loci were associated with IFNγ levels, three with IL-6 levels, five with IL-8 levels and six with TNFα levels. An intronic variant in *IL6R,* rs4129267, in complete LD with a missense variant, rs2228145, was associated with 0.11 SD higher IL-6 levels (95% CI, 0.08, 0.14; $P = 3.5x10^{-38}$) per copy of rs4129267 was replicated. Variants residing in the receptors of IFNγ (*IFNGR1*), IL-6 (*IL6R*) and IL-8 (*CXCR2, DARC*) were found to be associated with each respective cytokine and explained the greatest proportion of variance in their respective levels. Of these, only the *IL6R* locus showed robust evidence of colocalisation with cardiometabolic diseases (PP_{coloc} with CHD = 0.98).

Conclusion: This large cytokine GWAS further elucidates the genetic determinants of IL-6 and TNFα but most importantly, provides the first genome-wide evidence of variants associated with IL-8 and IFNγ levels.

3.1: Introduction

Evidence from observational studies and animal models points to a role for chronic inflammation in cardiometabolic disease pathophysiology[109,112,123,154,155,184–186,189]. Four key cytokines thought to play roles in chronic inflammation, atherosclerosis and insulin resistance are IFNγ, IL-6, IL-8 and TNFα[9,94,123,138,162,223,270]. In order to further our understanding of how these cytokines may affect cardiometabolic disease risk, I first need a better understanding of the genetics that underpin their levels. GWAS approaches systematically test the association between genetic variants and phenotypes of interest, highlighting loci of interest for follow-up.

The systematic literature review conducted in **Chapter 1** highlighted three loci associated with IL-6 levels *ABO*, *IL6R* and *RP11-314E23.1* whereas two loci, *DLEU1* and *ABO*, were associated with levels of TNFα. To-date, no significant loci have been shown to be associated with IL-8 or IFNγ levels in GWAS studies, despite both being investigated in a recent GWAS in Finnish participants[227]. To-date, the Finnish GWAS[227] investigating the levels of 27 cytokines and growth factors is the largest of its kind, as many previous GWAS efforts have been conducted in considerably smaller sample sizes.

This chapter describes the results of GWAS for the levels of the four cytokines of interest in the largest GWAS to date. Using the loci estimated to be associated with cytokine levels, loci of interest were identified for further follow-up. Finally, the respective heritability and pairwise genetic correlations between the four cytokines were estimated.

3.2: Methods

Phenotype preparation

Levels of IFNγ, IL-6, IL-8 and TNFα were measured in both EPIC-Norfolk[241] and Fenland[244] respectively, using the method described previously. Any samples where measurement was attempted but measures were below the LLOD were imputed using a random value between 0 and the LLOD. This imputation was done to retain statistical power. Samples where measurement was not attempted or failed due to technical failure were set to missing. Cytokine measurements were normalised using an inverse-rank normal transformation and residuals were generated by regression of the normalised values using age, age^2, sex and principal components as covariates. The generated residuals were then normalised using another round of inverse-rank normal transformation.

GWAS for IFNγ, IL-6, IL-8 and TNFα levels

Genome-wide genotyping in the Fenland cohort was performed in 3 sub-cohorts using either the Affymetrix genome-Wide Human variant Array 5.0, the Affymetrix UK Biobank Axiom Array or the Illumina CoreExome-24 v1 chip, with imputation to the Haplotype reference consortium v1.1[271], the 1000 genomes project[272] and the UK10K[273] reference panels. Samples from EPIC-Norfolk were genotyped using the Affymetrix UK Biobank Axiom Array and imputed to the same reference panels. GWAS analyses to estimate the association of genetic variants with cytokine levels were conducted in the Fenland and EPIC-Norfolk cohorts separately for IFNγ, IL-6, IL-8 and TNFα. variant associations were estimated using multivariable linear regression models in variantTEST[274]. Results were assessed for quality and consistency using several metrics. variants that had a minor allele count (MAC) lower than 10 were excluded along with those with evidence of deviation from Hardy Weinberg equilibrium ($p<10^{-6}$). Variants with beta estimates with absolute values above 10 were removed. variants with standard errors larger than 10 or with a value of 0 due to model instability were removed. Imputation quality was assessed using the "INFO" metric assigned during imputation using Impute2[275], where a value of 1 denotes perfect imputation or a directly genotyped variant. A cut-off value of 0.4 or above was used to retain variants of higher imputation quality. variants with sample sizes less than 50% of the maximum sample size were excluded.

Meta-analysis and quality control of GWAS analyses

GWAS were conducted in EPIC-Norfolk and Fenland separately and results were combined using fixed-effects inverse-variance weighted meta-analysis in METAL[276]. Following this, any variant that was not successfully genotyped for more than half of the sample size was removed. variants available in only 1 sub-study were dropped. variants that had MAC counts of less than 20 were removed. Insertions and deletions were removed from the analyses along with tri-allelic sites along with any variants with highly variable effect allele frequencies and corresponding standard errors. Lead variants were defined as those with the lowest p-values in the regions and loci were then defined based on 2Mb windows around each of the lead variants (1Mb either side of the lead variant). Manhattan plots were drawn using the EasyStrata[277] R-package under the R programming environment v3.3.1 (R core team 2017). A genome-wide significance threshold of $p<5x10^{-8}$ was used. QQ-plots were drawn for quality control assessment of the meta-analysis using STATA v14 (StataCorp, Texas, USA).

Regional association plots and independent signal selection

Regional association plots were drawn around each of the lead variants using LocusZoom v1.2[278] under the R programming environment v3.3.1 (R core team 2017) to examine the linkage disequilibrium (LD) structure within each 2Mb locus. LD structure was drawn from the

European populations of the 1000 Genomes Project[272] and genomic coordinates from the human genome assembly GRCh37 (hg19).

At each 2Mb locus defined around the lead variants, I sought to identify any independent secondary signals that met a locus-wide significance threshold of $P < 1x10^{-5}$, corresponding to the approximate number of variants in each window and has been used in previous studies[279,280]. A 10Mb window was used for a locus on chromosome 12 locus significantly associated with TNFα, as the LD structure spanned a region greater than 2Mb (**Figure S3.1**). Initially, independent variants were identified using GCTA cojo-select[281] which implemented a step-wise selection procedure, taking into account regional LD and the distance between genome-wide significant variants from GWAS summary data. At loci where there were multiple independent variants, association signals were decomposed using approximate conditional analysis using GCTA cojo-cond[281] by including the other independent variants within the region as covariates in a regression model. The method applied by GCTA uses R^2 LD metrics which capture the correlation between common variants well, however, do not perform well in scenarios where common and rare variants are correlated with one another. As a result of this, GCTA often found multiple independent variants at a locus, where one was common and the other rare. Given that the variants were often very rare and the p-values of were close to the significance threshold, a formal conditional analysis was used which was robust to the effects of rare variants and is not an approximation method.

Multivariable linear regression models were used to condition the variants at each non-*HLA* locus on the lead variant by including the genotype of the lead variant as a covariate. Genotypes from the largest Fenland genotyping sub-cohort of over 8,000 participants were used. Any variants that remained significant at $P< 1x10^{-5}$ were then added to the linear regression model as a further covariate. This process was repeated until no further variants were significant at the locus-wide threshold. All variants found to be significant at this threshold were considered secondary independent signals. Due to the extensive and complicated LD structure across the HLA region (chr6:28,477,797-33,448,354) signal decomposition and later fine mapping was not done in these regions. Following this, variants associated with levels of each respective cytokine were searched in the NHGRI-EBI GWAS catalog[282], a repository of published GWAS results, to determine the number of novel association signals.

Bayesian fine mapping

Independent signals were refined with the aim of narrowing each association signal to a causative variant set or single variant. Fine mapping was performed using 1Mb windows around each independent variant in each locus, as posterior probabilities have been previously

shown to decay to basal levels 500Kb either side of the association signal[283]. Functionally unweighted credible sets were defined, using a model that assumed only a single causal variant at each locus. variants were ranked according to their posterior probability of causality and included in credible sets until a cumulative posterior probability of 0.99 was reached. A prior of W = 0.02 was used as in previous studies[284] corresponding to the variance in allelic effects of standardised quantitative traits. Variant effect predictor (VEP) v96[285] annotations were used to annotate the regional and functional consequences of each variant.

Correlation between cytokines and variant-based heritability estimation

Pearson's correlations were estimated using the inverse-normal transformed residuals of the cytokine levels used to run the original GWAS. A correlation heat map was drawn using ggplot2[286] under the R programming environment v3.3.1 (R core team 2017) to graphically represent the relationship between each pair of cytokines.

LD-score regression[287] (LDSC) was used to estimate the genetic correlations across the four cytokines and their respective variant-based heritabilities. Prior to implementing LDSC, variants that displayed effect sizes larger than $\chi^2 > 80$ (corresponding to a Z-score of 8.9) were removed, as these outliers may bias the regression. Variants within the *HLA* region were also removed due to the complex LD structure in the region. To retain high imputation quality (> 0.9) the analysis was restricted to variants imputed to HapMap3, as these were considered to be well imputed across studies. A total of 1,217,312 variants were included in the LDSC regression, using European LD scores supplied with the LDSC package[287]. A second analysis without restricting the results to HapMap3 imputed variants was also conducted. Variant-based heritability estimates were then compared to estimates of other circulating biomarkers using the LD-hub database[288].

False positive associations may arise in GWAS due to genetic confounding, generally driven by population stratification or unaccounted relatedness between participants. The genomic control lambda (λ_{GC}) has been traditionally used to provide a measure of test statistic inflation and serve as a correction factor to counteract this. Values below 1.1 were considered indicative of suitable control of genetic confounding. However, the method is unable to distinguish genetic confounding from polygenic trait architecture. Polygenicity, as is often observed in common disease, is the presence of many associations across the genome with comparatively small effect sizes that cumulatively contribute to higher disease risk. The LDSC regression intercept, however, is a robust measure of confounding because it does not inflate in the same way as λ_{GC} does with larger sample sizes in the presence of polygenicity[287]. This robustness lends increased power when correcting GWAS estimates with the LDSC

intercept[287]. To this end, I examined the λ_{GC} and LDSC intercept estimates from a model which did not restrict the variants included to those imputed to HapMap3 only. Values for the intercept below 1.5 were considered to adequately control for genetic confounding in this GWAS.

Proportion of variance explained by variants associated with cytokine levels

To determine if particular variants explained more variance in the levels of each cytokine, the proportion of variance explained by each variant was estimated using the following formula:

$$PVE = \frac{2\beta^2 MAF(1-MAF)}{2\beta^2 MAF(1-MAF) + (SE^2)2NMAF(1-MAF)}$$

Where proportion of variance explained is denoted by PVE, β relates to the variant's effect on cytokine levels, SE relates to the standard error of the variant, N represents the number of participants and MAF represents the minor allele frequency of the variant. Variants were ordered from greatest proportion of variance explained to smallest per cytokine and a stacked bar plot was drawn to show the incremental increases in variance explained by each additional variant. Analyses were conducted using R v3.6.0 and plots were drawn using the ggplot2 package.

Colocalisation of variants associated with cytokine levels and cardiometabolic diseases

I employed a pairwise Bayesian genetic colocalisation framework to estimate whether variants explaining the greatest proportion of variance in cytokine levels were shared with either T2D or CHD. Using the above estimates, variants in *IFNGR1*, *IL6R*, *DARC* and *LIPC* explained the greatest proportion of variance in IFNγ, IL-6, IL-8 and TNFα levels respectively. The 1Mb regions either side of each variant were extracted from publicly available GWAS summary statistics for both diseases[22,289] and the respective cytokine GWAS from this study. GWAS summary statistics were first aligned to the cytokine-raising allele and any indels were removed. Any variants with standard errors of "0" were also removed from the analysis. Only variants present in all datasets were considered. Bayesian colocalisation analysis was then conducted using the COLOC[290] R package between each pair of traits using beta estimates and corresponding trait variances. Cytokine levels were treated as quantitative traits and both T2D and CHD were treated as case-control traits. The case proportion was used as input for case-control traits and standard deviations were used for quantitative traits. Colocalisation was done to estimate posterior probabilities (PP_{coloc}) denoting evidence of colocalisation: H0 – no signal; H1 – signal unique to trait 1; H2 – signal unique to trait 2; H3 – two independent causal variants in the same locus driving the association signal for the respective traits and H4 – presence of a causal variant shared between two traits. The prior probability that a variant

is associated with trait 1, p1, was set to $1x10^{-4}$. The same settings were used for p2. I assigned a prior probability of $1x10^{-5}$ for p12, which relates to the prior probability that a single variant is associated with both traits. Pairwise PP_{coloc} estimates were considered significant if they met the following criteria: (H4 + H3 ≥ 0.9 & H4/H3 ≥ 3. All data analysis was performed using R v3.6.0.

3.3: Results

GWAS and meta-analysis for loci associated with cytokine levels

Meta-analysis of the Fenland and EPIC-Norfolk cohorts resulted in a sample size of approximately 17,000 for each cytokine, with a mean of 16,248,771 variants analysed. **Table S3.1** details the number of variants removed during meta-analysis quality control procedures. These meta-analyses identified 153, 131, 1,004 and 710 variants that were associated with circulating IFNγ, IL-6, IL-8 and TNFα levels respectively at genome-wide significance ($P \leq 5 \times 10^{-8}$). **Figure 3.1** shows the Manhattan plots for each of the meta-analyses and **Figure S3.2** shows the QQ-plots for each meta-analysis. **Table 3.1** shows the lead variants of each association signal defined on the basis of distance-based pruning. This locus definition led to a total of 20 lead variants within 6 loci associated with IFNγ levels, 3 loci for IL-6, 5 loci for IL-8 and 6 loci for TNFα.

Table 3.1: Lead variants used for independent signal selection

Trait	Locus	Variant	Chr	Pos	EA	EAF	Beta (SE)	p-value
IFNγ	1	rs11754268	6	137540335	T	0.21	0.11 (0.013)	6.08×10^{-16}
IFNγ	2	rs76830965	3	159637678	A	0.12	0.13 (0.017)	1.17×10^{-14}
IFNγ	3	rs653178	12	112007756	T	0.53	-0.07 (0.011)	7.19×10^{-10}
IFNγ	4	rs2523609	6	31237255	A	0.62	0.07 (0.011)	1.11×10^{-09}
IFNγ	5	rs4492899	12	10598426	A	0.31	0.07 (0.012)	2.47×10^{-08}
IFNγ	6	rs2009581	2	111807677	A	0.27	-0.07 (0.012)	2.63×10^{-08}
IL-6	1	rs12730935	1	154419892	A	0.41	0.14 (0.011)	4.04×10^{-38}
IL-6	2	rs116457146	6	32578632	T	0.26	0.07 (0.012)	2.20×10^{-08}
IL-6	3	rs545129521	21	36823802	A	0.99	-0.71 (0.13)	4.27×10^{-08}
IL-8	1	rs12075	1	159175354	A	0.58	0.36 (0.011)	7.08×10^{-243}
IL-8	2	rs138840656	4	74567906	T	0.94	0.24 (0.024)	2.33×10^{-22}
IL-8	3	rs55799208	2	218999982	A	0.003	0.89 (0.118)	3.71×10^{-14}
IL-8	4	chr6:32634613:C>T	6	32634613	T	0.11	0.12 (0.02)	7.17×10^{-09}
IL-8	5	rs6503533	17	38184580	T	0.63	0.06 (0.011)	1.67×10^{-08}
TNFα	1	rs7161799	15	58770523	T	0.08	0.22 (0.018)	2.43×10^{-35}
TNFα	2[a]	rs3184504	12	111884608	T	0.47	0.11 (0.011)	3.09×10^{-22}
TNFα	3	rs28611443	6	31341411	A	0.08	-0.15 (0.02)	1.01×10^{-14}
TNFα	4[a]	rs11066320	12	112906415	A	0.42	0.08 (0.01)	2.95×10^{-13}
TNFα	5	rs146125856	15	50784990	T	0.90	0.12 (0.019)	8.83×10^{-10}
TNFα	6	rs66530140	4	187161211	T	0.49	0.07 (0.011)	1.77×10^{-09}

Abbreviations: IFNγ, Interferon-γ; IL-6, Interleukin-6; IL-8, Interleukin-8; TNFα, Tumour necrosis factor-α; Chr, Chromosome; Pos, Base-pair position; EAF, Effect allele frequency; EA, Effect allele; SE, Standard error.

a. Loci 2 and 4 were later combined into a single locus due to the LD structure in the surrounding region

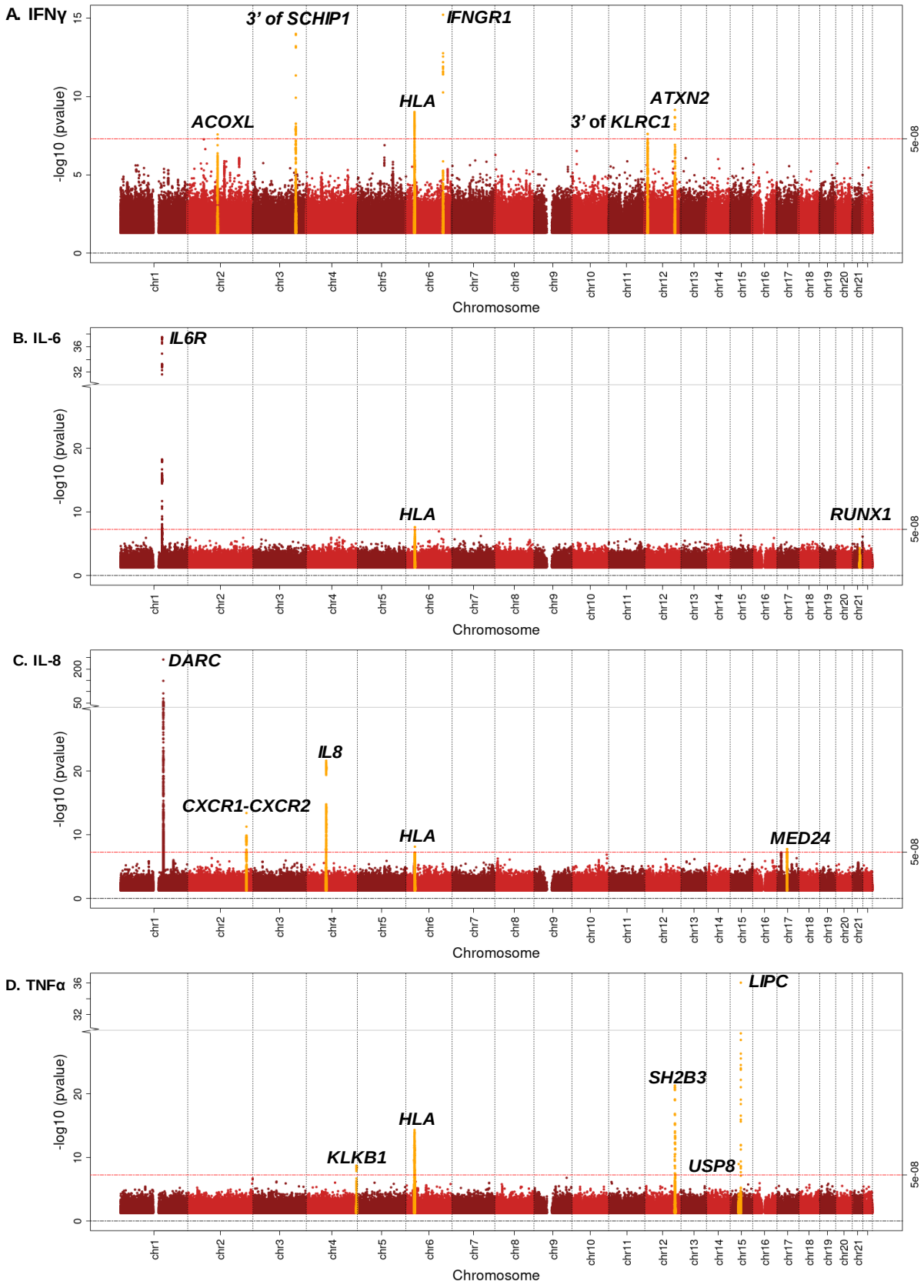


Figure 3.1: Manhattan plot for the meta-analysis of each cytokine. Relative base-pair positions of each variant are arranged by chromosome. The Y-axis depicts $\log_{10}$ P-values for each association with respective circulating cytokine levels. The red line denotes a genome-wide significance threshold of $P<5x10^{-8}$. The gene annotations for each lead variant from the distance-based pruning are shown. Novel signals are shown in orange.

Independent signal selection using formal conditional analysis

Independent variant selection identified 18 independent variants using a significance threshold of $P<1x10^{-5}$. **Table 3.2** shows 5 independent variants associated with IFNγ levels, 2 variants with IL-6 levels, 7 variants in 4 non-*HLA* loci with IL-8 levels and 4 variants with TNFα levels. Of the signals identified in this study, 16 were found to be novel associations with their corresponding cytokine as per the GWAS catalog[282]. Association estimates after having conditioned for other independent variants in the locus are presented.

Bayesian fine mapping of independent association signals

Bayesian fine mapping enabled the construction of credible sets that were highly likely to drive the observed association signals. Overall, the median number of variants and interquartile range (IQR) between the credible sets was 56 (4 – 135) variants. Median credible set size and IQR was 92,850 (8,221 – 247,015) base-pairs. In 13 of the 19 credible sets, the variant with the highest posterior probability was the same as the independent variant identified using conditional analysis (**Table 3.3**). Three 99% credible sets consisted of a single variant. At four loci a single variant accounted for more than 80% of the posterior probability, indicating high statistical evidence of causal association: Intronic variants in *IFNGR1* and *DARC* and missense variants in *LIPC* and *USP8*. In other credible sets, posterior probabilities were comparatively smaller making it challenging to determine the variant that was likely to be driving the association. This was reflected in comparatively larger window sizes and an increased number of variants within these credible sets. Instances where several common variants are in high LD make determination of the causal variant challenging, as in the case of rs4129267 in *IL6R* where the posterior probability of causality is split evenly across several variants in the region, resulting in lower posterior probabilities.

Table 3.2: Conditionally independent association signals for each cytokine													
Cytokine	**Locus**	**Selection Order [a]**	**Chr**	**Pos**	**Variant**	**EA**	**EAF**	**Beta[b]**	**SE[b]**	**P-value[b]**	**Meta-analysis P-value**	***HLA***	**Novel[d]**
IFNγ	1	1	6	137540335	rs11754268	T	0.21	0.10	0.019	$3.80x10^{-7}$	$6.08x10^{-16}$	N	Y
IFNγ	2	1	3	159637678	rs76830965	A	0.12	0.14	0.024	$3.63x10^{-9}$	$1.17x10^{-14}$	N	Y
IFNγ	3	1	12	112007756	rs653178	C	0.47	0.06	0.016	$7.32x10^{-5}$	$7.19x10^{-10}$	N	Y
IFNγ	4	N/A	6	31237255	rs2523609	A	0.62	0.07	0.01	$1.11X10^{-09}$	$1.11x10^{-09}$	Y	Y
IFNγ	5	1	12	10598426	rs4492899	A	0.31	0.06	0.017	0.001	$2.47x10^{-8}$	N	Y
IFNγ	6	1	2	111807677	rs2009581	G	0.73	0.07	0.018	$2x10^{-4}$	$2.63x10^{-8}$	N	Y
IL-6	1	1	1	154426264	rs4129267	T	0.41	0.11	0.016	$1.10x10^{-12}$	$3.52x10^{-38}$	N	N
IL-6	2	N/A	6	32578632	rs116457146	T	0.26	0.07	0.01	$2.20X10^{-08}$	$2.20x10^{-08}$	Y	Y
IL-6	3	1	21	36823802	rs545129521	T	0.01	0.75	0.173	$1.62x10^{-5}$	$4.86x10^{-8}$	N	Y
IL-8	1	1	1	159175354	rs12075	A	0.58	0.47	0.015	$1.25x10^{-193}$	$7.08x10^{-243}$	N	N[e]
IL-8	2	1	4	74567906	rs138840656	G	0.06	0.28	0.036	$1.37x10^{-14}$	$2.33x10^{-22}$	N	Y
IL-8	2	2	4	74636009	rs7689435	G	0.43	0.09	0.016	$2.35x10^{-8}$	$2.33x10^{-22}$	N	Y
IL-8	2	3	4	74675215	rs141761385	G	0.01	0.84	0.160	$1.61x10^{-7}$	$2.33x10^{-22}$	N	Y
IL-8	3	1	2	218999982	rs55799208	A	0.002	0.91	0.182	$5.62x10^{-7}$	$3.71x10^{-14}$	N	Y
IL-8	3	2	2	219020958	rs114050631	T	0.01	0.43	0.092	$3.47x10^{-6}$	$3.71x10^{-14}$	N	Y
IL-8	4	N/A	6	32634613	chr6:32634613:C>T	T	0.11	0.12	0.02	$7.17x10^{-9}$	$7.17x10^{-9}$	Y	Y
IL-8	5	1	17	38184580	rs6503533	T	0.63	0.07	0.016	$4.18x10^{-5}$	$1.67x10^{-8}$	N	Y
TNFα	1	1	15	58770523	rs7161799	T	0.08	0.27	0.030	$1.90x10^{-20}$	$9.00x10^{-37}$	N	Y
TNFα	3	N/A	6	31341411	rs28611443	C	0.92	0.15	0.02	$1.01x10^{-14}$	$1.01x10^{-14}$	Y	Y
TNFα	5	1	15	50784990	rs146125856	T	0.90	0.12	0.026	$1.52x10^{-6}$	$1.04x10^{-9}$	N	Y
TNFα	6	1	4	187163614	rs4253272	C	0.49	0.08	0.016	$1.20x10^{-7}$	$1.84x10^{-9}$	N	Y
TNFα	24[c]	1	12	111884608	rs3184504	T	0.47	0.10	0.016	$1.43x10^{-10}$	$4.72x10^{-22}$	N	Y

Abbreviations: IFNγ, Interferon-γ; IL-6, Interleukin-6; IL-8, Interleukin-8; TNFα, Tumour necrosis factor α; Chr, Chromosome; Pos, Position; EA, Effect allele; EAF, Effect allele frequency; SE, Standard error; N, No; Y, Yes

a. The order that variants were selected in within each locus, after conditioning on other variants in the same locus. *HLA* variants were not included in this analysis
b. Association statistics after conditioning
c. 10Mb locus created when loci 2 and 4 were merged into one
d. "Y" represents a novel association with levels of the corresponding cytokine in the GWAS catalog, "N" represents a previously identified signal
e. Has been previously reported at just below the genome-wide significance threshold

Table 3.3: Bayesian fine mapping results showing the lead variant of each 99% credible set and its corresponding posterior probability												
Cytokine	**Locus**	**Independent variant**	**Gene[a]**	**Haploreg Annotation[a]**	**Chr**	**Pos**	**Meta-analysis P-value**	**Variants in 99% credible set**	**Credible set window**	**Credible set lead variant**	**Credible set lead P-value**	**Credible set lead Posterior probability**
IFNγ	1	rs11754268	*IFNGR1*	Intronic	6	137540335	6.08×10^{-16}	2	35	rs11754268	6.08×10^{-16}	0.99
IFNγ	2	rs76830965	3' of *SCHIP1*	Intergenic	3	159637678	1.17×10^{-14}	4	34966	rs76830965	1.17×10^{-14}	0.44
IFNγ	3	rs653178	*ATXN2*	Intronic	12	112007756	7.19×10^{-10}	9	225769	rs653178	7.19×10^{-10}	0.37
IFNγ	5	rs4492899	3' of *KLRC1*	Intergenic	12	10598426	2.47×10^{-8}	92	93172	rs4492899	2.47×10^{-8}	0.12
IFNγ	6	rs2009581	*ACOXL*	Intronic	2	111807677	2.63×10^{-8}	135	247015	rs2009581	2.63×10^{-8}	0.39
IL-6	1	rs4129267	*IL6R*	Intronic	1	154426264	3.52×10^{-38}	11	8221	rs4129267	3.52×10^{-38}	0.16
IL-6	3	rs545129521	*RUNX1*	Intronic	21	36823802	4.86×10^{-8}	5,589	999854	rs545129521	4.86×10^{-8}	0.33
IL-8	1	rs12075	*DARC*	Missense	1	159175354	7.08×10^{-243}	1	0	rs12075	7.08×10^{-243}	0.99
IL-8	2	rs7689435	3' of *IL8*	Intergenic	4	74636009	2.35×10^{-8} [b]	80	97429	rs4279174	3.80×10^{-9}	0.06
IL-8	2	rs138840656	5' of *IL8*	Intergenic	4	74567906	1.37×10^{-14} [b]	92	45370	rs1951705	1.36×10^{-15}	0.11
IL-8	2	rs141761385	5' of *CXCL6*	Intergenic	4	74675215	1.61×10^{-7} [b]	4,455	999513	rs55648028	2.95×10^{-7}	0.03
IL-8	3	rs55799208	*CXCR2*	Missense	2	218999982	5.62×10^{-7} [b]	4,585	999742	rs1346768	3.38×10^{-4}	0.01
IL-8	3	rs114050631	3' of *CXCR1*	Intergenic	2	219020958	3.47×10^{-6} [b]	3,695	999273	rs1126580	1.16×10^{-5}	0.04
IL-8	5	rs6503533	*MED24*	Intronic	17	38184580	1.67×10^{-8}	135	92528	rs6503533	1.67×10^{-8}	0.04
TNFα	1	rs7161799	*LIPC*	Intronic	15	58770523	9×10^{-37}	1	0	rs7161799	9×10^{-37}	0.99
TNFα	5	rs146125856	*USP8*	Missense	15	50784990	1.04×10^{-9}	1	0	rs146125856	1.04×10^{-9}	0.99
TNFα	6	rs4253272	*KLKB1*	Intronic	4	187163614	1.84×10^{-9}	32	34744	rs4253272	1.84×10^{-9}	0.06
TNFα	24	rs3184504	*SH2B3*	Missense	12	111884608	4.72×10^{-22}	5	142707	rs3184504	4.72×10^{-22}	0.32

Abbreviations: IFNγ, Interferon-γ; IL-6, Interleukin-6; IL-8, Interleukin-8; TNFα, Tumour necrosis factor α; Chr, Chromosome; Pos, Position

a. Functional annotations from Variant Effect Predictor v96[285]

b. P-value conditional on other variants in the locus

Correlation and variant-based heritability across the four cytokines

Correlations between the four cytokines were measured on both a genetic and observational level. LDSC[287], was used to estimate the genetic correlations and is robust to sample overlap as this inflates the model intercept and not the regression slope which is used to estimate the correlation[291]. Correlation estimates between the four cytokines are shown in **Table 3.4**.

Table 3.4: Correlation estimates between the four cytokines

Cytokine 1	Cytokine 2	Genetic Correlation (SE)	P-value	Observational Correlation (SE)	P-value
IFNγ	IL-6	0.002 (0.2)	0.99	0.24 (0.008)	$< 2.2x10^{-16}$
IFNγ	IL-8	-0.15 (0.14)	0.26	0.12 (0.008)	$< 2.2x10^{-16}$
IFNγ	TNFα	0.39 (0.18)	0.03	0.34 (0.007)	$< 2.2x10^{-16}$
IL-6	IL-8	0.07 (0.16)	0.64	0.10 (0.008)	$< 2.2x10^{-16}$
IL-6	TNFα	0.38 (0.2)	0.06	0.26 (0.007)	$< 2.2x10^{-16}$
IL-8	TNFα	-0.14 (0.14)	0.32	0.24 (0.008)	$< 2.2x10^{-16}$

Abbreviations: IFNγ, Interferon-γ; IL-6, Interleukin-6; IL-8, Interleukin-8; TNFα, Tumour necrosis factor α; SE, Standard error

a. Genetic correlation from LDSC – shown with its corresponding SE and p-value
b. Pearson's correlation coefficient – estimated using both EPIC-Norfolk and Fenland

The four cytokines were significantly correlated with one another on an observational level, albeit with marginal estimates and were largely comparable to the genetic estimates. Although in the case of IFNγ and IL-8 levels, for example, directional inconsistencies were observed. Across both estimations, IFNγ levels were most correlated with TNFα levels, with both correlation estimates showing significance. IL-6 and TNFα levels were also correlated with one another, however, the genetic estimates were not significant despite having a higher estimate than the observational correlation.

The variant-based heritabilities for each cytokine were calculated as part of the LDSC regression (**Table 3.5**). The regression intercept (Intercept < 1.5) and genomic control lambda ($\lambda_{GC} < 1.1$) did not show evidence of test statistic inflation due to bias or population stratification and that any inflation is likely to be due to polygenic architecture (**Table 3.5**). Heritability estimates were in line with those of other biomarkers from the LD-hub database[288] such as LDL cholesterol from the global lipids consortium (GLGC): $h^2 = 0.13$ (0.02) and HbA1c from the meta-analysis of glucose and insulin-related traits consortium (MAGIC): $h^2 = 0.07$ (0.01).

Table 3.5: Narrow-sense heritability estimates from LDSC

Cytokine	h^2	SE	λ_{GC}	χ^2	Intercept	Number of variants
IFNγ	0.11	0.03	1.02	1.03	0.99	1,179,881
IL-6	0.06	0.02	1.02	1.02	0.99	1,179,857
IL-8	0.21	0.04	1.03	1.05	0.98	1,179,832
TNFα	0.09	0.03	1.03	1.04	1.00	1,179,873

Abbreviations: IFNγ, Interferon-γ; IL-6, Interleukin-6; IL-8, Interleukin-8; TNFα, Tumour necrosis factor α; h^2; Narrow-sense heritability; SE, Standard error; λ_{GC}, Genomic inflation factor; χ^2, Chi-squared statistic

Incremental variance explained by variants associated with cytokine levels

The proportion of variance explained for each variant associated with respective cytokine levels was estimated (**Figure 3.2**). Variants were ordered by their proportion of variance explained. Variants in *IFNGR1*, *IL6R*, *DARC* and *LIPC* explained the greatest proportion of variance in IFNγ, IL-6, IL-8 and TNFα levels respectively. In the case of *IL6R* and *DARC* variants, this was substantially greater than the remaining variants associated with the respective cytokine combined, 0.9% and 6.3% respectively.

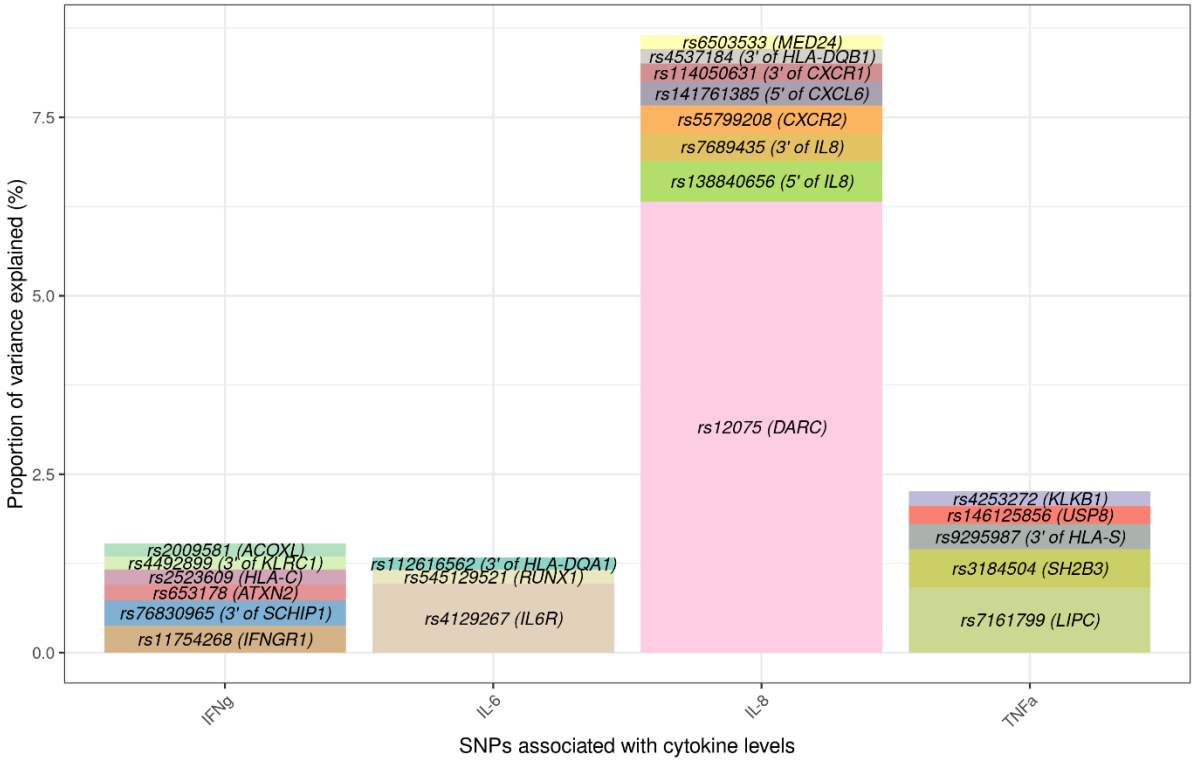


Figure 3.2: Stacked bar plot illustrating the incremental variance explained by each variant associated with cytokine levels. The four cytokines are shown along the X-axis and the proportion of variance explained by each variant in percent is shown on the Y-axis. Variants associated with each respective cytokine are ordered by their variance explained from largest (bottom of the bar) to smallest (top of the bar). Abbreviations: IFNγ, Interferon-γ; IL-6, Interleukin-6; IL-8, Interleukin-8; TNFα, Tumour necrosis factor α; SNP, Single nucleotide polymorphism.

Colocalisation of cytokine variants with cardiometabolic diseases

The genetic colocalisation between variants that explained the greatest proportion of variance in the respective cytokine levels and cardiometabolic diseases was estimated (**Figure 3.3**). T2D results for IFNγ, IL-6 and IL-8 robustly showed that cytokine and disease traits were driven by independent variants at the *IFNGR1, IL6R* and *DARC* loci, whereas only TNFα was estimated to have a significant association at the *LIPC* locus. Comparatively at the *IFNGR1*, *DARC* and *LIPC* loci, only IFNγ, IL-8 and TNFα were estimated to have significant association signals at the respective loci. In contrast, the *IL6R* locus showed robust evidence that IL-6 levels and CHD colocalise (posterior probability of colocalisation, PP_{coloc} = 0.98).

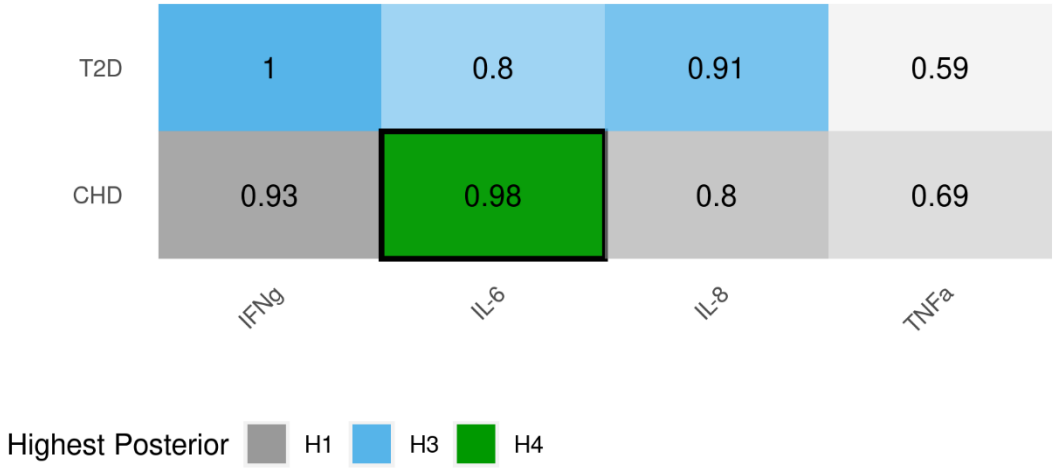


Figure 3.3: Heatmap of the pairwise colocalisation between cytokine levels and cardiometabolic diseases. A heatmap matrix depicting the largest pairwise colocalisation estimate between cytokine levels and both T2D and CHD. The 1Mb regions either side of the variant explaining the greatest proportion of variance in the levels of each cytokine was used. Each colocalisation hypothesis is coloured differently with the colour saturation showing the evidential strength. Trait-pairs with significant posterior probability estimates of colocalisation (H4 + H3 ≥ 0.9 & H4/H3 ≥ 3) were outlined in black. To discriminate between H1 and H2 hypotheses, traits along the X-axis were used as "Trait 1" in the analysis and traits listed on the Y-axis were used as "Trait 2". H0 – no signal; H1 – signal unique to trait 1; H2 – signal unique to trait 2; H3 – two independent causal variants in the same locus driving the association signal for the respective traits and H4 – presence of a causal variant shared between two traits.

Abbreviations: H, Hypothesis; CHD, Coronary heart disease; T2D, Type 2 diabetes; IFNγ, Interferon-γ; IL-6, Interleukin-6; IL-8, Interleukin-8; TNFα, Tumour necrosis factor α.

3.4: Discussion

In what is the largest GWAS to date for cytokine levels, I identified 22 variant-cytokine associations within 19 loci, 16 of which are novel associations with the corresponding cytokine and not in the *HLA* gene region. This study also provides the first estimates of heritability which are comparable to estimates for other biomarkers. Genetic correlations between the four cytokines were also estimated and are in line with observational findings.

Of the associations identified in this study, variants at the *USP8, LIPC, DARC* and *IFNGR1* loci were likely to be the causal variant for their respective association signals with TNFα (*USP8* and *LIPC*), IL-8 (*DARC*) and IFNγ (*IFNGR1*) levels based on fine-mapping. The rs146125856 variant induces a missense L776P substitution within the ubiquitin carboxyl-terminal hydrolase domain of *USP8*, responsible for the cleavage of conjugated ubiquitin chains. *USP8* is a deubiquitinating enzyme thought to function in the turnover of endosomal transmembrane proteins and T-cell antigen receptor signalling[292]. While other members of the *USP* protein family have been shown to inhibit TNFα signalling, however, the role of *USP8* is unclear[293]. The rs7161799 variant lies within an intronic region of the hepatic lipase C (*LIPC*) gene. Primarily expressed in the liver, LIPC functions as a triglyceride hydrolase that partially converts HDL to LDL and acts as a ligand for receptor-mediated lipoprotein absorption[294]. Previous work has shown that variants in *LIPC* are associated with HDL levels, however, contrasting effect directions have been reported[295–297]. However, none of the reported variants are in LD with rs7161799 ($R^2 < 0.05$) suggesting that this represents a separate pathway that impacts TNFα levels. TNFα has previously been shown to interfere and regulate mechanisms that function in hepatic lipid homeostasis[298,299] despite this, the impact of *LIPC* on TNFα expression is still unclear.

A missense variant, rs12075 (G42D), was found to be associated with IL-8 levels at the Duffy antigen chemokine receptor (*DARC*) locus. The association with lower IL-8 levels at this locus has been previously described[300]. Previous studies have described associations with rs12075 and MCP1 levels[227]. *DARC* encodes a non-specific, membrane bound chemokine reservoir that assimilates chemokines from inflamed tissue to maintain homeostasis and is central to monocyte and neutrophil infiltration[301–304]. The association of IL-8 levels in cardiovascular disease has long been established, however, contrasting results have been found regarding the direction of this association[305]. This difference may be due to the temporally distinct roles that IL-8 levels play in atherosclerosis[306] and the fact that large-scale clinical trials have not been conducted. IL-8 levels function as a chemoattractant for circulating leukocytes and neutrophils to the vessel wall which initiate the atherosclerotic process[306]. In the later stages, DARC on the surface of erythrocytes functions as a regulator of local inflammatory processes

and aids in the clearance of IL-8 from circulation as high levels of IL-8 have been found to be bound to erythrocytes after myocardial infarction[307]. These results therefore replicate previous findings of the association between *DARC* and IL-8 levels and may point towards DARC assimilation and release of IL-8 being a mediator of the temporal association between IL-8 levels and cardiovascular disease.

The results highlight several associations of cytokines with their respective cellular receptors at the *IFNGR1, IL6R* and *DARC/CXCR1/2* loci. The implications of this are intuitive, as variants affecting receptor expression or function will impact circulating levels of the respective cytokines. In addition, *IFNGR1, IL6R* and *DARC* explained the greatest proportion of variance in IFNγ, IL-6 and IL-8 levels respectively. The rs11754268 variant was found to lie within the promoter of *IFNGR1* in a transcription factor binding site for RNA polymerase II and other transcription factors[308]. This may influence IFNGR1 protein expression and thereby contribute to higher IFNγ levels as IFNγ is not internalised. variants in the *IFNGR1* gene have previously been found to be associated with higher BMI[309], however, the variant identified in this study is independent of this signal ($R^2 < 0.05$). At the *IL6R* locus, refinement of causality is challenging due to the LD structure and therefore posterior probabilities are shared between 11 variants. The lead variant rs4129267 is in high LD with rs2228145, described in **Chapter 1**. Considering the experimental evidence that rs2228145 is functional[310], it is likely to be the variant driving the observed association. However, this presents a challenge to interpreting the causal association between variants leading to higher cytokine levels and risk of cardiometabolic diseases. It may seem intuitive to assume that the cytokine-raising allele would reflect higher inflammation mediated by an increase in cytokine levels. However, in the case of the *IL6R* locus (and possibly other loci found here), the opposite occurs where the missense variant rs2228145 leads to increased IL-6 levels but lower IL-6 mediated inflammation due to lower signalling through the IL-6 receptor[310]. Thus, combining cytokine-raising alleles that reflect opposing effect directions on downstream inflammation into a combined instrument may render the MR approach difficult to interpret and even invalid in some instances. In these instances, it may be more appropriate to interpret results of individual loci separately.

IL-8 has three known cellular receptors namely CXCR1, CXCR2 and DARC/ACKR1[311]. This study identified independent associations of variants within or in close proximity to all three receptors with IL-8 levels. Of the identified variants, rs7689435 and rs138840656 have both been identified as expression quantitative trait loci (eQTL) that lead to significantly higher expression of the IL-8 gene (*CXCL8*) in blood[312]. C-X-C motif chemokine receptor 2 (*CXCR2*) is the receptor for all members of the IL-8 chemokine family and CXCR1 is a receptor for CXCL6-8[311]. CXCR2 mediates neutrophilic migration to sites of inflammation and many of the

angiogenic properties of IL-8[311]. Both CXCR1 and CXCR2 are G-protein coupled receptors, which make them attractive drug targets[311]. CXCR2 in particular, has been investigated as a possible therapeutic target for asthma and a number of related respiratory conditions including chronic obstructive pulmonary disease (COPD), clinical trials are ongoing[311]. The rare missense variant rs55799208 induces an R153H substitution in the second intracellular loop of CXCR2, an area crucial for G-protein interactions and receptor activation[313]. The rs114050631 variant downstream of CXCR1, the receptor for CXCL6-8[311], lies within a STAT3 transcription factor binding site and is an established eQTL for lower *CXCR1* expression in blood[312]. Both of these variants have previously been associated with lower counts of several white blood cell types[314]. Neither the IL-8 polygenic score nor the individual IL-8 raising variants were associated with either disease outcome. Given that these variants reside in biologically plausible loci (including *CXCL8*) and their associations with IL-8 levels are statistically robust, this result reduces the likelihood that IL-8 levels have a strong causal effect in cardio-metabolic disease. This furthers our knowledge of the involvement of IL-8 in cardio-metabolic disease as prior to this, increases in IL-8 levels have been demonstrated in CHD and T2D patients in isolated reports but the causal relevance of IL-8 to these diseases had yet to be assessed[224].

The pleiotropic *ATXN2/SH2B3* locus on chromosome 12q24.1 was found to be associated with both IFNγ and TNFα levels respectively. The lead variants within each gene, rs653178 and rs3184504 respectively, were found to be in high LD ($R^2 > 0.9$) and both were within the 99% credible set for each association signal. Ataxin-2 (ATXN2) plays a role in RNA processing and function and the formation of stress granules[315]. Under normal conditions, the SH2B adaptor protein 3 (SH2B3) protein negatively regulates TNFα signalling by modulating the activities of both PI3K and MAPK ERK1/2[316]. Interestingly, a recent study showed that SH2B3 regulates adipose tissue expansion and activation of innate lymphoid cells, together these roles function to reduce the risk of T1D[317]. Loss of SH2B3 in a murine model led to adipose inflammation and glucose intolerance, however, this intolerance was not linked with concurrent impaired insulin secretion which is the hallmark of T1D, instead insulin resistance was noted which is more typical of T2D[317]. The rs3184504 variant, associated with TNFα levels, induces a R262W missense substitution within the pleckstrin homology domain of the SH2B3 protein which is crucial for protein translocation to the cell membrane and regulation of intracellular signalling[318]. This missense substitution is not predicted to induce a loss-of-function phenotype by LoFtool[319] therefore it is unlikely that carriers of rs3184504 would exhibit the phenotypes associated with loss of SH2B3 but given that the variant induces an amino acid substitution in a key domain, carriers may exhibit differential protein activity as a result. Both variants are associated with a number of autoimmune diseases including celiac

disease[320,321] and Crohn's disease[322,323] and have also been associated with higher risk of both CHD[289,324] and T1D[325,326]. A similar impact on IFNγ levels has been described using *SH2B3*$^{-/-}$ knockout in murine CD8^{+} T cells which lead to higher IFNγ secretion[317,327]. These results taken together with the complex LD structure at the *ATXN2/SH2B3* locus (**Figure S3.1**) make the resolution of causality complex and illustrate some of the challenges in applying an MR framework using cytokine levels as risk factors. The association with CHD for both IFNγ and TNFα deserves two considerations. Firstly, the association at the *SH2B3/ATXN2* locus may reflect horizontal pleiotropy in which case the locus would be an invalid instrument for MR[71]. Secondly, only some of the mechanisms which act in pathways leading to higher levels of these cytokines may be causally implicated in CHD. Indeed, this locus is clearly highly pleiotropic and is associated with several anthropometric and lipid traits which contribute to cardiometabolic diseases via disparate pathways[76]. It is therefore likely that the *ATXN2/SH2B3* locus is associated with both TNFα and IFNγ levels due to shared biology which may manifest via a shared regulatory mechanism, however, the exact mechanism and causal relationship underlying this mechanism is still not understood.

There are several limitations to this study. Firstly, using an inverse-rank normal transformation of the cytokine levels limits the comparability of the association results of this study to those from previous studies that often use natural log transformations. However, this does facilitate direct comparison across the four cytokines as all effect estimates will be standardised to a 1 SD increase and therefore risk of cardiometabolic diseases can be compared across the four cytokines. Secondly, the Bayesian fine mapping done here assumes only a single causal variant which may in some cases be an incorrect assumption of the underlying genetic architecture at a given locus. Methods that allow for multiple causal variants at the same locus may have produced different results at loci where complex LD structure hindered the resolution of the association signal to a single causal variant[328,329]. Thirdly, several factors prohibited the use of MR to estimate the causal relevance of cytokine levels to disease. The relatively low number of variants found to be associated with cytokine levels, compared to other biomarker traits such as triglyceride levels[330], meant that such an approach would likely be underpowered and liable to bias by variants with large effects on the trait of interest. In addition, loci such as *ATXN2-SH2B3* are established pleiotropic loci, invalidated one of the assumptions of MR as the locus is known to influence multiple traits through discrete pathways, characteristic of horizontal pleiotropy[62]. This study also identified associations between variants in cytokine receptors and levels of the respective cytokine. While variants such as rs2228145 in *IL6R* may not necessarily invalidate the assumptions of MR, they may hinder the interpretation of results through seemingly inverse associations between cytokine levels and disease, due to a lack of receptor mediated inflammatory signalling. Finally, this

study found that variants in cytokine receptors explain substantially greater proportion of variance in their respective cytokine levels compared to other variants. Colocalisation between these variants and cardiometabolic diseases may warrant further investigation on a single locus basis, however, for the majority of cytokines no colocalisation was observed with cardiometabolic diseases. The *IL6R* locus was the only region where robust colocalisation with cardiometabolic diseases was estimated, the relevance of this to cardiometabolic diseases will be investigated in detail in the following chapter.

Chapter 4: Meta-analysis Investigating the Role of Interleukin-6 Mediated Inflammation in Type 2 Diabetes

Contributions and collaborations

I designed the study, analysed the data and wrote the first draft of the report. Claudia Langenberg, Nicholas J. Wareham and Luca A. Lotta supervised the study. Rupal L. Shah and I performed the literature search and reviewed the identified articles. Stephen J. Sharp provided statistical support. I collated the data and performed the data analysis with assistance from Luca A. Lotta, Jian'an Luan, Isobel D. Stewart, and Eleanor Wheeler. Manuel A. R. Ferreira and Aris Baras provided data from the MyCode Study from the DiscovEHR Collaboration between Regeneron Genetics Center and Geisinger Health System.

Publications related to this chapter

Nicholas Bowker; Rupal L. Shah; Stephen J. Sharp; Jian'an Luan; Isobel D. Stewart; Eleanor Wheeler; Regeneron Genetics Center; DiscovEHR Collaboration; Manuel A. R. Ferreira; Aris Baras; Nicholas J. Wareham; Claudia Langenberg; Luca A. Lotta. **Meta-analysis investigating the role of interleukin-6 mediated inflammation in type 2 diabetes**. *Published at EBioMedicine.* 2020; 61: 103062

Abstract

Background: Evidence from animal models and observational epidemiology points to a role for chronic inflammation, in which interleukin 6 (IL-6) is a key player, in the pathophysiology of type 2 diabetes (T2D). However, it is unknown whether IL-6 mediated inflammation is implicated in the pathophysiology of T2D.

Methods: I performed a meta-analysis of 15 prospective studies to investigate associations between IL-6 levels and incident T2D including 5,421 cases and 31,562 non-cases. I also estimated the association of a loss-of-function missense variant (Asp358Ala) in the IL-6 receptor gene (*IL6R*), previously shown to mimic the effects of IL-6R inhibition, in a large trans-ethnic meta-analysis of six T2D case-control studies including 260,614 cases and 1,350,640 controls.

Findings: In a meta-analysis of 15 prospective studies, higher levels of IL-6 (per log pg/mL) were significantly associated with a higher risk of incident T2D (1·24 95% CI, 1·17, 1·32; $P=1x10^{-12}$). In a trans-ethnic meta-analysis of 260,614 cases and 1,350,640 controls, the *IL6R* Asp358Ala missense variant was associated with lower odds of T2D (OR, 0·98; 95% CI, 0·97, 0·99; $P=2x10^{-7}$). This association was not due to diagnostic misclassification and was consistent across ethnic groups. IL-6 levels mediated up to 5% of the association between higher body mass index and T2D.

Interpretation: Large-scale human prospective and genetic data provide evidence that IL-6 mediated inflammation is implicated in the aetiology of T2D but suggest that the impact of this pathway on disease risk in the general population is likely to be small.

Funding: The EPIC-Norfolk study has received funding from the Medical Research Council (MRC) (MR/N003284/1, MC-UU_12015/1 and MC_PC_13048) and Cancer Research UK (C864/A14136). The Fenland Study is funded by the MRC (MC_UU_12015/1 and MC_PC_13046).

4.1: Introduction

Chronic inflammation, which is partly mediated via the interleukin 6 (IL-6) pathway, has been hypothesized to play a critical role in the pathophysiology of type 2 diabetes (T2D)[112,123]. In obesity, the enhanced production of IL-6 and other pro-inflammatory mediators as a result of higher BMI is associated with insulin resistance[138,139,171] and with a higher risk of T2D[154,186] and coronary heart disease[189]. In addition, evidence from animal models and observational epidemiology has linked chronic inflammation with metabolic disease risk factors including disruption of lipid metabolism,[138,223] hypertension[331–333] and impaired insulin secretion[123,139,270,334]. On this basis, it has been hypothesised that therapeutically targeting inflammatory pathways may reduce the risk of T2D or improve glycaemic control in people with diabetes[9,112,186,335].

While small scale clinical trials had provided initial supportive evidence of possible benefit,[123,218] this notion has been strongly challenged by the negative findings of the Canakinumab™ Anti-inflammatory Thrombosis Outcome Study[226,336] (CANTOS). In a pre-specified secondary analysis of this trial of 10,061 people with previous myocardial infarction and high CRP, treatment with the interleukin-1β inhibitor Canakinumab™ substantially reduced circulating inflammatory markers, but did not significantly reduce risk of new-onset diabetes in participants without diabetes at recruitment, or improve glycaemia in participants with diabetes[221].

However, the possible therapeutic benefit on diabetes risk of directly inhibiting the IL-6 pathway has not been directly evaluated in large trials. In cardiovascular medicine, biomarker studies showing associations of higher IL-6 levels with higher cardiovascular disease incidence[89,90,189,190] and genetic studies[88,160,337] showing robust associations of an experimentally-validated[310] partial loss-of-function variant in the IL-6 receptor gene with protection against coronary disease have provided the evidence base for ongoing clinical trials that are evaluating the effects of IL-6 receptor inhibition in myocardial infarction.

In this study, biomarker and human genetic data from large-scale population-based cohorts were combined to investigate the possible etiologic role of IL-6 and IL-6 receptor (IL-6R) related pathways in T2D risk.

4.2: Methods

Study design

Three sets of analyses were used to investigate the relationships between the IL-6 pathway and T2D (**Table 4.1 and Figure 4.1**). In stage 1, the associations between IL-6 levels, glycaemic and anthropometric traits, and risk of incident T2D were estimated in two population-based cohort studies and a meta-analysis of 15 prospective studies (**Figure 4.1A, Table 4.1 and Table S4.1**). In stage 2, the association of a missense variant (Asp358Ala) in the interleukin 6 receptor gene (*IL6R*) with T2D was estimated in a trans-ethnic meta-analysis of six case-control studies (**Figure 4.1B, Table 4.1 and Table S4.1**). In stage 3, the proportion of the association between body mass index (BMI) and risk of cardio-metabolic disease outcomes mediated by IL-6 levels was estimated in three population-based cohort studies (**Figure S4.1, Table 4.1 and Table S4.1**). All studies were approved by local institutional review boards and ethics committees and participants gave written informed consent.

Table 4.1: Summary of the study design.

Stage and Aim	Independent Variables	Primary Outcome Variables	Outcome Data Sources	Outcome measure and statistical significance
Stage 1: Estimate the association of IL-6 levels with incident type 2 diabetes	IL-6 levels	Type 2 diabetes	EPIC-Norfolk (individual-level data); 14 prospective studies (summary statistics)	Risk ratio and 95% CI P < .05
Stage 2: Estimate the association of *IL6R* Asp358Ala with type 2 diabetes	Genotypes of Asp358Ala	Type 2 diabetes	UK Biobank (individual-level data); DIAMANTE[22], MVP[337], Finngen[338], Geisinger[339], Suzuki *et al.*[340] (summary statistics)	Odds ratio and 95% CI P < .05
Stage 3: Estimate the proportion of the associations between BMI and cardiometabolic diseases and BMI and fasting insulin that are mediated by IL-6	Independent variable: BMI (observational); polygenic score for BMI (genetic) Mediator: IL-6 levels	Type 2 diabetes, coronary heart disease and fasting insulin levels	EPIC-Norfolk, UK Biobank, Fenland (individual-level data) and MAGIC (summary data)	Mediation percentage and 95% CI (%)

Abbreviations: IL-6, interleukin 6; EPIC, European prospective investigation into cancer and nutrition; DIAMANTE; Diabetes meta-analysis of trans-ethnic association studies; MVP, Million veteran program; CI, confidence interval
Data sources for secondary analyses are summarised in **Table S4.1**.

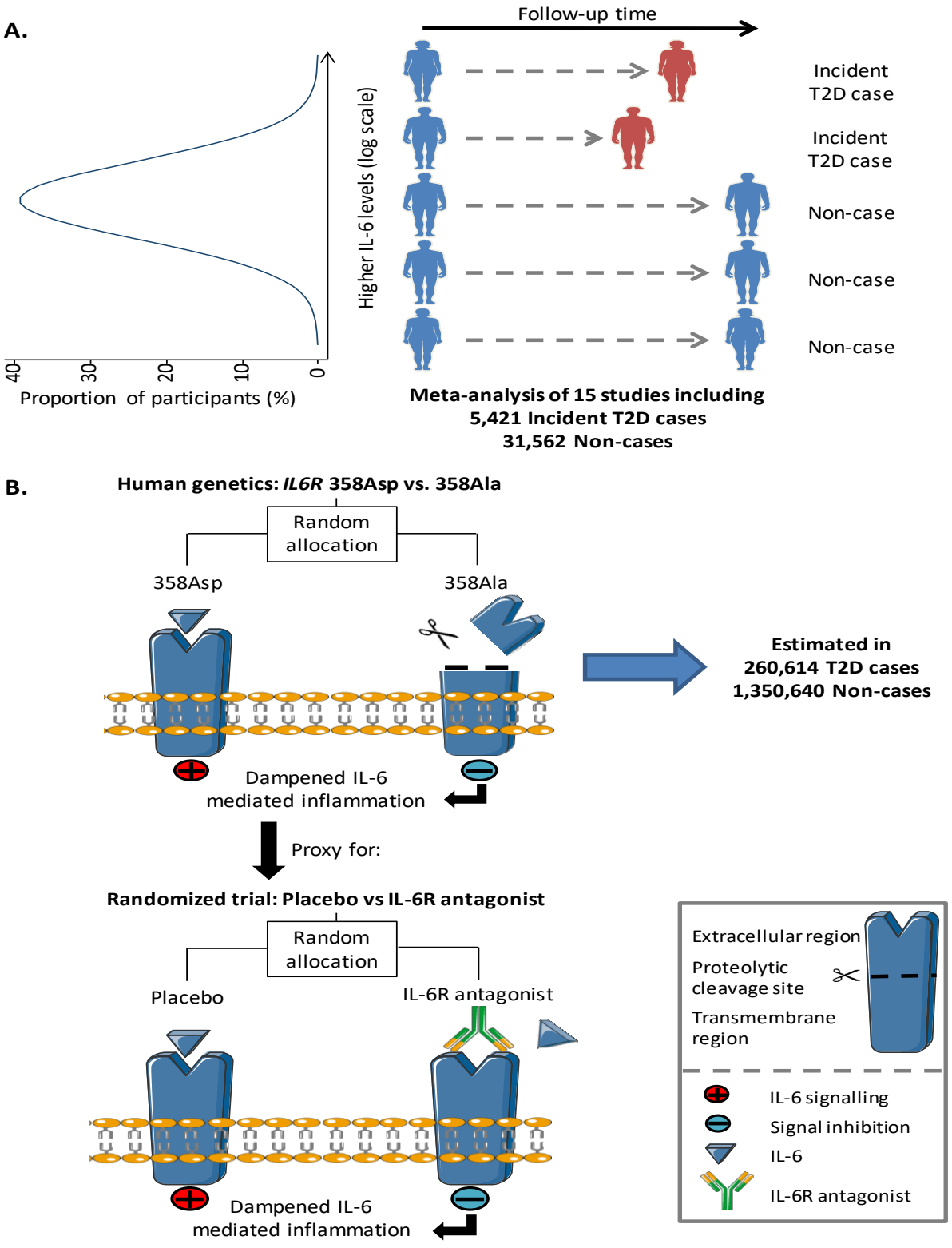


Figure 4.1: Study design. Panel A. Prospective association between IL-6 levels and risk of incident type 2 diabetes. **Panel B.** Mechanism by which IL6R Asp358Ala leads to dampened IL-6 mediated inflammation. Asp358Ala increases the proteolytic cleavage of the IL-6R, depicted with a pair of scissors, resulting in dampened classical IL-6 signalling. This is comparable to the effects of an IL-6R antagonist, shown at the bottom of the panel. Abbreviations: T2D, type 2 diabetes; IL-6, interleukin-6; IL-6R, interleukin 6 receptor (protein); Asp, aspartic acid; Ala, alanine; gp130, Glycoprotein 130.Modified from Servier Medical Art, licensed under a Creative Commons Attribution 3.0 Generic License from http://smart.servier.com.

Studies and participants

Participants from EPIC-Norfolk[241] and Fenland[244] are described in **Chapter 2** and outlined in **Table 1**. UK Biobank[66] (data collection: 2006-2019; **Table 1**) is a population-based cohort study of individuals recruited from 22 rural and urban recruitment centres in the United Kingdom. European ancestry participants with available genome-wide genotyping and phenotypic data were included in this study. Ethical approval for the UK Biobank study was given by the North West - Haydock Research Ethics Committee (16/NW/0274). This research was conducted using application 44448. Participants gave their electronic consent to use their anonymised data and samples for health-related research, to be re-contacted for further sub-studies, and for access to their health-related records.

The association of IL-6 levels with incident T2D was estimated in a meta-analysis of prospective studies including EPIC-Norfolk[241] and 14 other studies[186,225,239,240] (data collection: 1998-2018) as described in **Methods S4.1 and Tables S1-S4**. Estimates from EPIC-Norfolk were generated as part of this study and contributed 8% of the total incident cases and 22% of the non-cases to the analysis. The association of *IL6R* Asp358Ala with prevalent T2D was estimated in a trans-ethnic meta-analysis of results from individual-level data in UK Biobank, and summary-level data from the European ancestry analysis of the Diabetes Meta-analysis of Trans-ethnic Association Studies[22] (DIAMANTE; data collection: 2018-2019; **Table S4.1**), Million Veteran Program[337,341] (MVP; data collection: 2011-2019; **Table S4.1**), Finngen[338] (Finngen; data collection: 1996-2016; **Table S4.1**), MyCode Study from the DiscovEHR Collaboration between Regeneron Genetics Center and Geisinger Health System[339,342] (Geisinger; data collection: 2007-2020; **Table S4.1**) and a recent Japanese meta-analysis[340] (Suzuki *et al*.; data collection: 2003-2012; **Table S4.1**). The association of *IL6R* Asp358Ala with prevalent type 1 diabetes was estimated in a meta-analysis of results from individual-level data in UK Biobank, and summary-level data from the Type 1 Diabetes Genetics consortium[326] (T1DGC; data collection: 2004-2019; **Table S4.1**) and MVP[341] (data collection: 2011-2019; **Table S4.1**). Summary-level results from 10 previously published genome-wide association studies were used in different analyses of the study (**Table S4.1**).These included associations with BMI, BMI-adjusted waist-to-hip ratio (WHR), unadjusted WHR, waist- and hip-circumference from a meta-analysis of Genetic Investigation of Anthropometric Traits[343] (GIANT) consortium and UK Biobank[30], associations with fasting glucose, glucose at 2 hours after an oral glucose challenge and fasting insulin from the Meta-analyses of Glucose and Insulin-related Traits consortium[344–346] (MAGIC).The association with HbA1c was estimated in a meta-analysis of UK Biobank[66] and MAGIC[347] (**Table S4.1**).The association with non-fasting glucose was estimated in UK Biobank[66].

Exposure and outcome variables

In stage 1, the exposure variable was the circulating level of IL-6 (log-transformed and expressed in log-pg/mL), measured as described in **Chapter 2**. Log-transformed IL-6 measures were preferred to inverse-rank normal transformed measures as in **Chapter 2** as (1) I aimed to meta-analyse the EPIC-Norfolk estimates of the association between IL-6 levels and T2D with estimates previously reported in log-pg/mL, and (2) in **Chapter 2** I aimed to compare association estimates across cytokines, whereas in this analysis only IL-6 levels were of interest.

The outcome variable used in the meta-analysis was incident T2D from the EPIC-Norfolk study. Fasting glucose, glucose at 2 hours after an oral glucose challenge, HbA1c, fasting insulin, waist-to-hip ratio (WHR), BMI, iron, transferrin, ferritin, body fat percentage, abdominal fat mass, gluteofemoral fat mass, leg fat mass, abdominal to gluteofemoral fat mass ratio, peripheral fat mass, visceral fat mass and subcutaneous fat mass from the Fenland study were used in observational analyses. Incident T2D was defined as described in **Chapter 2**. Fasting glucose, two-hour glucose, fasting insulin, HbA1c, weight, height, BMI, waist and hip circumferences, WHR, and DEXA overall and compartmental fat masses were measured as described in **Chapter 2** (**Table S4.5**). Iron (in µmol/L), transferrin (in g/L) and ferritin (in µg/L) were measured using the Dimension RxL Integrated Chemistry System (Siemens, Germany).

In stage 2, the exposure variable was the genotype of rs2228145 (HUGO Gene Nomenclature Committee gene name, *IL6R*; transcript change, NCBI transcript identifier NM_000565.3 c.1073A>C; protein change, Asp358Ala). The variant 358Ala protein (encoded by the minor allele C; allele frequency in European ancestry participants in the 1000 Genomes Project, 36%) has been experimentally shown to be more susceptible to proteolytic cleavage by ADAM-proteases[348], resulting in enhanced shedding of soluble IL-6R, reduced expression of the membrane-bound form of the receptor[310] and impaired cellular responsiveness to IL-6[310]. On this basis, the variant allele has been hypothesized to impair signalling of the IL-6 receptor via its classical pathway in a way that mimics the effects of IL-6R inhibition using therapeutic monoclonal antibodies approved for the treatment of autoimmune conditions[88,160]. The pattern of association of this variant with a variety of intermediate traits and outcomes is highly consistent with the effects of IL-6R inhibitory drugs on those same traits in randomized clinical trials[88,160]. Hence, this variant has been used in several previous studies as a genetic instrument to study the likely consequences of pharmacological IL-6R inhibition. In UK Biobank, the variant genotype was obtained via imputation to the Haplotype Reference Consortium v1.1,[271] UK10K[273] and 1000 genomes phase 3[349] reference panels as previously

described. The imputation quality score was 1 (i.e. the maximum possible score), indicating excellent imputation quality comparable with direct genotyping.

The outcome variables were T2D in the primary analysis and type 1 diabetes, fasting glucose, 2-hour glucose, non-fasted glucose, fasting insulin, HbA1c, BMI, WHR, BMI-adjusted WHR, waist and hip circumference in secondary analyses. In UK Biobank, binary definitions of prevalent disease and a case-control analytical design were used as previously described[30]. Participants were classified as cases of T2D if they had: (1) self-reported T2D diagnosis or (2) self-reported use of oral anti-diabetic medications at nurse interview or at digital questionnaire, or (3) an electronic health record consistent with T2D (ICD-10 code E11). Participants were classified as cases of type 1 diabetes if they had: (1) self-reported type 1 diabetes diagnosis or (2) self-reported use of insulin but not oral anti-diabetic medications at nurse interview or at digital questionnaire, or (3) an electronic health record consistent with type 1 diabetes (ICD-10 code E10). Participants with evidence of diabetes but for whom the type of diabetes was unclear (e.g. participants reporting T2D but treated with insulin only or participants with electronic health records of both type 1 and T2D, code E10 in some electronic records but E11 in other records) were excluded. Controls were participants who (1) did not self-report a diagnosis of diabetes of any type, and (2) did not take any diabetes medications, and (3) did not have an electronic health record of diabetes of any type. Summary results for the association with glycaemic and anthropometric traits were extracted from the results of previous genome-wide association studies (**Methods S4.2, Table S4.1 and Table S4.6**).

In stage 3, the exposure variables were BMI and a polygenic score for higher BMI. The latter was derived using 97 BMI-associated common variants from a previous genome-wide association study[309] (**Table S4.7**). For each individual, the polygenic score was computed by adding the number of copies of each contributing genetic variant weighted by its association estimate in kg/m^2 units of BMI per allele.

The outcome variables were T2D, coronary heart disease and fasting insulin levels. T2D was defined as described above. In EPIC-Norfolk, incident coronary heart disease was defined as new-onset coronary heart disease in a participant without coronary heart disease at the time of IL-6 measurement based on an electronic health record consistent with ischemic heart disease (ICD-10 codes I20-25). Participants with coronary heart disease at the time of IL-6 measurement were excluded. In UK Biobank, prevalent coronary heart disease was defined as either (1) myocardial infarction or coronary disease documented in the participant's medical history at the time of enrolment by a trained nurse or (2) an electronic health record of acute myocardial infarction or its complications (ICD-10 codes I21-I23). Controls were participants

who did not meet any of these criteria. Summary-level results from MAGIC[346] were used for fasting insulin levels.

The mediator variable was the level of IL-6 in plasma measured by electrochemiluminescence as described above.

Statistical analysis

In stage 1, Cox regression models were used to estimate the association between IL-6 levels and T2D, as described in **Chapter 2**. Incident rates were calculated as number of incident cases divided by the total person-years of follow-up in people at risk. Absolute rate differences were obtained for each decile of the IL-6 distribution in EPIC-Norfolk relative to bottom decile (i.e. lowest IL-6 level) and 95% confidence intervals were calculated as ± 1.96 * √(number of incident cases in decile/(total follow-up in decile)^2 + number of incident cases in bottom quantile/(total follow-up in bottom decile^2)). The interaction between IL-6 levels and BMI on risk of incident type 2 diabetes was estimated by including an interaction term for BMI in the Cox regression model and adjusting for covariates as before, with the exclusion of BMI as a covariate. Estimates of association from the EPIC-Norfolk study and other 14 prospective studies[186,225,239,240] were pooled using inverse-variance weighted fixed-effect meta-analysis, making the assumption that the study-specific estimates of association all approximated the relative risk. Linear regression models were used to estimate associations between IL-6 and glycaemic traits. Models were adjusted using three methods: (1) for age and sex, (2) for age, sex and DEXA-derived detailed anthropometric variables, (3) for age, sex, DEXA variables and iron variables (iron, ferritin and transferrin) to mitigate the effects of IL-6 levels on hepcidin and iron metabolism.

In stage 2, the association between *IL6R* Asp358Ala and IL-6 levels in Fenland and EPIC-Norfolk were estimated using linear regression adjusting for age, sex, genotyping platform and the first 10 genetic principal components. The associations of *IL6R* Asp358Ala in UK Biobank were estimated using linear (continuous outcomes) or logistic (binary outcomes) regression. To minimize genetic confounding, association analyses were restricted to European ancestry individuals, identified by combining k-means clustering of genetic principal components with self-reported ancestry. To control for relatedness, analyses were either clustered using family structure data (third degree relatives) and adjusted for 40 genetic principal components or performed using linear mixed-effects models adjusting for a genomic kinship matrix. All analyses were adjusted for age, sex and genotyping array. Association estimates from UK Biobank and other studies were pooled using inverse-variance weighted fixed-effect meta-analysis.

Various secondary and sensitivity analyses, described below, were conducted to provide context for the association of *IL6R* Asp358Ala with T2D. In UK Biobank, the proportion of variance in T2D and coronary heart disease explained by the *IL6R* Asp358Ala variant was estimated as the difference in R^2 between a regression model including Asp358Ala and covariates (age, sex, genotyping array and genetic principal components) and one including only the covariates. The heritability explained by the variant was estimated by rescaling the variance explained by the heritability in T2D and coronary heart disease estimated using linkage disequilibrium score regression[287] (**Methods S4.2**). The association of Asp358Ala with T2D was also estimated while conditioning on an independent association signal upstream of Asp358Ala.

To assess whether the association with T2D might be influenced by misclassification of cases of type 1 diabetes (i.e.an etiologic subtype of diabetes where immune pathways have an established causal role) as cases of T2D, the association of *IL6R* Asp358Ala with T2D was tested in individuals below the median value of a polygenic score for type 1 diabetes[350]. This approach has been shown to enable the exclusion of over 95% type 1 diabetes cases present in a dataset (sensitivity of values above the median to detect type 1 diabetes, 96%; false negative rate [i.e. type 1 diabetes cases left in individuals below the median], 4%)[350]. Simulations revealed that, when using this approach, only 46 (0.4%) misclassified cases out of over 11,000 cases of T2D would remain in the analysis, even in the extreme scenario that 100% of type 1 diabetes cases in UK Biobank were misclassified as T2D (**Methods S4.2**).

To assess whether genetic predisposition to lower IL-6R signalling due to *IL6R* Asp358Ala genotypes interact with genetic predisposition to known risk factors for T2D, the association of *IL6R* Asp358Ala with T2D was estimated in individuals above or below the median value for 10 polygenic scores capturing genetic predisposition to higher fasting glucose[345,346], fasting insulin[345,346], HbA1c via glycaemic-related mechanisms[347], insulin resistance[345,351], impaired insulin secretion[351], BMI[343], BMI-adjusted WHR[30], BMI-adjusted WHR via lower gluteofemoral fat[30], BMI-adjusted WHR via higher abdominal fat[30] and lower BMI via a β-arrestin biased gain-of-function variants in *MC4R*[352] (**Table S4.7**). The analysis was then repeated using HbA1c levels in UK Biobank as the outcome and a polygenic score for higher BMI as the interaction variable.

While both experimental[310] and genetic[88,160] association studies have shown that rs2228145-C (358Ala) has phenotypic associations that mimic the effects of IL-6R inhibitory therapy, an association of this variant with lower risk of T2D does not necessarily mean that IL-6R

inhibitory therapy will produce clinically meaningful reduction in the risk of T2D. Even when assuming that the effects of a genetic variant and that of a drug on the target are qualitatively the same, two differences remain. The first is that genetic variants are usually associated with small differences in the activity of the target gene, while drugs usually have large effects on target activity. The second is that the small differences in target gene activity associated with a genetic variant are lifelong, while the effect of drugs is usually assessed in trials of short duration. Statistical approaches have been proposed to model these differences and formulate approximate projections of a range of possible effects of drug treatment on the basis of the magnitude of genetic associations[222,353]. In this study, a projected range of possible effects for IL-6R blocking therapy on incident T2D in a primary prevention setting were formulated using (a) the genetic association of *IL6R* Asp358Ala with diabetes in this study, (b) the genetic association of *IL6R* Asp358Ala with C-reactive protein (used as biomarker of IL-6R mediated inflammation) in this and other studies[88] and (c) the effects of IL-6R blocking therapy on C-reactive protein (the most-widely used biomarker of target engagement for IL-6R blocking therapy) in randomized clinical trials[160]. Assumptions and details of the methods employed in these calculations are in **Methods S4.2**.

In stage 3, two mediation models were utilized (**Figure S4.1**). The first estimated the proportion of the association between measured BMI and cardio-metabolic disease outcomes mediated by IL-6 levels (observational mediation). The second estimated the proportion of the association between genetic predisposition to higher BMI and cardio-metabolic disease outcomes mediated by IL-6 levels (genetic mediation). The genetic analysis was also repeated using fasting insulin levels as the outcome. As genetic associations are robust to non-genetic forms of confounding, I hypothesized that the second model would be more conservative in estimating the role of IL-6 levels as a possible mediator.

The proportion of association mediated by IL-6 was estimated using the Baron-Kenny[354] method. To obtain a 95% confidence interval for this proportion, 80% of individuals from each sample population used in each step of the mediation analysis were randomly sampled and the respective analyses were performed with 1000 iterations per arm of the framework. Sampled individuals were replaced prior to the next iteration. Mediation estimates were calculated for the 1000 iterations and the 2.5th and 97.5th percentile mediation estimates respectively were used as the lower and upper bounds of the 95% confidence interval.
P-values <.05, or the Bonferroni correction corresponding to P<.05 for each analysis of the study with multiple exposures or outcomes, were considered statistically significant. All reported P-values were from 2-tailed statistical tests. Statistical analyses were performed

using STATA v14.2 (StataCorp, College Station, Texas 77845 USA), R v3.2.2 (The R Foundation for Statistical Computing), BOLT-LMM v2.3.2[355] and METAL v2011-03-25[276].

4.3: Results

IL-6 levels, anthropometric and glycaemic measures and risk of T2D

Among 7,421 participants of the EPIC-Norfolk cohort, median levels of IL-6 were 0.59 pg/mL (interquartile range [IQR], 0.42, 0.85 pg/mL; **Table 4.2**).

Table 4.2: Study participants

Study	Fenland	EPIC-Norfolk	UK Biobank
Participants, N	10,344	7,421	447,491
Age at baseline, mean years (SD)	48 (8)	62 (9)	57 (8)
Women, N (%)	5,528 (53)	4,536 (61)	243,114 (54)
Men, N (%)	4,816 (47)	2,885 (39)	204,377 (46)
Current smokers, N (%)	1,236 (12)	574 (8)	46,459 (10)
BMI in kg/m^2, mean (SD)	26.7 (4.5)	26.4 (3.8)	27.4 (4.7)
Waist-to-hip ratio, mean (SD)	0.88 (0.09)	0.84 (0.09)	0.87 (0.09)
Systolic blood pressure in mmHg, mean (SD)	122 (15)	134 (18)	138 (19)
Diastolic blood pressure in mmHg, mean (SD)	74 (10)	81 (11)	82 (10)
IL-6 in pg/mL, median (IQR)	0.51 (0.34, 0.74)	0.59 (0.42, 0.85)	N/A
IL-6 in log-pg/mL, mean (SD)	-0.77 (0.94)	-0.55 (0.83)	N/A

In UK Biobank genotyping was performed using the Affymetrix 500K and Affymetrix UK BiLEVE genotyping chips and imputation was performed using the Haplotype Reference Consortium v1.1, UK10K and 1000 Genomes phase 3 reference panels. IL-6 levels were not measured in UK Biobank.
Abbreviations: N/A, not available; N, number of participants; SD, standard deviation; BMI, body mass index; IL-6, Interleukin-6; IQR, Interquartile range

A total of 411 incident cases of T2D occurred over 122,115 person-years of follow-up (incidence rate, 3.4 per 1000 person-years). Higher levels of IL-6 were associated with higher incidence of T2D (hazard ratio [HR] per log-pg/mL higher IL-6 levels, 1.25; 95% CI, 1.10, 1.41; P<.001), with participants in the top decile of IL-6 levels having approximately 2-fold higher hazard compared to the bottom decile (HR, 2.14; 95% CI, 1.21, 3.81; P=.009; absolute rate difference in incident cases per 1000 person-years, 4.2; 95% CI, 2.8, 5.7; **Figure S4.3**). The difference between these results and those presented in **Chapter 2** is that these are natural log transformed, whereas those in **Chapter 2** were inverse-rank normal transformed to enable comparison across cytokines. For comparison, participants in the top decile of BMI had approximately 6-fold higher hazard compared to those in the bottom decile (HR, 5.64; 95% CI, 2.77, 11.49; P=2x10^{-6}; **Figure S4.4**). No evidence of an interaction between IL-6 levels and BMI was found on incident type 2 diabetes risk ($P_{interaction}$=0.86). A meta-analysis of 15 prospective studies including 5,421 incident cases and 31,562 non-cases showed a robust association between higher IL-6 levels and higher incident T2D, with a pooled hazard ratio that was approximately the same as that of the EPIC-Norfolk analysis (**Figure 4.2A**). In sensitivity analyses, no evidence of publication bias or bias induced via differing IL-6 quantitation methods used between studies was found (P_{Egger} = 0.44; **Methods S4.1**). In

10,344 participants of the Fenland cohort, IL-6 levels were positively correlated with DEXA-measures of overall adiposity and abdominal fat distribution, which accounted for 9% of the variance in IL-6 levels (**Table S4.8**). After extensive adjustment for adiposity measures, higher IL-6 levels remained associated with higher HbA1c levels, but not with fasting glucose, 2-hour glucose or fasting insulin (**Figures 4.2B and 4.2C**). This remained after further adjustment for iron-related traits to mitigate the effects of IL-6 on hepcidin and iron metabolism (beta in % per log-pg/mL higher IL-6 levels, 0.04; 95% CI, 0.03, 0.05; P=<.001).

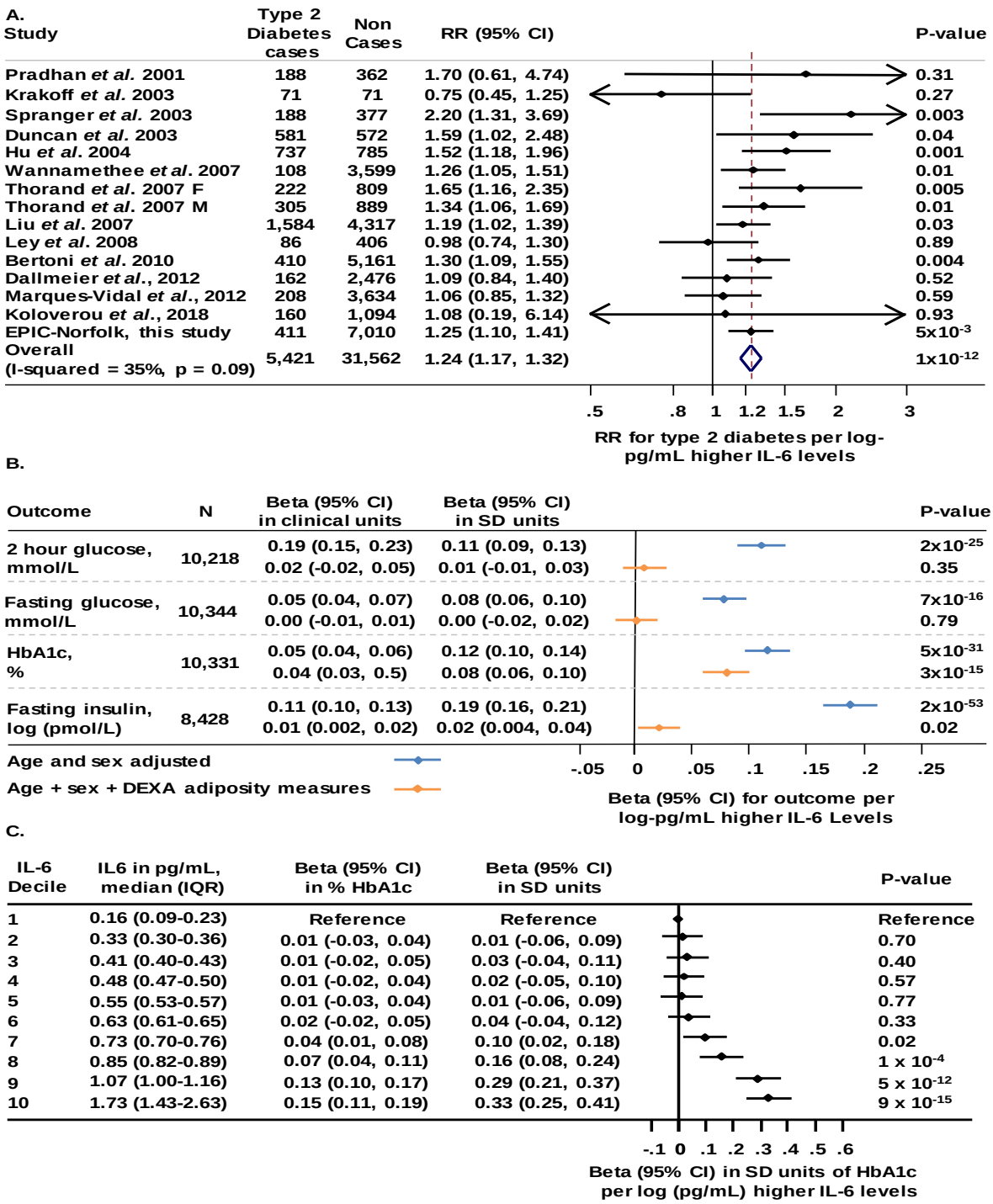


Figure 4.2: Association of IL-6 levels with type 2 diabetes and glycaemic traits. Panel A. Association between higher IL-6 levels and type 2 diabetes in 15 prospective studies. The study-specific estimates of association (here labelled "Relative Risk" (RR)) and their 95% confidence intervals are shown per log-pg/mL higher IL-6 levels. Thorand *et al.*, 2007 is depicted with sex-stratified estimates as different IL-6 quantitation methods were used in men and women. **Panel B.** Association between IL-6 levels and glycaemic traits in the Fenland study. Estimates are presented in standard deviation and clinical units of continuous outcome per log-pg/mL higher IL-6 levels. **Panel C.** Association between IL-6 levels (in deciles) and HbA1c.Estimates are shown for a given decile compared to the bottom decile (reference group) and are adjusted for age, sex, DEXA adiposity measures and iron-related traits. Abbreviations: RR, risk ratio; CI, confidence interval; F, females; M, males; pg, picograms; mL, millilitres; L, litres, mmol; millimoles; pmol, picomoles; HbA1c, glycated haemoglobin; DEXA, dual-energy X-ray absorptiometry; IQR, inter-quartile range; SD, standard deviation.

Genetically impaired IL6R signalling and T2D risk

In 14,528 participants from the Fenland and EPIC-Norfolk cohorts, each copy of the 358Ala partial loss-of-function variant in *IL6R* was associated with 0.11 log-pg/mL (95% CI, 0.09, 0.13; P<.001) higher IL-6 levels. In 260,614 cases and 1,350,640 controls, each copy of the 358Ala partial loss-of-function variant in *IL6R* was associated with an odds ratio for T2D of 0.98 (95% CI, 0.97, 0.99; P<.001; **Figure 4.3A**). In UK Biobank, the odds of T2D was lowest in carriers of two copies of the 358Ala allele (OR compared with non-carriers, 0.92; 95% CI, 0.89, 0.96; P<.001; **Figure S4.5**), with no statistical evidence of deviation from an additive association ($P_{non\text{-}linearity}$=0.13; **Figure S4.5**). In the UK Biobank study, estimates of association with T2D for Asp358Ala were comparable in odds ratio, variance explained and heritability to those for coronary heart disease (**Figure S4.5 and Table S4.9**). The association of Asp358Ala with diabetes was conditionally independent of an association peak for T2D led by rs2481065, which is located 115,059 base pairs upstream of the Asp358Ala variant (OR per copy of 358Ala conditioned on rs2481065, 0.97; 95% CI, 0.95, 0.99; P=.007; **Table S4.10**). The association with T2D of the Asp358Ala variant was not the consequence of case-admixture (diagnostic misclassification) between type 1 and T2D. First, in an analysis including 24,209 cases and 758,240 controls, each copy of the 358Ala allele was associated with an odds ratio for type 1 diabetes of 0.97 (95% CI, 0.95, 0.99; P=.007; **Figure S4.6**). Given the ~10-fold higher prevalence of type 2 vs type 1 diabetes in the general population[350] and similar magnitude of association, only very high levels of misclassification might lead to a spurious association with T2D (**Figure S4.7**). To further address this, a sensitivity analysis was conducted in UK Biobank where possible cases of type 1 diabetes were excluded on the basis of a cut-off of a polygenic score for type 1 diabetes (sensitivity to detect type 1 diabetes cases for exclusion >95%)[350]. In this analysis, estimates of the association with T2D were approximately the same as those in the main analysis (OR for T2D, 0.96; 95% CI, 0.94, 0.99; P=.008; **Figure 4.3B**).

In continuous metabolic trait analyses, the 358Ala allele was associated with lower HbA1c (beta in SD units per copy of 358Ala, -0.007; 95% CI, -0.012, -0.003; P=.002; **Figure S4.8**) and numerically lower BMI-adjusted WHR (beta in SD units per copy of 358Ala, -0.004; 95% CI, -0.008, 0; P=.03; **Figure S4.8**). The latter association was not statistically significant after accounting for the number of tests and was unlikely to fully explain the association with T2D (**Figure S4.8**).

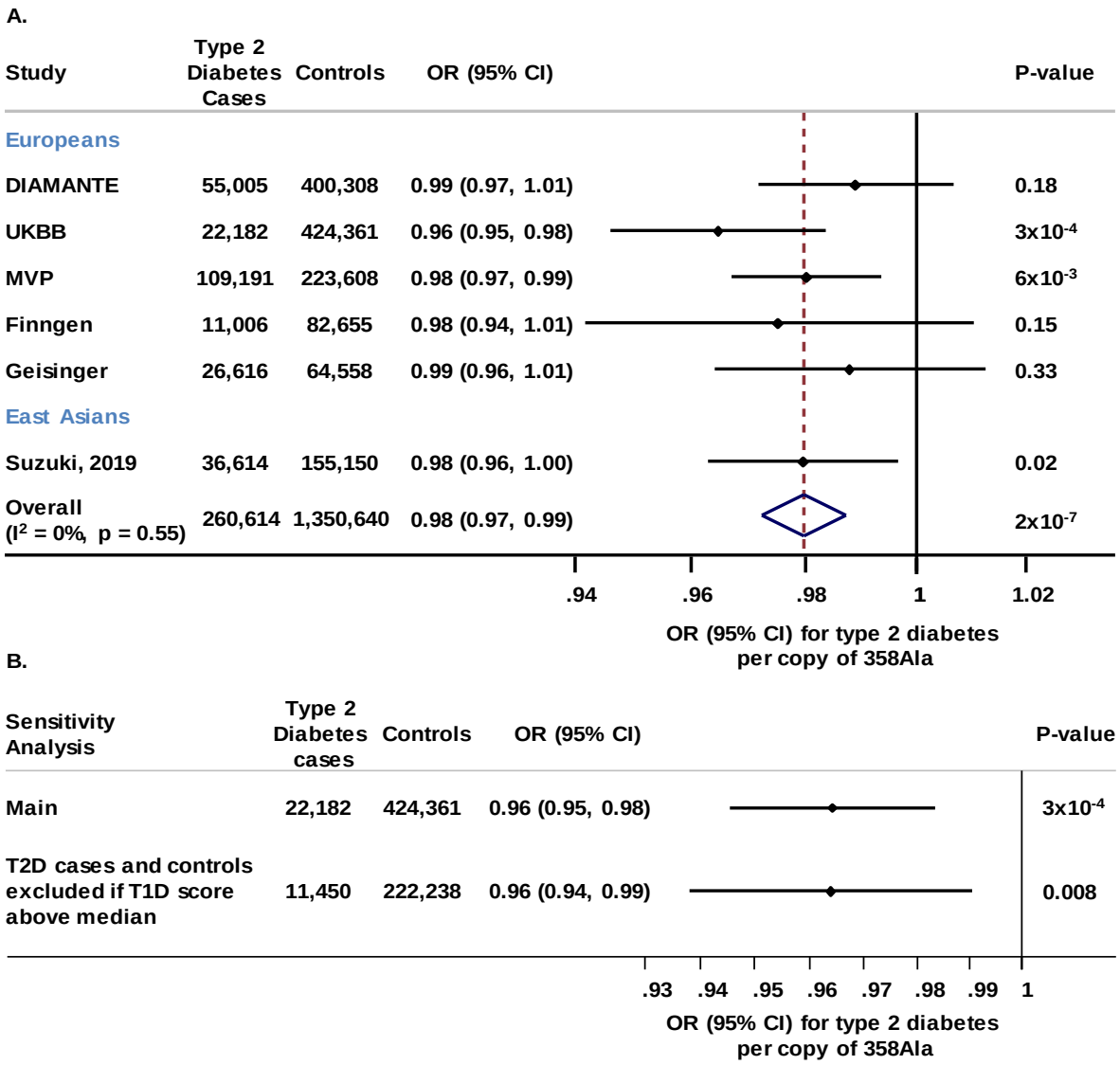


Figure 4.3: Association of *IL6R* Asp358Ala with type 2 diabetes, glycaemic and anthropometric traits. Panel A. Odds ratios and their 95% confidence intervals for type 2 diabetes are presented per copy of 358Ala. **Panel B.** Association of Asp358Ala with type 2 diabetes in UK Biobank before and after excluding participants based on the values of a type 1 diabetes polygenic risk score. Both cases and controls were excluded if the type 1 diabetes score was above the median of the value among diabetes cases. Abbreviations: DIAMANTE, Diabetes meta-analysis of trans-ethnic association studies; UKBB, United Kingdom biobank; MVP, Million veteran program OR, odds ratio; CI, confidence interval; Asp, aspartic acid; Ala, alanine; T2D, type 2 diabetes; T1D, type 1 diabetes.

The association of Asp358Ala genotype was similar in people above or below the median of 9 polygenic scores for T2D risk factors, with no evidence of interaction (**Figure S4.9A**). There was, however, evidence of a significant interaction between Asp358Ala genotype and a polygenic score for higher BMI on HbA1c levels ($P_{interaction}$=0.009; **Figure S4.9B**). Projections of the range of possible effects of IL-6R inhibition therapy in the primary prevention of T2D were generated based on results from genetic studies and biomarker data from trials. These projections suggested that treatment might attain benefits above 10% relative risk reduction,

corresponding to a relative risk of 0.90, only when assuming long-term treatment duration with high dose inhibitors (**Figure S4.10**).

Mediation by the IL-6 pathway of cardio-metabolic disease risk associated with higher adiposity

Both higher measured BMI or genetic predisposition to higher BMI via a 97-variant polygenic score were associated with higher circulating IL-6 levels (0.14 log-pg/mL higher IL-6 levels per 4.8 kg/m^2 [corresponding to one standard deviation] higher measured BMI; 95% CI, 0.12, 0.16; P<.001; 0.13 log-pg/mL higher IL-6 levels per 4.8 kg/m^2 higher BMI due to the polygenic score; 95% CI, 0.05, 0.21; P=.001). The percentage of the BMI/T2D association mediated by IL-6 was estimated to be 4% (95% CI, 3, 5%) in observational analyses and 3% (95% CI, 2, 5%) in genetic analyses. The percentage of the BMI/cardio-metabolic disease association mediated by IL-6 was estimated to be 18% (95% CI, 14, 25%) in observational analyses and 10% (95% CI, 7, 15%) in genetic analyses. The percentage of the association between BMI and fasting insulin levels mediated by IL-6 was estimated to be 3% (95% CI, 1, 6%) in genetic analyses.

4.4: Discussion

This large genetic and biomarker study supports the hypothesis that IL-6 mediated inflammation plays a role in the aetiology of T2D, but it also suggests that the impact of this pathway on disease risk in the general population may be small. Early evidence of a possible role of inflammation in metabolic disease dates back to over 25 years ago[139]. Findings of an inflammatory response accompanying obesity[109,138,223,356], insulin resistance[138,223,270] and diabetes[109,223,334,356], and the support of experimental results from cellular and animal models have led to the hypothesis that T2D may be viewed as an "inflammatory disease"[123].

This study provides large-scale human genetics evidence that supports the role of chronic inflammation via the IL-6 pathway in T2D. It also shows that IL-6 mediates in small part the links between obesity, insulin resistance and cardio-metabolic diseases, which has been a prevalent hypothesis in the field due to the evidence that the adipose tissue of insulin-resistant obese people is infiltrated with inflammatory cells[137,138] and displays an enhanced secretion of IL-6, tumour necrosis factor alpha and other pro-inflammatory mediators[109,270,356]. Biomarker[89,90,189,190] and human genetic[88,160,337] results with similar strength to those of this study have formed part of the evidence basis for efforts to re-purpose IL-6R inhibitory therapy in cardiovascular disease (e.g. clinicaltrials.gov registered trials number NCT03004703, NCT01491074, or NCT02419937).

However, several findings from this study suggest that the impact of the IL-6 pathway on risk of T2D in the general population may be small. Firstly, people in the top decile of IL-6 levels had a 2-fold (hazard ratio, 2.14) higher risk of incident T2D than people in the bottom decile. This is approximately three times smaller than the relative risk associated with being in the top decile of BMI in this study (hazard ratio, 5.64). While the association between Asp358Ala and lower T2D risk was statistically robust, consistent across studies and ancestries and was unaffected by possible misclassification of type 1 diabetes cases, the odds ratio of 0.98 for disease risk was lower for carriers of one copy and the odds ratio of 0.92 was lower for carriers of two copies of this partial loss-of-function allele compared to non-carriers. Finally, the proportion of BMI-associated risk of diabetes mediated by IL-6 levels was below 5%, ~3-4 times smaller than for coronary disease. In line with this, the proportion of the association between BMI and fasting insulin levels mediated by IL-6 was around 3%. Taken together, these results suggest that removing inflammation via this pathway may not substantially impact on diabetic risk in obesity but still represents a biologically significant effect.

This study has limitations. First, as this study is based on observational epidemiology findings and not randomised controlled trial data, it cannot definitively establish causality. Second, this

study did not exclude patients with pre-existing inflammatory conditions or acute infections at baseline. Although I cannot exclude the possibility that IL-6 levels may have been impacted by acute infections at baseline, to affect the observational results presented, IL-6 levels would have to differ significantly at the time of prospective type 2 diabetes diagnosis to have an effect. In addition, the study design included follow-up of these observational analyses with genetic analyses which are robust to this type of confounding. Despite these efforts to mitigate this potential confounding, capturing infection-mediated inflammation in observational studies is challenging and is therefore a limitation of this study. Third, some of the findings of this study assume that the *IL6R* Asp358Ala variant mimics the effects of IL-6R inhibition. While the *in vitro* consequences[310] and the pattern of association[88,160] of this variant with a variety of phenotypes have been shown to be consistent with the effects of IL-6R inhibitory drugs, differences cannot be excluded which could affect some of the estimates reported here. Using re-scaling methods to project the possible efficacy of IL-6R blocking therapy in primary prevention based on estimates of lifelong genetic associations, the projected relative risk reduction was below 10% in most simulated scenarios. In comparison, relatively safe and inexpensive lifestyle interventions[28], metformin[28] or other hypoglycaemic drugs[357] have been associated with reductions in the incidence of diabetes ranging from 30% to 70% in previous primary prevention trials. However, due to the underlying assumptions, this estimate is likely to be conservative and may underestimate the effect of IL-6R antagonists, which profoundly inhibit the pathway. In contrast to pharmacological antagonism, the *IL6R* Asp358Ala variant results in a partial loss of IL-6 mediated signalling and may therefore not be the most informative model to estimate the possible effects IL-6R blockade on glucose metabolism achieved through complete or near-complete pharmacological inhibition of this pathway. In addition, IL-6 is a pleiotropic cytokine that has established pro- and anti-inflammatory effects that act via the classical and trans signalling pathways[358]. It is therefore conceivable that the Asp358Ala variant acts upon the anti-inflammatory effects of IL-6 signalling and may have impacted upon the results of this study. Fourth, while some of the results from this study suggest that IL-6R inhibition might have limited efficacy for the primary prevention of T2D, these findings focus on first occurrence of disease in people from the general population. Hence, these findings do not exclude that IL-6R inhibition may yield clinically significant benefits on glucose metabolism in groups of people at risk for T2D, or on glycaemic levels and risk of complications in patients with T2D. In fact, in a post-hoc meta-analysis of three randomized controlled trials of the IL-6R inhibitor sarilumab including 1,982 patients with rheumatoid arthritis, Genovese and colleagues showed statistically-significant improvements in HbA1c in patients randomized to sarilumab as opposed to placebo (least squares mean difference in HbA1c ranging from −0.21% to −0.69%), which were more pronounced at higher dosages and in patients who also had diabetes[212]. This finding is further reinforced by results

demonstrating a significant 0.4% reduction in HbA1c in 67 rheumatoid arthritis patients treated with tocilizumab[359]. Therefore, it is possible that IL-6R inhibition may yield clinically meaningful changes in glycaemia in people with inflammatory or immune conditions which are linked with diabetes. Results from the interaction analysis between Asp358Ala and BMI on HbA1c levels suggest that reductions in HbA1c may be more pronounced in individuals below the median of the BMI polygenic score. However, as this result was estimated in a population-based cohort, this requires validation in a cohort selected for participants with existing inflammatory conditions. Finally, while the associations observed in this study are statistically robust and despite a growing body of research in the topic, the molecular mechanisms linking the IL-6 inflammatory pathway with glucose metabolism remain only partly understood. In this study, IL-6 levels and the *IL6R* Asp358Ala variant were more strongly associated with HbA1c, which reflects average glycaemic levels in the three months prior to measurement, than with other glycaemic traits. Infection and other inflammatory challenges are associated with reactive hyperglycaemia[360], which may be aggravated in people with higher IL-6 levels[361]. It is possible that enhanced glycaemic response to inflammatory stimuli may explain part of the associations between higher IL-6 mediated inflammation and higher diabetes risk observed in this study.

In conclusion, this large human genetics and biomarker study supports the hypothesis that IL-6 mediated inflammation is implicated in the pathophysiology of T2D but suggests that the impact of this pathway on disease risk in the general population is likely to be small.

Chapter 5: Estimating the Inflammatory Proteins and Pathways Implicated in Cardiometabolic diseases: A Multi 'Omics Approach

Contributions and collaborations

I designed the study, collated all input datasets and performed all analyses described in this chapter aside from the GWAS for each of the protein targets covered by the SomaLogic v4 assay. The SomaLogic GWAS formed part of a larger effort to characterise the role of circulating proteins on diverse health outcomes which is being led by Dr. Eleanor Wheeler. Both Dr. Wheeler and Dr. Maik Pietzner provided methodological support during this work. I wrote the first draft of this chapter which was later reviewed by both Dr. Langenberg and Dr. Pietzner.

Abstract

Background: Evidence from *in vitro* studies, animal models and observational epidemiology points to an inflammatory component in the aetiology of cardiometabolic diseases. However, the proteins and pathways involved are sparsely characterised and the direction of effect is a matter of debate, i.e. it is unclear whether inflammation is causal for or a result of disease.

Methods: Using a combination of GO terms and reciprocal lookup, inflammatory proteins with suggestive evidence ($P \leq 1x10^{-5}$) of a genetic association with cardiometabolic diseases were identified. Next, the colocalisation between all identified proteins and cardiometabolic diseases was estimated using HyPrColoc, a Bayesian multi-trait colocalisation framework. Following this, the levels of proteins which colocalised with cardiometabolic diseases were tested for association with polygenic scores for established risk factors for cardiometabolic diseases using multivariable linear regression. Finally, gene set enrichment analysis was performed to estimate whether inflammatory proteins which colocalised with cardiometabolic diseases were enriched in particular pathways.

Findings: A total of 3,214 human proteins were identified using GO terms, of which 1,358 were captured by the SOMAscan v4 assay. Reciprocal lookup of sentinel variants for disease and inflammatory protein levels in GWAS summary statistics resulted in 335 inflammatory proteins with at least suggestive evidence of an association with either of the two diseases. A total of 10 and 18 loci were identified where inflammatory proteins colocalised with T2D and CHD, respectively. T2D was estimated to colocalise with a total of 164 inflammatory proteins whereas CHD colocalised with 181 proteins. However, closer examination showed that 95% and 85% of proteins which colocalised with T2D and CHD colocalised at a minimum of one established pleiotropic locus. In addition, 42% and 36% of proteins which colocalised with T2D and CHD respectively were associated with at least one polygenic score (PGS) for an established cardiometabolic risk factor, with triglyceride-PGS associated with the levels of the most proteins in each case. Finally, gene set enrichment analysis showed that inflammatory proteins which colocalised with cardiometabolic diseases were significantly enriched in the human complement cascade.

Conclusion: This study identified multiple loci where inflammatory proteins colocalised with cardiometabolic diseases, however, this was largely at pleiotropic loci. The strong association of PGS for established cardiometabolic risk factors with many candidate proteins suggests that inflammation may be resultant of the genetic predisposition to either the disease or its associated risk factors. Therefore, the inflammatory component of cardiometabolic diseases is unlikely to be an independent causal risk factor for disease and is likely to be resultant of a combination of pleiotropic effects and established cardiometabolic risk factors.

5.1: Introduction

Modern technologies have facilitated the simultaneous, large scale quantification of many biomarker levels. This has increased interest in their relevance to disease as putative drug targets and how they may be used to predict disease risk[68]. The inflammatory component of cardiometabolic diseases is a central aspect of the aetiology of cardiometabolic diseases that is still yet to be fully elucidated[4,109,112,123,125,223,268]. As outlined in **Chapter 1**, evidence from *in vitro* studies[111,137–139], animal models[123,223,270] and observational epidemiology[94,154,155,168,181] has pointed to a role for chronic inflammation in cardiometabolic diseases. However, whether this is causal or is a result of either the disease itself or related risk factors is yet to be established. Related to this, the proteins that function in the inflammatory component of cardiometabolic diseases as well as the putative pathways through which they act to modulate disease risk are largely unknown.

GWAS studies have led to the identification of thousands of genetic loci that are associated with clinical endpoints and phenotypes of interest[284,362]. However, identifying the underlying causal variant and the molecular basis of these associations is challenging. One method of addressing this problem is by testing whether genetic associations are shared between traits, an approach known as genetic colocalisation[284]. The presence of a shared variant increases the likelihood that the traits share a common causal mechanism or pathway[284]. This approach uses a Bayesian framework to estimate the likelihood of a causal variant underlying a genetic association being shared between pairs of traits[284]. However, this method only estimates whether the genetic association is likely to be shared between two traits and does not refine this to a single candidate causal variant. Secondly, this approach can quickly become inefficient when considering hundreds of traits as independent pairwise analyses would need to be conducted per locus to enumerate all possible trait combinations.

Recent methodological extensions to this method[362,363] enable the rapid enumeration of colocalisation across multiple traits. In particular, HyPrColoc[362] uses a Bayesian framework to estimate clusters of traits which share a genetic association and refines this to distinct causal variants by employing a fine mapping approach (described in **Chapter 3**). This is therefore an efficient and powerful method of identifying shared loci across traits while estimating the likely causal variant underlying the association for future follow-up. Application of such a framework[362] in this setting helps to elucidate and contextualise the role of inflammation in cardiometabolic diseases by providing information on two levels. Firstly, the method[362] is able to fine map shared association signals between traits to single candidate variants and provide empirical estimates of the likelihood that the candidate variant is shared between traits of interest. This is of interest as these candidate variants tag biological pathways that may

explain how they affect disease risk. Secondly, the proteins which are estimated to colocalise with disease are of interest as they highlight shared aetiology and therefore provide biological context surrounding the role of inflammatory proteins in the aetiology of cardiometabolic diseases.

In this study, I aimed to characterise the genetics underlying the inflammatory component of cardiometabolic diseases and infer the biological pathways. I leveraged large-scale aptamer-based proteomic data in a multi-trait colocalisation framework to identify loci where clusters of inflammatory proteins, cardiometabolic diseases and risk factors colocalise at shared causal variants. I then further characterise these protein clusters by estimating their association with established cardiometabolic risk factors and estimating the effects of pleiotropic loci on inflammation in cardiometabolic diseases.

5.2: Methods

This work made use of GWAS summary statistics that formed part of a larger effort to study the genetic underpinnings of plasma protein levels and how these relate to diseases. The methods are described briefly in the following section.

Profiling of the plasma proteome

Fasted EDTA-plasma samples from 12,084 participants from the Fenland[244] study (described in **Chapter 2**) were subjected to proteomic profiling by SomaLogic Inc. (Boulder, US) using an aptamer-based technology (SOMAscan v4). The relative abundances of 4,775 human proteins were measured using 4,979 SOMAmers, as previously described[364]. To account for within run hybridisation variability, control probes were used to generate a scaling factor for each sample. Differences in total signal between samples as a result of variation in overall protein concentration or technical variability such as reagent concentration, pipetting or assay timing, were accounted for using the ratio between each SOMAmer's measured value and a reference value. The median of these ratios was computed for each dilution set (40%, 1% and 0.005%) and applied to each dilution set. Samples were removed if they failed SomaLogic QC measures or did not meet the acceptance criteria of between 0.25-4 for all scaling factors. A total of 10,078 samples had available genotype data and were used in this study. Aptamer target annotations and mapping to UniProt accession numbers as well as gene identifiers were provided by SomaLogic.

GWAS and meta-analysis of human plasma protein levels

Genotyping and imputation in the Fenland study were performed as described in **Chapter 2**. Related individuals and ancestry outliers were removed, leaving 10,708 participants with both protein measures and genetic data for association testing (OMICS=8,350, Core-Exome=1,026, GWAS=1,332). Prior to GWAS, relative abundances of each SOMAmer were transformed within each genotyping subset using the rank-based inverse normal transformation. These were then regressed on age, sex, sample collection site and 10 principal components to generate residuals. GWAS was performed using BGENIE (v1.3), assuming an additive model. Variants with minor allele frequency < 0.001, imputation quality < 0.4 or Hardy Weinberg Equilibrium $p < 1x10^{-7}$ in any of the genotyping subsets were excluded from further analyses. Results from the three genotyping subsets were later combined in a fixed-effects meta-analysis using METAL[276]. Only variants present in the largest genotyping subset were taken forward for further analysis, variants specific to either of the smaller subsets were not considered.

Locus definition and *cis/trans* annotation

A genome-wide significance threshold of $P<1x10^{-11}$ was used, adjusting the conventional genome-wide significance threshold to account for 5,000 tests. Non-overlapping loci were defined as 500Kb regions either side of all genome-wide significant variants, any overlapping regions were merged. The MHC region (chr6:25.5–34.0Mb) was treated as a single region. Within each region, sentinel variants were defined as the variant with the lowest P-value in the region. Variants were classified as *cis*-acting pQTLs if the variant was less than 500Kb from the respective protein encoding gene.

Conditional analysis

Conditional analysis was performed as described in **Chapter 3**, independently for each aptamer. As an additional quality control measure, all variants identified by GCTA[365] in a given region were jointly modelled using individual level data from the largest available genotyping subset. Any variants that were no longer genome-wide significant were not taken forward.

Inflammatory protein identification

To identify proteins which play a role in inflammation, defined as the response of the immune system to an exogenous or endogenous stimulus, gene ontology (GO)[366,367] terms were used. GO terms[366,367], accessed via the Amigo2 browser[368], have a tree hierarchical structure split into three main classes: biological processes, cellular components and molecular functions. In order to obtain the broadest and most comprehensive list, all human proteins annotated to function under the biological process annotation “immune system process” were selected. Proteins were then filtered to retain only human proteins; any micro RNA or long non-coding RNA annotations were removed. Finally, the list of identified proteins was searched by Uniprot ID in the SomaLogic dataset to identify inflammatory proteins of interest. This process yielded a total of 1,358 candidate protein targets.

Genetic correlation between inflammatory proteins and cardiometabolic traits

Using the inflammatory proteins identified above, pairwise genetic correlation was estimated between the proteins and cardiometabolic traits of interest using LDSC v1.0.0[287,291]. Prior to implementing LDSC, variants within the *HLA* region as well as those that displayed effect sizes larger than $\chi^2 > 80$ (corresponding to a Z-score of 8.9) were removed, as described in **Chapter 3**.To retain high imputation quality (> 0.9) the analysis was restricted to variants imputed to HapMap3, as these were considered to be well imputed across studies. The cardiometabolic traits considered were BMI, WHRadjBMI, T2D, CHD, HbA1c, fasting insulin, fasting glucose, 2-hour glucose, HDL, LDL and triglycerides. Fasting insulin, glucose and 2-hour glucose were all adjusted for BMI. Several autoimmune diseases were included as comparators with the aim of delineating inflammatory proteins more correlated with autoimmune disease, these

included rheumatoid arthritis, Crohn's disease, ulcerative colitis and inflammatory bowel disease. Details of the included studies can be found in **Table 5.1**.

Firstly, heatmaps were drawn to illustrate the pairwise correlations between cardiometabolic traits and diseases using the pheatmap package in R v3.6.3. Next, a second heatmap illustrating the pairwise correlation between cardiometabolic traits and inflammatory proteins was drawn. Inflammatory proteins that were significantly correlated with at least one cardiometabolic trait were included. P-values smaller than $P<1.4x10^{-4}$, accounting for the 350 traits tested, were considered statistically significant.

Table 5.1: Summary of the included traits and their data sources					
Trait class	**Outcome [a]**	**Cases overall, N**	**Non-cases (for case-control studies) or participants (for continuous trait studies) or number of variants (for polygenic scores) overall, N**	**Participating study**	**PubMed ID for cohort description**
Cardiometabolic disease	Type 2 diabetes	74,124	824,006	DIAMANTE	30297969
	Coronary heart disease	34,541	261,984	CARDIoGRAMplusC4D, UK Biobank	28714975, 25826379
Anthropometry	BMI		738,628	GIANT, UK Biobank	25673413; 25826379
	WHRadjBMI		636,282		
Glycaemic	HbA1c		451,782	UK Biobank; InterAct	25826379
	Fasting insulin adjBMI		141,325	MAGIC	DOI: 10.1101/2020.07.23.217646
	Fasting glucose adjBMI		190,384		
	2-hour glucose adjBMI		62,183		
Lipid	HDL		450,957	UK Biobank; InterAct	25826379
	LDL		375,774		
	Triglycerides		450,625		
Cytokines	IFNγ		17,030	EPIC-Norfolk; Fenland	**Chapters 2 and 3**
	IL-6		17,318		
	IL-8		16,353		
	TNFα		17,330		
Autoimmune disease	Rheumatoid arthritis	29,880	73,758	Meta-analysis of 26 studies	24390342
	Crohn's disease	12,194	28,072	Cambridge MREC; 26192919	28067908
	Ulcerative colitis	12,366	33,609		
	Inflammatory bowel disease	25,042	34,915		
Plasma proteins	4,979 proteins		10,708	Fenland	27841877
Polygenic scores	T2D		386	DIAMANTE	30297969
	T2DadjBMI		191		
	CHD		73	CARDIoGRAMplusC4D, UK Biobank	28714975, 25826379
	BMI		96	GIANT	25673413
	Hip		22	UK Biobank; GIANT	30575882
	Waist		36		
	WHR		202		
	HipadjBMI		22		
	Waist adjBMI		36		
	WHRadjBMI		202		
	Insulin resistance		53	MAGIC; GLGC	27841877
	Fasting insulin adjBMI		30	MAGIC	DOI: 10.1101/2020.07.23.217646

Abbreviations: N, Number of participants; HbA1c; Glycated haemoglobin; BMI; Body mass index; DIAMANTE, Diabetes meta-analysis of trans-ethnic association studies; MAGIC, Meta-Analyses of glucose and insulin-related traits consortium; EPIC, European prospective investigation of cancer; GLGC, Global lipid genetics consortium

a. Traits are continuous unless otherwise stated as a disease or a polygenic score

Reciprocal lookup of independent variants in inflammatory protein targets and cardiometabolic disease GWAS summary statistics

Next, I aimed to determine whether established T2D and CHD loci are protein quantitative trait loci (pQTLs) for inflammatory proteins and *vice versa*, whether pQTLs for inflammatory proteins were associated with disease risk. To do this, a reciprocal lookup approach was used. Firstly, using published GWAS summary statistics for T2D[22] and CHD[289], the independent variants associated with disease risk were extracted from the GWAS summary statistics of each inflammatory protein. Similarly, independent variants associated with levels of inflammatory proteins were extracted from GWAS summary statistics for each cardiometabolic disease. In cases where the independent variant itself was unable to be found in the reciprocal GWAS, proxy variants in LD ($R^2 > 0.8$) with the variant of interest were used. Following this, independent variants associated with disease ($P<5x10^{-8}$) were prioritised to take forward any variant with at least suggestive evidence of an association with protein levels ($P<1x10^{-5}$). Similarly, only independent variants associated with inflammatory protein levels ($P<5x10^{-8}$) that were within or near the protein-coding gene (*cis* variants) were prioritised to take forward any variant with at least suggestive evidence of an association with disease ($P<1x10^{-5}$). This was done to facilitate biological interpretation of the effects of the variant. Loci were then defined, as described in **Chapter 3**, as 500Kb regions either side of the variant with the lowest P-value in the region. Any overlapping loci were merged into a single locus, yielding 85 loci in total.

Multi-trait colocalisation between inflammatory proteins and cardiometabolic traits

I used multi-trait colocalisation (HyPrColoc)[362] within each of the loci defined above to identify inflammatory proteins which share a common causal variant with cardiometabolic diseases. In short, HyPrColoc employs a Bayesian framework to estimate whether multiple traits colocalise within a single locus at a shared genetic signal and creates clusters of traits which colocalise around distinct causal variants in the region. HyPrColoc was run using a variant-specific prior configuration where default settings were as follows 1) the prior probability that a variant is associated with a single trait, prior 1, was set at $1x10^{-4}$, and 2) prior 2, the prior probability that a variant is associated with an additional trait, given that it is already associated with one trait, was set at 0.02[369]. A regional threshold of 0.5 was used, denoting the probability that all traits share an association with one or more variants in the region. Related to this, an alignment threshold of 0.5 was also used, relating to the probability that all traits align and share a single causal variant.

The 500Kb regions either side of the lead variant at each locus were extracted from GWAS summary statistics for the 335 inflammatory protein targets with suggestive evidence of an

association with a disease, 4 cytokines and 11 cardiometabolic traits of interest. The cardiometabolic traits consisted of disease outcomes, anthropometric traits, lipids and glycaemic measures as previously outlined.

To assess sensitivity in the number and size of clusters identified to the specification of algorithm parameters, increasingly stringent prior and threshold configurations were used. The values of prior 2 considered were 0.02 and 0.01, threshold values of 0.5, 0.6, 0.7, 0.8 and 0.9 were considered. Heatmaps based on similarity matrices were drawn to illustrate how often trait pairs were clustered together across all algorithm parameter choices. This protocol is notably stringent therefore results that passed thresholds such as these have robust evidence for colocalisation. Only variants present in all included traits were considered for each respective locus and any variants with a reported standard error of 0 were removed. T2D and CHD were considered as binary case-control traits whereas all others were considered quantitative. To estimate the posterior probability that the candidate variant is the causal variant (PP_{causal}), I multiplied the PP for colocalisation (PP_{coloc}) by the PP explained by the candidate variant ($PP_{explained}$).

In order to prioritise high confidence clusters of inflammatory proteins and cardiometabolic traits, several prioritisation criteria were used, 1) only clusters where inflammatory proteins colocalised with a cardiometabolic disease were considered, 2) clusters were retained if they showed sufficient evidence for colocalisation with disease i.e. $PP_{coloc} \geq 0.5$ and regional posterior probability ($PP_{regional}$) ≥ 0.8. Gene and consequence information for each candidate variant were annotated using VEP v99[285].

Diverse traits colocalising at a given locus do not necessarily imply a common casual mechanism. I therefore established a classification system for each of the proteins to characterise the levels of specificity. Pleiotropic loci were defined based on the literature[370–373] as any of *APOE*, *GCKR*, *TM6SF2*, *ABO* and *SH2B3*, all of which have been associated with diverse outcomes. Non-pleiotropic loci were considered as none of the above genes. In detail, proteins only colocalising at well-known pleiotropic loci, as defined, were classified as pleiotropic. Those colocalising at pleiotropic and non-pleiotropic loci were classified as "mixed", and most notably, proteins colocalising at specific loci were classified as non-pleiotropic. In order to assess the contribution of pleiotropic loci to the colocalisation between inflammatory proteins and cardiometabolic diseases, the number of proteins that colocalised at non-pleiotropic, pleiotropic and a mixture of non- and pleiotropic loci were counted. Pleiotropic classifications were conducted for each disease independently.

Association of candidate variants with all human proteins

For candidate variants of interest arising from the HyPrColoc analysis, the association of the variant and all 4,979 human proteins on the SOMAscan v4 platform was estimated using multivariable linear regression adjusting for age, sex, sample collection site and the first ten genetic principal components. Test site refers to the testing centre where blood samples were taken from participants. As sample processing and storage protocols varied by test site, internal quality control testing showed that protein levels varied by test site as well. Thus, to avoid spurious associations, this was included as a covariate. Volcano plots to illustrate the association of the variant with each protein were drawn using the ggplot2 R package. All data analysis was performed using R version 3.6.3.

Association between polygenic scores for cardiometabolic traits and inflammatory protein levels

It is currently unclear whether the inflammation observed in cardiometabolic diseases is resultant of the disease itself or its related risk factors. To estimate whether cardiometabolic risk factors and disease endpoints are associated with inflammatory protein levels and may therefore be mediating the observed inflammation in cardiometabolic diseases, the associations between polygenic scores for 12 cardiometabolic traits (**Table 5.1**) and levels of the inflammatory proteins were estimated. The four cytokines of interest were also included in the list of inflammatory proteins.

Firstly, the variants and weighting for each score was extracted from the respective publication (**Table 5.1**), except for the CHD score, which was accessed using the ieugwasr R package (ID: ebi-a-GCST005195). Each score was then depleted for variants within 200Kb either side of a pQTL for one of the inflammatory proteins. This was done to avoid spurious associations between scores and inflammatory protein levels resulting from variants that affect both traits. Following this, genotype dosages for each of the remaining variants were extracted in Fenland participants genotyped using the Affymetrix UK Biobank Axiom Array, the largest genotyping sub-cohort. Dosages were aligned to the score-raising allele and weighted by their effect on the respective cardiometabolic trait. A score for each cardiometabolic trait was generated per individual by summing the respective weighted variants contributing to the score. Finally, inflammatory protein levels were inverse-rank normal transformed so that each protein had a mean of 0 and a standard deviation of 1. The association between the polygenic score and inflammatory protein levels was estimated using multivariable linear regression adjusted for age, sex, test site and the first ten genetic principal components.

Volcano plots were used to visualise the associations between polygenic scores and inflammatory protein levels. These were coloured according to their colocalisation with cardiometabolic diseases and whether this colocalisation was estimated to be driven by known pleiotropic loci (*APOE*, *GCKR*, *TM6SF2*, *SH2B3* or *ABO*)[370–373]. A Bonferroni significance threshold of $P<1.5x10^{-4}$ was used to account for 339 tests, corresponding to 335 inflammatory proteins with suggestive evidence of an association with cardiometabolic diseases and the four cytokines of interest. All data analysis was performed using R version 3.6.3.

Gene-set enrichment analysis

In order to estimate whether the proteins which colocalised with T2D and CHD were enriched in terms of their pathway involvement, gene-set enrichment analyses were conducted for each disease independently. Enrichment analysis was conducted using the functional mapping and annotation (FUMA) v1.3.6a[374] web application GENE2FUNC process. Curated gene sets for testing enrichment were downloaded from the molecular signatures database (MSigDB) v7[375]. As pathway annotations are notably database specific, MSigDB combines data from several sources including KEGG, reactome and Biocarta to circumvent this. In this analysis, gene-set enrichments were only considered significant if the same gene set from at least two databases was significant.

The protein-encoding gene for each protein which colocalised with disease were used as input. The genes for all cytokines and inflammatory proteins present in the SomaLogic dataset that were identified using GO terms were used as the background for the analysis. Only gene-sets consisting of more than 2 genes were included. Significant overlap between input genes and gene-sets was performed using linear regression[376]. Multiple testing correction using the Benjamini-Hochberg method at a false discovery rate of $\alpha = 0.05$ was applied.

5.3: Results

A total of 3,214 human proteins were identified using GO terms as playing a role in an "immune system process". Of these, 1,358 proteins (42%) were covered by the SomaLogic v4 assay. The median number of variants estimated to be significantly associated with protein levels was 2 (IQR: 1-4). Genetic correlation estimates between autoimmune diseases and cardiometabolic traits showed a clear cluster of cardiometabolic traits which was separate from a cluster of autoimmune diseases (**Figure S5.1**). Genetic correlation estimates showed that autoimmune diseases and cardiometabolic traits were significantly correlated with levels of 68 proteins (**Figure 5.1**). However, no clear delineation in the correlation of inflammatory proteins with autoimmune diseases and cardiometabolic traits was observed. Total triglycerides were significantly correlated with levels 38 of these proteins (56% of the significant proteins; 3% of the total), the most of any trait, whereas T2D and CHD were significantly correlated with levels of 24 and 18 proteins, respectively. The most significant correlation with both T2D and CHD was transforming growth factor beta receptor 3 (TGF-b R III) with comparable correlation estimates of 0.37 (genetic correlation estimate (r_g)=-0.37; 95% CI, -0.47, -0.26; P=1.6x10^{-12}) and 0.36 (r_g=-0.36; 95% CI, -0.47, -0.26; P=2.3x10^{-11}), respectively. However, the strongest correlation with T2D was insulin like growth factor binding protein 2 (IGFBP-2) (r_g=-0.65; 95% CI, -0.87, -0.44; P=2.1x10^{-9}). The strongest positive correlation with T2D was with leptin levels (r_g=0.53; 95% CI, 0.36, 0.69; P=1.1x10^{-10}), a correlation that was also the strongest for BMI. The strongest positive correlation with CHD was with IL-6 levels (r_g=0.37; 95% CI, 0.18, 0.55; P=1.1x10^{-4}), IL-6 levels were also strongly correlated with T2D and were ranked second only to leptin in terms of correlation strength (r_g=0.43; 95% CI, 0.24, 0.62; P=8.9x10^{-6}). However, while genetic correlation estimates provide insight into common genetic architecture between traits, the direction of effect is unable to be inferred.

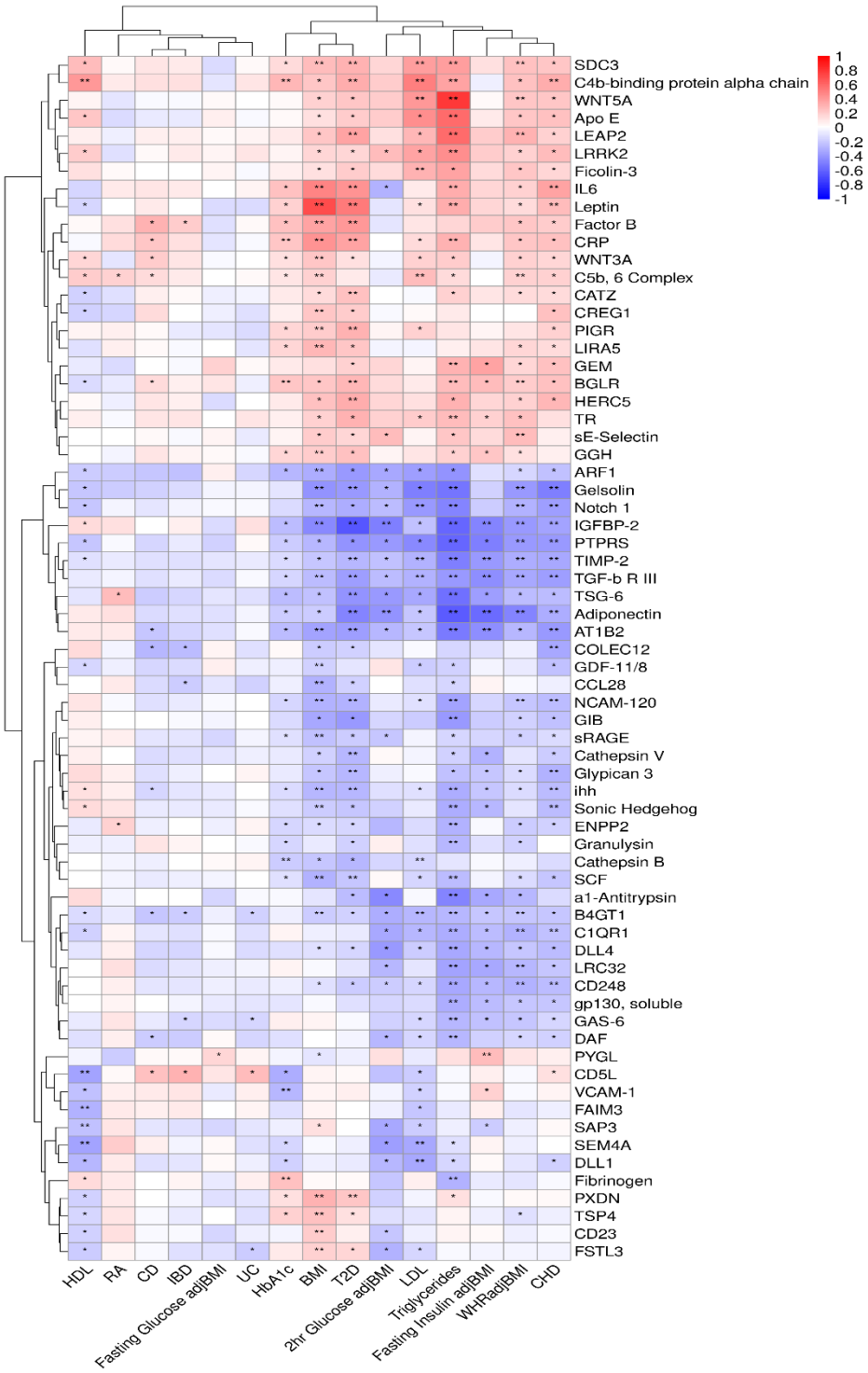


Figure 5.1: Correlation heatmap illustrating the pairwise genetic correlation of autoimmune diseases and cardiometabolic traits with inflammatory protein levels. Only proteins that are significantly correlated with one autoimmune or cardiometabolic trait below a Bonferroni significance threshold of P< 1.4 x10^{-4} are shown. The data sources of each trait are outlined in **Table 5.1**. Positive correlations are shown in red whereas inverse correlations are shown in blue. Colour saturation denotes the strength of the correlation between traits. Significant correlations at a nominal significance threshold of P< 0.05 are shown with a single asterisk "*", whereas correlations significant at a Bonferroni significance threshold are shown with a double asterisk "**".

Abbreviations: BMI, Body mass index; HbA1c, Glycated haemoglobin; T2D, Type 2 diabetes; CHD, Coronary heart disease; HDL, High density lipoprotein; LDL, Low-density lipoprotein; RA, Rheumatoid arthritis; UC, Ulcerative colitis; IBD, Inflammatory bowel disease.

Reciprocal lookup of cardiometabolic and inflammatory GWAS sentinels

A total of 386 and 118 independent variants for T2D and CHD, respectively, were extracted from the GWAS summary statistics for each of the 1,358 inflammatory proteins. Following this, 2,067 unique pQTLs associated with inflammatory protein levels were extracted from GWAS summary statistics for the two diseases. Once results were filtered to retain variants with P-values less than $P<1x10^{-5}$, 290 and 280 of the T2D and CHD variants met this threshold in the inflammatory protein GWAS. Comparatively fewer inflammatory sentinel variants were retained in the disease GWASs, of the 2,067 variants, only 73 and 94 variants were found in the T2D and CHD GWAS results, respectively. This was reduced to 11 variants extracted from each disease GWAS once only *cis* variants were considered. Only proteins with suggestive evidence ($P<1x10^{-5}$) of an association with cardiometabolic diseases were retained, resulting in 335 inflammatory proteins of interest (**Table S5.1**). Finally, a total of 85 1Mb loci were defined and taken forward for multi-trait colocalisation analysis (**Table S5.2**).

Multi-trait colocalisation between inflammatory proteins and cardiometabolic traits

Multi-trait colocalisation was conducted at each of the 85 loci (**Table S5.2**) to estimate whether subsets of inflammatory proteins colocalised with cardiometabolic diseases. Across the 85 loci, a median of 2,717 (IQR: 3,665 – 2,153) variants were included. At 10 and 18 loci, respectively, T2D and CHD were estimated to colocalise with at least one inflammatory protein where $PP_{coloc} \geq 0.5$ and $PP_{regional} \geq 0.8$ (**Table 5.2** and **Table 5.3**). T2D was estimated to colocalise with a total of 164 inflammatory proteins whereas CHD colocalised with 181 proteins. At 6 (60%) and 13 (72%) loci for T2D and CHD respectively, the candidate variant was estimated to explain more than 80% of the posterior probability ($PP_{explained} > 0.8$). Interestingly, inflammatory proteins were not estimated to colocalise with disease at any of the inflammatory sentinels, i.e. protein-encoding genes, included in the analysis, except for those that were already established disease sentinels.

Table 5.2: Loci where T2D colocalised with inflammatory proteins										
Sentinel	**Locus**	**Traits**	**PP coloc**	**Candidate variant**	**PP explained**	**N SNPs**	**Prior 2**	**Thresh**	**Gene**	**Consequence**
rs7412	6	HERC5, CRP, ARGI1, GGT2, MMP-8, T2D, HbA1c, WHRadjBMI	0.97	rs429358	1	2,771	0.98	0.9	***APOE***	Missense
rs72802342	11	REG1B, DLL4, TLR4:MD-2 complex, GIB, GRN, Trypsin 2, IL-1 sRI, T2D, HbA1c, Fasting Glucose adjBMI	0.68	rs72802342	1	3,401	0.98	0.8	*CTRB2*	Downstream
rs1260326	16	LRRK2, IgA, SELS, CD59, GEM, HERC5, WNT5A, B4GT1, PXDN, GRB14, M-CSF R, KIF3A, Ephrin-A2, EFNB2, C1QR1, RNF8, GAS-6, SAP3, ABP1, CD248, RBP, SPLC2, PACAP, Alpha-1B-glycoprotein, FADD, SDC3, SEM4A, L-plastin, CD46, CRIP1, CEAM8, ARF1, UB2D3, SP-D, TIMP-2, Apo E, Myeloperoxidase, Cystatin C, gp130, soluble, suPAR, EMAP-2, GDF-11/8, Fibrinogen, Lipocalin 2, DAN, TNFSF15, MBL, C9, IL-18 BPa, TNF sR-II, Coagulation Factor VII, CD109, CD48, DLL4, FCG3B, TSP4, FSTL3, b2-Microglobulin, TAFI, Adiponectin, a1-Antitrypsin, Angiostatin, Thrombin, IGF-I sR, NCAM-L1, MSP, NCAM-120, sL-Selectin, Epithelial cell kinase, Coactosin-like protein, C1QBP, TSG-6, DAF, ILT-2, ILT-4, JAG1, Notch 1, SLAF6, DLL1, Fas, soluble, sLeptin R, DR6, ALCAM, FGL1, LEAP2, SEM4D, PTPRS, CECR1, GGT2, LIRA3, GP116, CPN2, MFAP5, FAIM3, HBD-1, DRB3, Semaphorin-7A, LRC32, LIRA5, ROR2, CRTAM, NKp46, IGFBP-2, Leptin, NR1H2, CATF, C1RL1, GGH, C4b-binding protein alpha chain, LPLC1, VEGF-D, Gelsolin, ENPP2, BT3A3, TLR5, Apo A-I, TGF-b R III, ADAM 9, CLM6, TCCR, PGRP-L, NPFF, CM35H, FCRL1, EMBP, BMI, LDL, T2D, HbA1c, HDL, Triglycerides, 2hr Glucose adjBMI, Fasting Glucose adjBMI, Fasting Insulin adjBMI	0.65	rs1260326	1	1,570	0.98	0.8	***GCKR***	Missense; splice region
rs1050362	58	Hexosaminidase B, Angiostatin, Haptoglobin, Mixed Type, Galectin-3, T2D	0.78	rs217181	1	2,297	0.98	0.8	*HPR*	Downstream
rs8107974	35	WNT5A, NEUR1, Hexosaminidase B, BGLR, CRIP1, Apo E, Apo B, Angiostatin, CATZ, Cathepsin D, prosaposin, GGT2, LIRB4, Cathepsin B, CATF, CREG1, GGH, ADAM 9, MIP-3b, LDL, T2D, CHD, HbA1c, HDL, Triglycerides, WHRadjBMI	0.73	rs58542926	0.99	1,934	0.98	0.8	***TM6SF2***	Missense
rs507666	24	PSP, IL-1 sRII, REG3G, Complement receptor type 2, SPLC2, REG1B, Complement receptor type 1, 6Ckine, suPAR, Calpain I, AIF1, Midkine, MBL, DC-SIGN, sL-Selectin, GRN, TR, NFASC, LRC32, IL-1F6, T2D	0.68	rs529565	0.83	3,161	0.98	0.8	***ABO***	Intron
rs10882101	2	FCG3B, T2D	0.94	rs12219514	0.66	2,365	0.98	0.5	*Y RNA*	Downstream
rs2972144	4	cIAP-1, CLC4G, WNT5A, C1QBP, PTPRS, LDL, T2D, CHD, HbA1c, Triglycerides, 2hr Glucose adjBMI, Fasting Glucose adjBMI, Fasting Insulin adjBMI	0.68	rs2943646	0.56	2,347	0.98	0.8	3' of *IRS1*	Intergenic
rs2072633	47	Osteocalcin, T2D	0.86	rs7382085	0.30	4604	0.98	0.6	*HLA-DRA*	Downstream
rs2268382	85	Leptin, T2D, HbA1c	0.84	rs61462211	0.28	2,195	0.98	0.8	5' of *KLF14*	Intergenic

Abbreviations: Coloc, Colocalisation; BMI, Body mass index; HbA1c, Glycated haemoglobin; T2D, Type 2 diabetes; CHD, Coronary heart disease; HDL, High density lipoprotein; LDL, Low-density lipoprotein.

1. Gene names coloured in red are known pleiotropic loci
2. Loci are reported at the lowest prior 2 and threshold configuration where $PP_{coloc} \geq 0.5$ and $PP_{explained} \geq 0.8$, clusters may alter with increasingly stringent prior and threshold configurations

Table 5.3: Loci where CHD colocalised with inflammatory proteins										
Sentinel	**Locus**	**Traits**	**PP coloc**	**Candidate variant**	**PP Expl.**	**N SNPs**	**Prior 2**	**Thresh**	**Gene**	**Consequence**
rs11591147	14	CRIP1, Apo B, PAFAH, LDL, CHD, HDL	0.80	rs11591147	1	3,203	0.98	0.5	*PCSK9*	Missense
rs55730499	1	CRIP1, PAFAH, SLAF6, PTPRS, EMBP, LDL, CHD, HDL	0.78	rs10455872	1	2,906	0.98	0.8	*LPA*	Intron
rs7412	6	cIAP-1, CLC4G, LRRK2, GRB14, SDC3, HO-1, Syntenin 1, CRIP1, gpIIbIIIa, TECK, Apo B, b2-Microglobulin, C1QBP, PAFAH, SLAF6, PSME1, FAIM3, TR, C1RL1, PGM1, KMT2C, TIGIT, BT3A3, MLF1, UB2D2, ihh, Sonic Hedgehog, CD5L, ADAM 9, NPFF, STOM, NR1H4, p15-INK4b, LAG-3, LDL, CHD, HDL	0.78	rs7412	1	2,771	0.98	0.7	***APOE***	Missense
rs4845625	30	Fibrinogen, CRP, Trypsin 2, LIRB4, Cathepsin B, SAA, IL-6 sRa, CHD, HbA1c, HDL, IL6	0.74	rs2228145	1	1,934	0.98	0.8	*IL6R*	Missense
rs964184	36	LRRK2, SELS, WNT5A, B4GT1, GRB14, GAS-6, LL-37, SPLC2, HO-1, CRIP1, UB2D3, gpIIbIIIa, Apo E, EMAP-2, Apo B, C1QBP, PAFAH, DR6, GGT2, MMP-8, IFIT2, TIGIT, GBP2, Apo A-I, STOM, LDL, CHD, HbA1c, Triglycerides	0.68	rs964184	1	3,234	0.98	0.8	*ZNF259 (APOA5)*	3' UTR
rs2954029	15	M-CSF R, IL-1 sRII, GAS-6, LL-37, SPLC2, gpIIbIIIa, Cystatin C, TECK, GDF-11/8, Apo B, C9, FSTL3, sE-Selectin, TAFI, CPN2, HBD-1, GGH, TLR5, Apo A-I, Galectin-3, BMI, LDL, CHD, HbA1c, HDL, Triglycerides	0.67	rs28601761	1	2,390	0.98	0.8	3' of *TRIB1*	Intergenic
rs507666	24	ACE, CLC4G, SECTM1, M-CSF R, EFNB2, CD8A, BGLR, HD-5, HO-1, CD46, gp130, soluble, TECK, Apo B, CCL28, SARP-2, LRP8, CD30 Ligand, sE-Selectin, GNS, TLR4:MD-2 complex, VEGF sR2, HSP 70, P-Selectin, IGF-I sR, NCAM-L1, sICAM-1, GIB, PAFAH, DAF, ILT-2, MO2R1, OX2G, sICAM-5, sICAM-2, SEM4D, Cripto, MANBA, GP116, ICAM4, DRB3, S100A13, GPNMB, Uromodulin, CATE, TLR5, MICA, PIGR, LGMN, IL-1 R4, CLM6, CM35H, FCRL1, Sema E, AT1B2, BTNL8, CRIP2, SCF, LDL, CHD, HDL, Triglycerides, 2hr Glucose adjBMI, Fasting Glucose adjBMI	0.66	rs507666	1	3,161	0.98	0.8	***ABO***	Intron
rs6905288	52	SPLC2, BMI, LDL, CHD, HbA1c, Triglycerides, WHRadjBMI, 2hr Glucose adjBMI, Fasting Insulin adjBMI	0.59	rs998584	1	2,426	0.98	0.5	*VEGFA*	Downstream
rs3184504	8	CO1A1, CARD9, TPSN, SECTM1, B4GT1, M-CSF R, C1QR1, CD8A, BGLR, Complement receptor type 2, SEM4A, Complement receptor type 1, TIMP-2, GDF-11/8, Fibrinogen, Midkine, DAN, DC-SIGN, IL-18 BPa, CD48, FCG3B, granzyme A, b2-Microglobulin, P-Selectin, NCAM-L1, sL-Selectin, IgD, DAF, sICAM-5, sICAM-2, SEM4D, PTPRS, CECR1, MFAP5, FAIM3, Semaphorin-7A, NKp46, IL-23, STAT1, KI2S2, CD7, VEGF-D, Gelsolin, RAB4A, KI2L3, BT3A3, Collagen Type III, CXCL16, soluble, VCAM-1, IL-1 sRI, TGF-b R III, I-TAC, CD5L, CLC4K, Cathepsin V, Kallikrein 7, Thrombopoietin Receptor, GCP-2, Lymphotoxin a1/b2, IP-10, C1-Esterase Inhibitor, GP1BA, CLM6, COLEC12, TRML2, Epo, Sema E, CRIP2, AZGP1, LAG-3, LDL, CHD, HbA1c, IFNg, TNFa, Triglycerides, WHRadjBMI, Fasting Insulin adjBMI	0.68	rs7310615	0.99	1,056	0.98	0.8	***SH2B3***	Intron
rs8107974	35	WNT5A, NEUR1, Hexosaminidase B, BGLR, CRIP1, Apo E, Apo B, Angiostatin, CATZ, Cathepsin D, prosaposin, GGT2, LIRB4, Cathepsin B, CATF, CREG1, GGH, ADAM 9, MIP-3b, LDL, T2D, CHD, HbA1c, HDL, Triglycerides, WHRadjBMI	0.73	rs58542926	0.99	1,934	0.98	0.8	***TM6SF2***	Missense
rs10512861	72	6Ckine, CHD	0.89	rs11925382	0.99	2,854	0.98	0.8	*DNAJC13*	Intron
rs2814993	53	DR6, GPNMB, LDL, CHD, HDL	0.85	rs76967117	0.99	2,371	0.98	0.8	*ILRUN*	Intron
rs15285	41	CRIP1, GGT2, LDL, CHD, HbA1c, WHRadjBMI	0.53	rs3208305	0.82	3,665	0.98	0.5	*LPL*	3' UTR
rs2246942	17	prosaposin, LGMN, CHD	0.97	rs1412444	0.60	3,060	0.98	0.5	*LIPA*	Intron
rs668948	29	Apo B, CHD, HDL	0.93	rs541041	0.60	2,294	0.98	0.8	5' of *APOB*	Intergenic
rs2972144	4	cIAP-1, CLC4G, WNT5A, C1QBP, PTPRS, LDL, T2D, CHD, HbA1c, Triglycerides, 2hr Glucose adjBMI, Fasting Glucose adjBMI, Fasting Insulin adjBMI	0.68	rs2943646	0.56	2,347	0.98	0.8	3' of *IRS1*	Intergenic
rs34855406	51	LGP2, CHD	0.98	rs2074164	0.45	1,297	0.98	0.7	*KAT2A/MLX*	Intron
rs6511720	5	Apo B, CHD	0.94	rs56289821	0.30	2,255	0.98	0.9	5' of *LDLR*	Intergenic

Abbreviations: BMI, Body mass index; HbA1c, Glycated haemoglobin; T2D, Type 2 diabetes; CHD, Coronary heart disease; HDL, High density lipoprotein; LDL, Low-density lipoprotein.

1. Gene names coloured in red are known pleiotropic loci
2. The causal gene for rs964184 has been refined to *APOA5* using eQTL data in other studies[362]
3. Loci are reported at the lowest prior 2 and threshold configuration where $PP_{coloc} \geq 0.5$ and $PP_{explained} \geq 0.8$, clusters may alter with more stringent prior and threshold configurations
4. The candidate variant at *SH2B3*, rs7310615, and the missense variant in the same gene, rs3184504, are in perfect LD ($R^2 = 1$)

A total of four loci (*APOE*, *TM6SF2*, *ABO* and *IRS1*) were shared between T2D and CHD, of these, both diseases colocalised with inflammatory proteins at the same candidate variant at *TM6SF2* and *IRS1*. At *ABO* the LD between rs529565 and rs507666 was moderate (R^2=0.39). Comparatively, at *APOE* the LD between the respective candidate variants rs429358 and rs7412 was negligible (R^2=0.01), indicating that the two variants are independent of one another. Interestingly, T2D was estimated to colocalise with inflammatory proteins at rs429358 whereas CHD colocalised with inflammatory proteins at rs7412 (**Tables 5.2** and **5.3**), suggesting independent mechanisms of action at each variant. Closer examination of the effects of both variants on all inflammatory proteins estimated to colocalise with T2D and CHD at *APOE*, showed a clear delineation in the effects of each variant with respect to the proteins they are estimated to colocalise with (**Figure 5.2**).

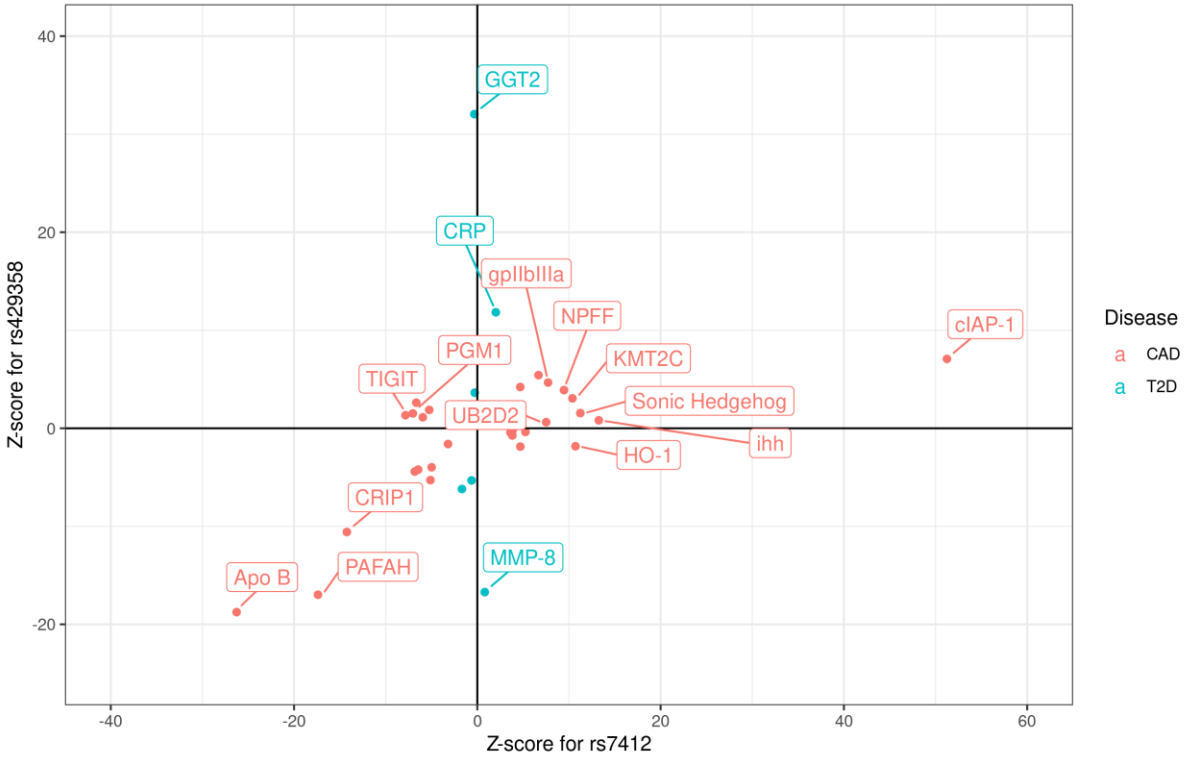


Figure 5.2: Scatterplot illustrating the effects of rs7412 and rs429358 on the set of inflammatory proteins which colocalised with T2D and CHD in *APOE*. The Z-score of the association between each variant and the inflammatory proteins are plotted. Proteins colocalising with T2D at rs429358 are shown in blue whereas proteins colocalising with CHD at rs7412 are shown in red. Any protein with an absolute Z-score greater than 10 with one of the variants is labelled. A clear delineation in the effects of each variant on the subset of proteins estimated to colocalise with T2D or CHD at either variant can be seen.
Abbreviations: CHD, Coronary heart disease; T2D, Type 2 diabetes; Apo B; Apolipoprotein B; PAFAH, Platelet-activating factor acetylhydrolase; CRIP1, Cysteine-rich protein 1; MMP-8, Matrix metalloproteinase-8; TIGIT, T-cell immunoreceptor with Ig and ITIM domains; PGM1, Phosphoglucomutase-1; UB2D2; Ubiquitin-conjugating enzyme E2 D2; HO-1, Heme oxygenase 1; ihh, Indian hedgehog protein; KMT2C, Histone-lysine N-methyltransferase 2C; NPFF, Pro-FMRFamide-related neuropeptide FF; gpIIbIIIa, Integrin alpha-IIb: beta-3 complex; CRP, C-reactive protein; GGT2, Inactive gamma-glutamyltranspeptidase 2; cIAP-1, Baculoviral IAP repeat-containing protein 2.

Of the 18 loci where CHD was estimated to colocalise with inflammatory proteins, candidate variants at 11 loci (61%) are in or nearby genes with diverse functions in lipid metabolism and homeostasis (*PCSK9, LPA, APOE, ZNF259 [APOA5], TRIB1, TM6SF2, LPL, LIPA, APOB, LDLR, ILRUN*). At each of these genes aside from *LDLR* and *LIPA*, CHD colocalised with at least one lipid measure, further highlighting their involvement in the regulation of lipid levels. Recent evidence[372,377] has suggested that two of these genes, *ZNF259* (*APOA5*) and *TRIB1* have pleiotropic associations with several metabolic traits ranging from body fat distribution[377] to lipid levels[372]. In addition to these, the candidate variants at *APOE*, *GCKR*, *TM6SF2*, *ABO* and *SH2B3* are established pleiotropic loci because of their associations with many diverse traits[370–373]. Closer examination of the number of proteins that colocalise with T2D and CHD highlighted that the majority of proteins, 95% and 85% respectively, colocalised with both diseases at a minimum of one pleiotropic locus (mixed + pleiotropic loci; **Table 5.4**). Only 5% and 15% of all proteins colocalise with T2D and CHD at non-pleiotropic loci.

Table 5.4: Number of proteins which colocalise with disease at the 3 pleiotropy classes

Disease	Number of proteins which colocalise at non-pleiotropic loci, (%)	Number of proteins which colocalise at mixed loci, (%) [a]	Number of proteins which colocalise at pleiotropic loci, (%) [b]	Total colocalised proteins
T2D	8 (5%)	11 (7%)	145 (88%)	164
CHD	28 (15%)	34 (19%)	119 (66%)	181

Abbreviations: T2D, Type 2 diabetes; CHD, Coronary heart disease

a. Proteins which colocalise at mixed loci do so at a combination of non-pleiotropic and pleiotropic loci. These proteins are not double counted in the other two groupings
b. *APOE*, *GCKR*, *TM6SF2*, *ABO* and *SH2B3* are pleiotropic loci

CHD was estimated to colocalise with inflammatory proteins at rs2228145 (Asp358Ala), the missense variant in *IL6R* (PP_{coloc} = 0.74; $PP_{explained}$ = 1). As outlined in **Chapter 4**, this variant is associated with HbA1c levels. These results demonstrate that HbA1c colocalises with IL-6 levels at Asp358Ala, providing further evidence of the link between IL-6 and HbA1c levels. Examining the wider associations of Asp358Ala with all 4,979 human proteins highlighted the specificity of Asp358Ala for IL-6R levels and showed associations with levels of galectin-3 binding protein (LG3BP), S100 Calcium Binding Protein A7 (S100A7), Calcium/Calmodulin Dependent Protein Kinase I (CAMK1) and Mannose-binding lectin (**Figure 5.3**). Only mannose-binding lectin was included as an inflammatory protein in this analysis but was not estimated to colocalise with CHD at rs2228145 (**Table 5.3**). Consistent with an effect of reduced IL6R-mediated inflammation, Asp358Ala was associated with -0.11 SD lower CRP levels (95% CI, -0.13, -0.08; $P=2x10^{-15}$) per copy of Asp358Ala.

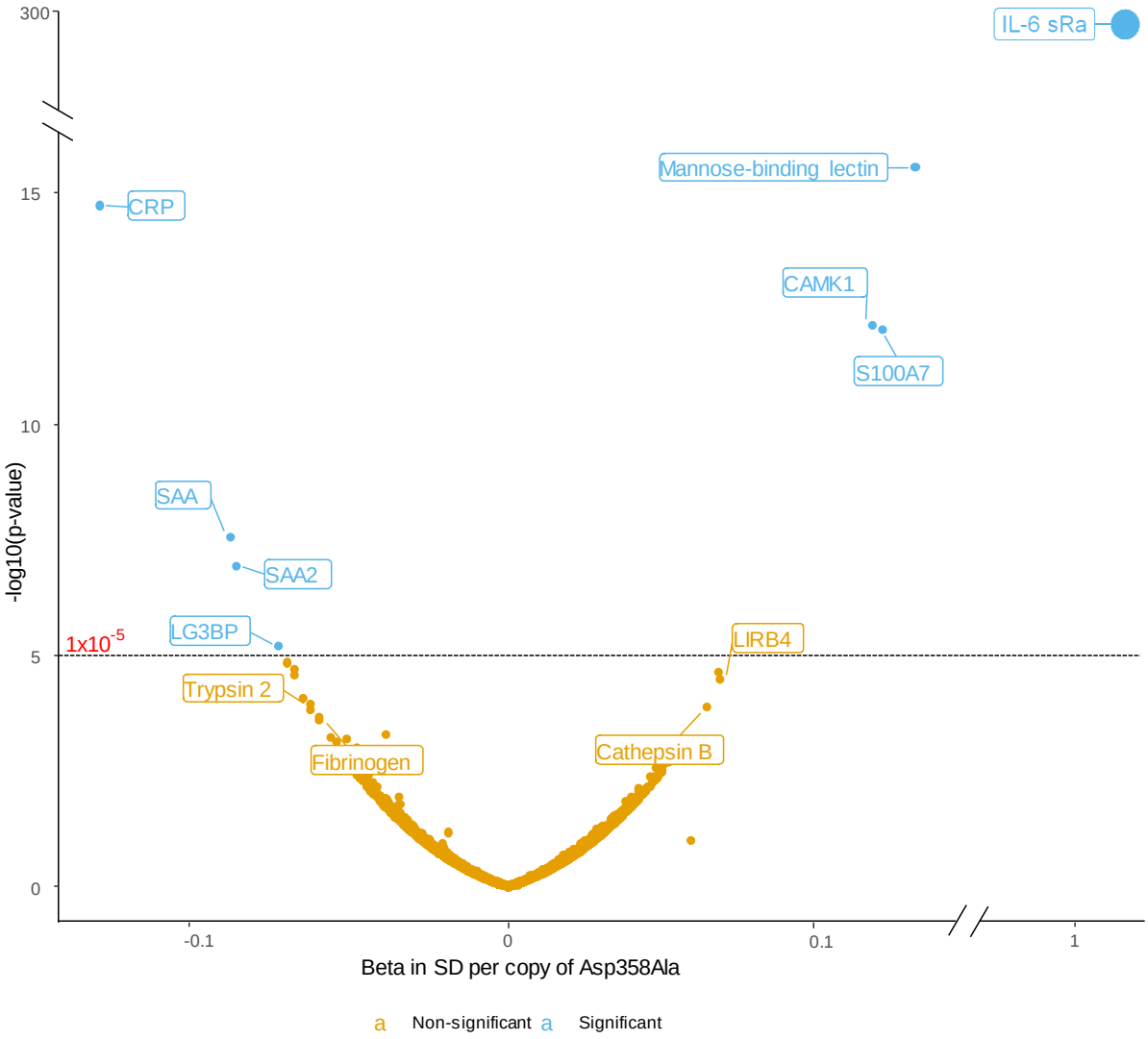
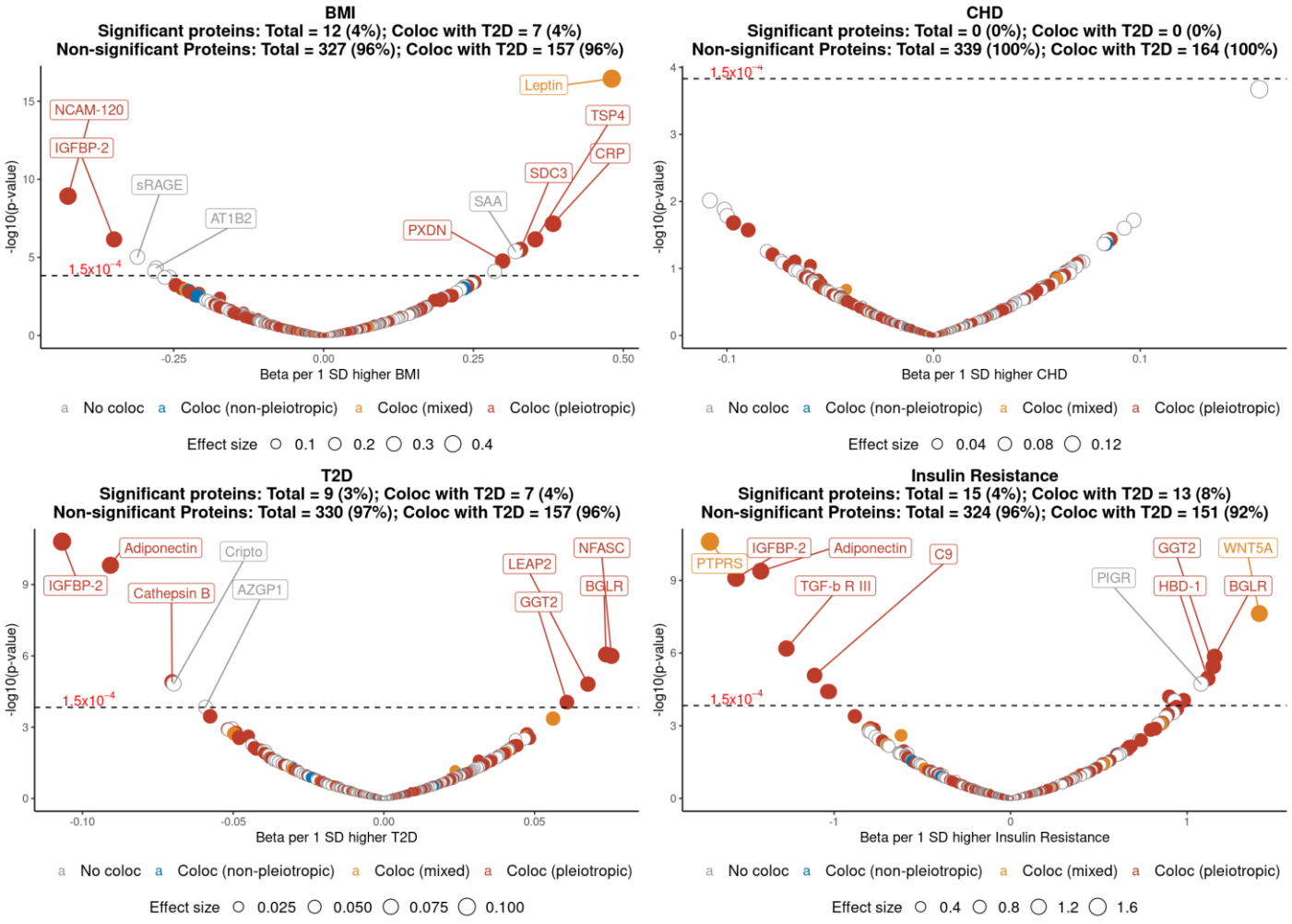


Figure 5.3: Volcano plot of the association between the *IL6R* missense variant (Asp358Ala; rs2228145) and all 4,979 human proteins from the SOMAscan v4 platform. Due to the large association of Asp358Ala with higher IL-6 sRa levels, the axes are broken to facilitate its display. Marker size for each protein is relative to its absolute effect size (beta) of Asp358Ala on levels of the respective protein. Proteins are coloured according to their significance with Asp358Ala. Proteins coloured in blue are significant at a Bonferroni significance threshold of $P < 1x10^{-5}$, denoted by the dashed line. All significant associations with Asp358Ala are labelled as well as any inflammatory proteins which were estimated to colocalise with CHD at Asp358Ala (**Table 5.3**).

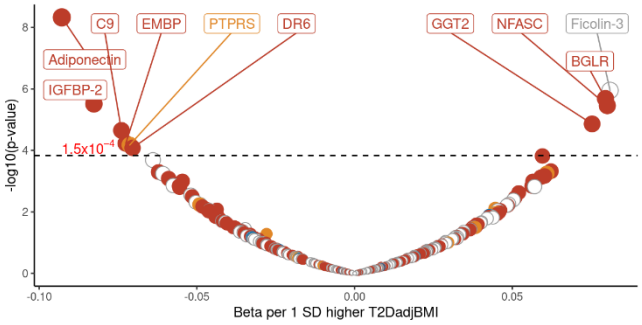
Abbreviations: IL-6 sRa, Interleukin-6 receptor subunit alpha; LIRB4, Leukocyte Immunoglobulin Like Receptor B4; CAMK1, Calcium/calmodulin-dependent protein kinase type 1; LG3BP, Galectin-3-binding protein; SAA, Serum amyloid A-1 protein; SAA2, Serum amyloid A-2 protein; CRP, C-reactive protein

Association of polygenic scores for cardiometabolic traits with inflammatory protein levels

The association of 13 polygenic scores for cardiometabolic risk factors and disease outcomes with inflammatory protein levels were estimated in up to 7,959 participants from Fenland. Overall, the polygenic scores were significantly associated ($P<1.5x10^{-4}$) with the levels of 100 inflammatory proteins, 29% of the total 339 proteins investigated. Of the 164 proteins which colocalised with T2D (**Table 5.2**), the levels of 69 proteins (42% of the total) were associated with at least one polygenic score (**Figure 5.4**). Of the proteins not significantly associated with any score, only Galectin-3, IL-1 receptor type 1 (IL-1 sR1), Osteocalcin, Trypsin-2, phospholipase A2 (GIB) and the TLR4:MD-2 complex colocalised with T2D at a non-pleiotropic locus. Thus, 94% of proteins not associated with any polygenic score colocalised with T2D at a minimum of one pleiotropic locus. The TLR4:MD-2 complex and GIB colocalised with T2D at *CTRB2* whereas Galectin-3 and Osteocalcin colocalised at *HPR* and *HLA-DRA* respectively.



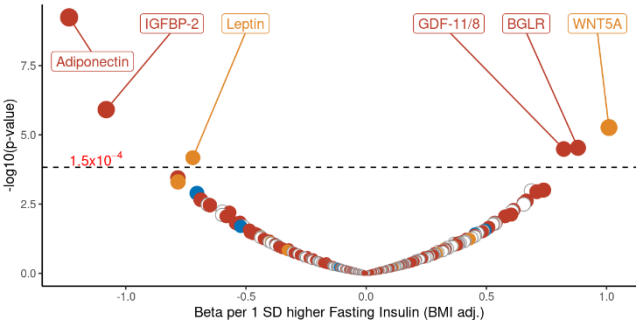
T2DadjBMI
Significant proteins: Total = 10 (3%); Coloc with T2D = 9 (5%)
Non-significant Proteins: Total = 329 (97%); Coloc with T2D = 155 (95%)



a No coloc a Coloc (non-pleiotropic) a Coloc (mixed) a Coloc (pleiotropic)

Effect size ○ 0.025 ○ 0.050 ○ 0.075

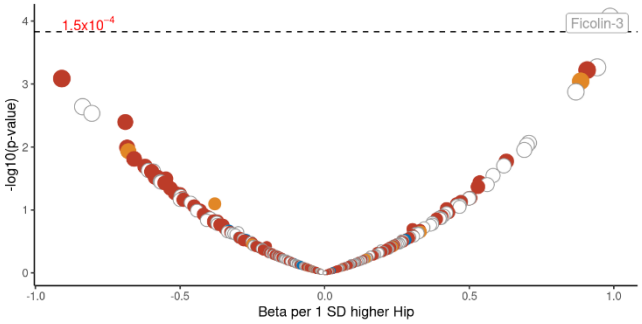
Fasting Insulin (BMI adj.)
Significant proteins: Total = 6 (2%); Coloc with T2D = 6 (4%)
Non-significant Proteins: Total = 333 (98%); Coloc with T2D = 158 (96%)



a No coloc a Coloc (non-pleiotropic) a Coloc (mixed) a Coloc (pleiotropic)

Effect size ○ 0.3 ○ 0.6 ○ 0.9 ○ 1.2

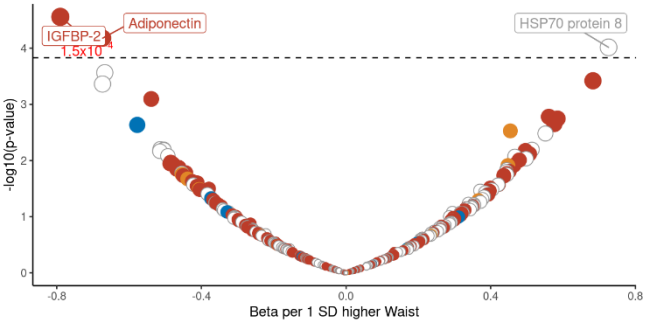
Hip
Significant proteins: Total = 1 (0%); Coloc with T2D = 0 (0%)
Non-significant Proteins: Total = 338 (100%); Coloc with T2D = 164 (100%)



a No coloc a Coloc (non-pleiotropic) a Coloc (mixed) a Coloc (pleiotropic)

Effect size ○ 0.25 ○ 0.50 ○ 0.75

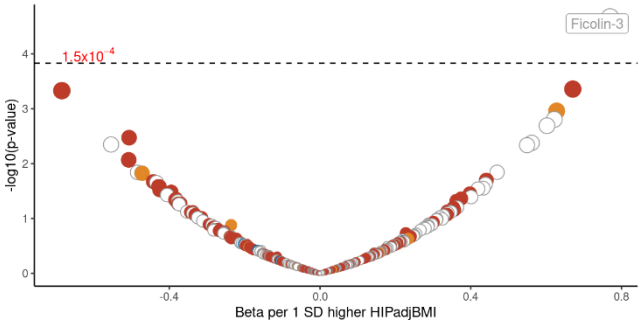
Waist
Significant proteins: Total = 3 (1%); Coloc with T2D = 2 (1%)
Non-significant Proteins: Total = 336 (99%); Coloc with T2D = 162 (99%)



a No coloc a Coloc (non-pleiotropic) a Coloc (mixed) a Coloc (pleiotropic)

Effect size ○ 0.2 ○ 0.4 ○ 0.6

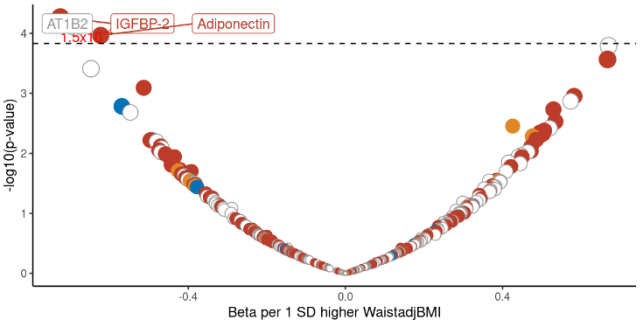
HIPadjBMI
Significant proteins: Total = 1 (0%); Coloc with T2D = 0 (0%)
Non-significant Proteins: Total = 338 (100%); Coloc with T2D = 164 (100%)



a No coloc a Coloc (non-pleiotropic) a Coloc (mixed) a Coloc (pleiotropic)

Effect size ○ 0.2 ○ 0.4 ○ 0.6

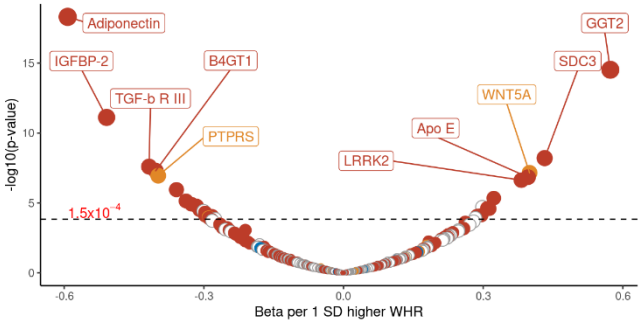
WaistadjBMI
Significant proteins: Total = 3 (1%); Coloc with T2D = 2 (1%)
Non-significant Proteins: Total = 336 (99%); Coloc with T2D = 162 (99%)



a No coloc a Coloc (non-pleiotropic) a Coloc (mixed) a Coloc (pleiotropic)

Effect size ○ 0.2 ○ 0.4 ○ 0.6

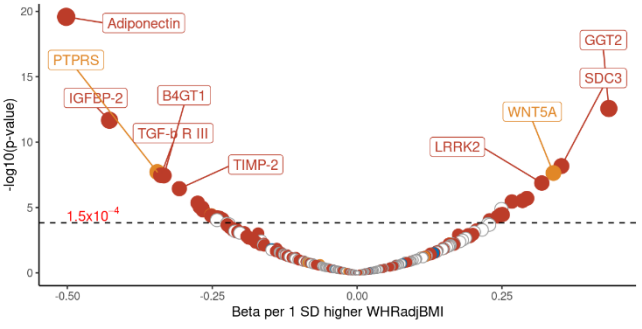
WHR
Significant proteins: Total = 31 (9%); Coloc with T2D = 27 (16%)
Non-significant Proteins: Total = 308 (91%); Coloc with T2D = 137 (84%)



a No coloc a Coloc (non-pleiotropic) a Coloc (mixed) a Coloc (pleiotropic)

Effect size ○ 0.1 ○ 0.2 ○ 0.3 ○ 0.4 ○ 0.5

WHRadjBMI
Significant proteins: Total = 31 (9%); Coloc with T2D = 29 (18%)
Non-significant Proteins: Total = 308 (91%); Coloc with T2D = 135 (82%)



a No coloc a Coloc (non-pleiotropic) a Coloc (mixed) a Coloc (pleiotropic)

Effect size ○ 0.1 ○ 0.2 ○ 0.3 ○ 0.4 ○ 0.5

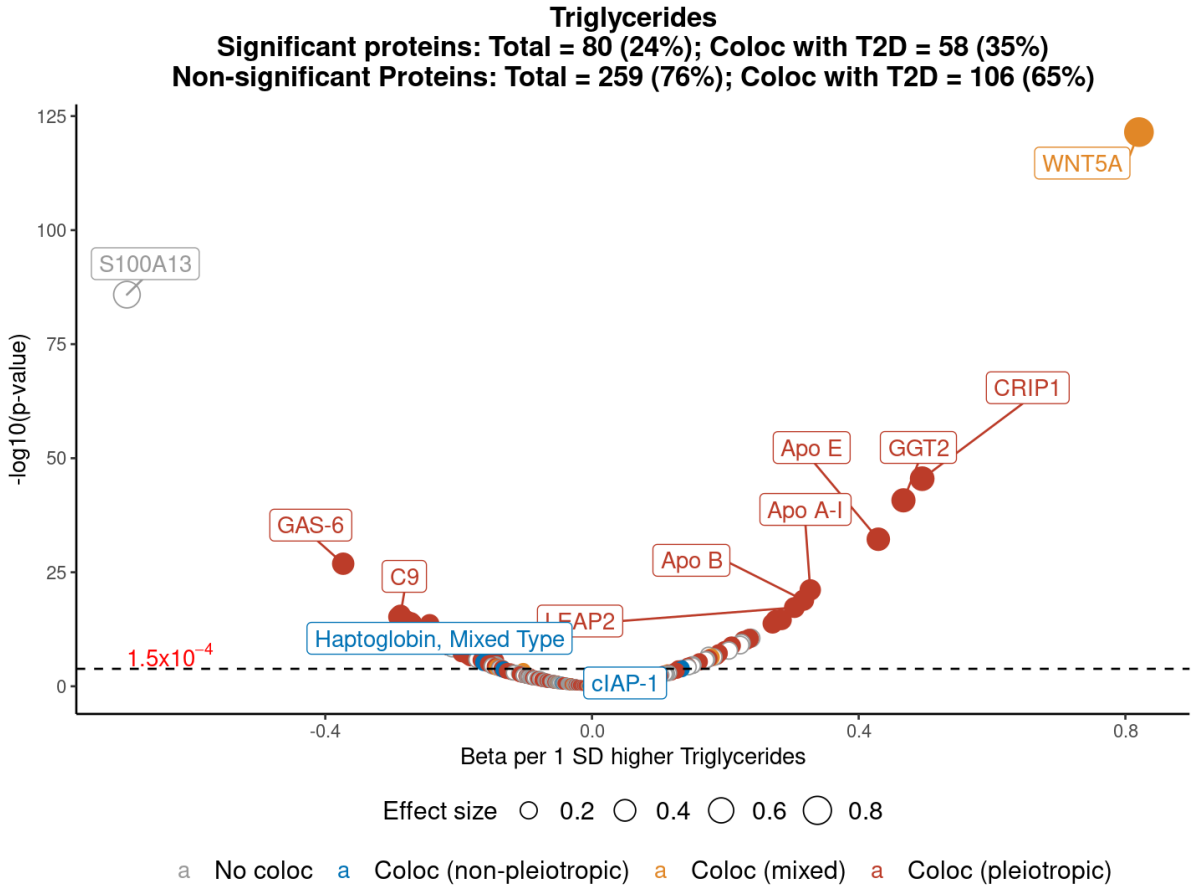
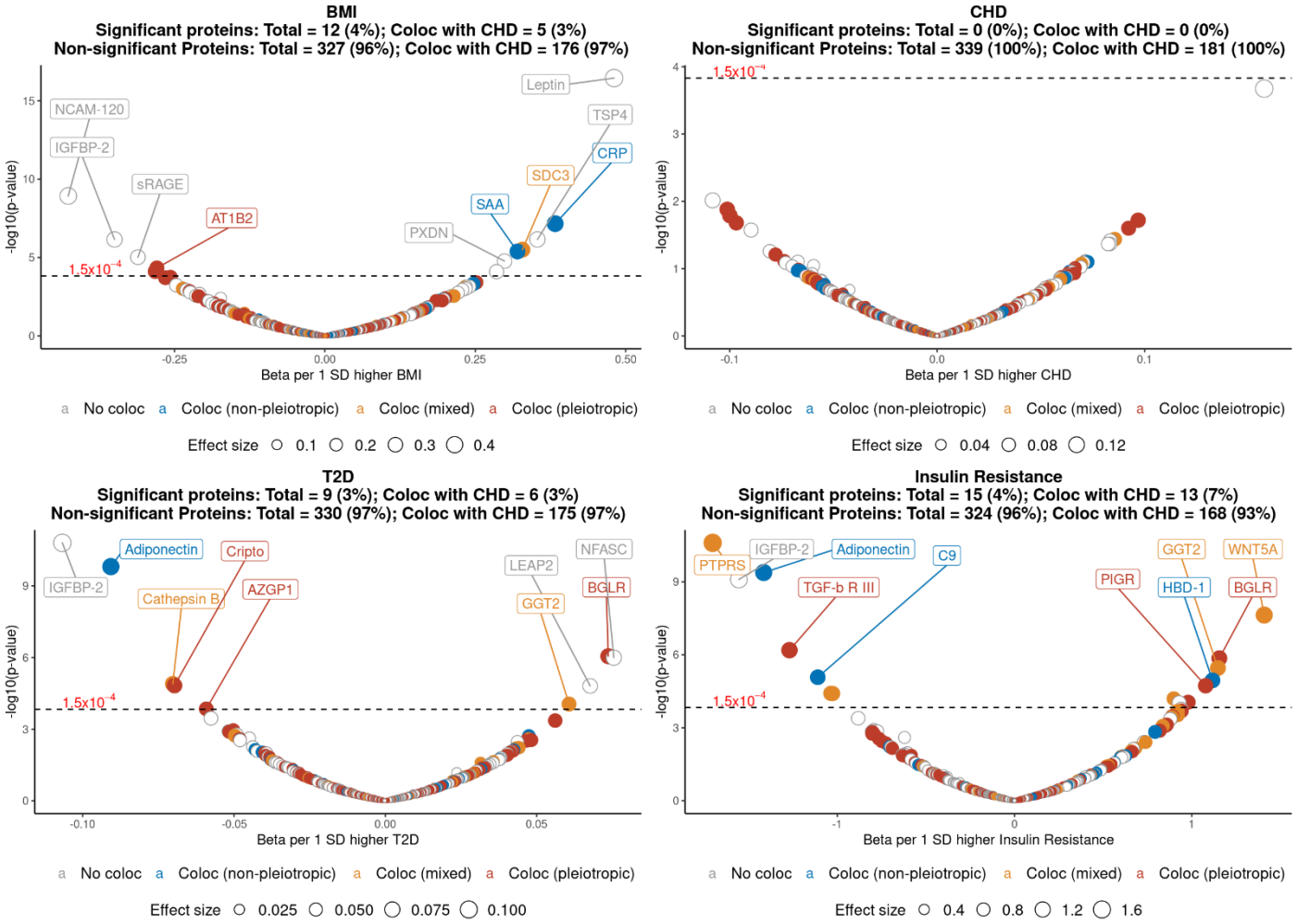


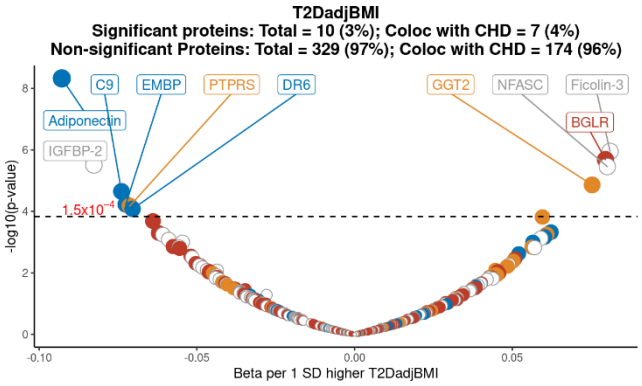
Figure 5.4: Volcano plots illustrating the association of polygenic scores for 13 cardiometabolic traits with levels of the 335 inflammatory proteins and four cytokines. The number of proteins significantly associated ($P<1.5x10^{-4}$) with each polygenic score is shown. Markers for each protein are sized according to their association with the respective polygenic score and coloured according to their colocalisation with T2D. Proteins that do not colocalise with T2D are shown as open grey circles, those that colocalise with T2D at non-pleiotropic loci are shown in blue, those that colocalise with T2D at both pleiotropic and non-pleiotropic loci are deemed "mixed" and shown in gold and those that colocalise with T2D at pleiotropic loci (*APOE*, *GCKR*, *TM6SF2* or *ABO*) are shown in red. The top 10 proteins significantly associated with each score are labelled as well as any proteins significantly associated with the polygenic score that also colocalised with T2D at a non-pleiotropic locus.
Abbreviations: Coloc, Colocalisation; SD, Standard deviation; T2D, Type 2 diabetes

The scores significantly associated with the levels of the largest number of proteins were the triglycerides, WHR and WHRadjBMI scores. The triglycerides score was significantly associated with the levels of 80 of the 339 inflammatory proteins, 24% of the total and the most of any score. Of these, 58 proteins colocalised with T2D, 96% of which colocalised at pleiotropic loci. The triglycerides score was significantly associated with levels of Haptoglobin (mixed type) and Baculoviral IAP repeat-containing protein 2 (cIAP-1), two out of eight proteins which colocalised with T2D at non-pleiotropic loci (**Figure S5.2**). Thus, the triglycerides score was significantly associated with 35% of the proteins estimated to colocalise with T2D. Similarly, the WHRadjBMI score was associated with the levels of 31 proteins, the second-highest number of any score. Of these, 29 proteins were estimated to colocalise with T2D.

Thus, the WHRadjBMI score was significantly associated with 18% of the proteins estimated to colocalise with T2D.

Of the 181 proteins which colocalised with CHD (**Table 5.3**), the levels of 67 proteins (36% of the total) were associated with at least one polygenic score (**Figure 5.5**). Of the proteins not significantly associated with any score, 9 colocalised with CHD at non-pleiotropic loci. Thus, 92% of proteins not associated with any polygenic score colocalised with CHD at a minimum of one pleiotropic locus. Similar to the T2D results and in line with the finding that a considerable proportion of the loci identified by colocalisation with CHD are involved in lipid metabolism, the triglycerides score was significantly associated with levels of 54 (30%) of the proteins that colocalised with CHD, the most of any score (**Figure 5.5**). Indeed, the triglycerides score was significantly associated with 16 (57%) of the 28 proteins which colocalised with CHD at non-pleiotropic loci (**Figure S5.3**). In line with the T2D results, the WHR and WHRadjBMI scores were associated with levels of the second-highest number of proteins (**Figure 5.5**), 13% of the proteins which colocalised with CHD respectively.

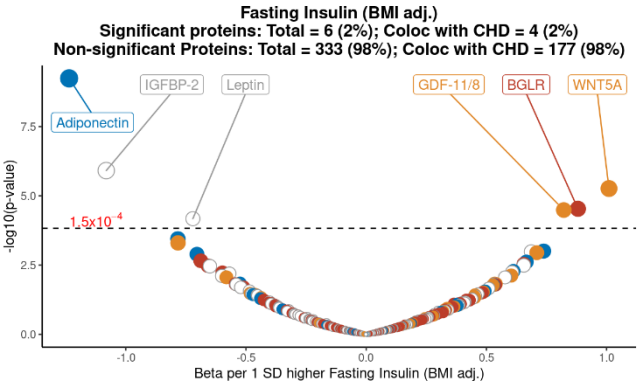




T2DadjBMI
Significant proteins: Total = 10 (3%); Coloc with CHD = 7 (4%)
Non-significant Proteins: Total = 329 (97%); Coloc with CHD = 174 (96%)

a No coloc a Coloc (non-pleiotropic) a Coloc (mixed) a Coloc (pleiotropic)

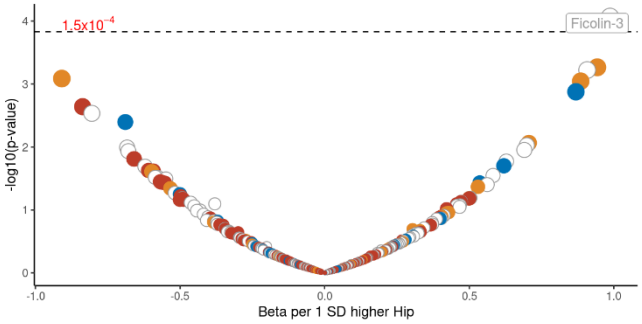
Effect size ○ 0.025 ○ 0.050 ○ 0.075



Fasting Insulin (BMI adj.)
Significant proteins: Total = 6 (2%); Coloc with CHD = 4 (2%)
Non-significant Proteins: Total = 333 (98%); Coloc with CHD = 177 (98%)

a No coloc a Coloc (non-pleiotropic) a Coloc (mixed) a Coloc (pleiotropic)

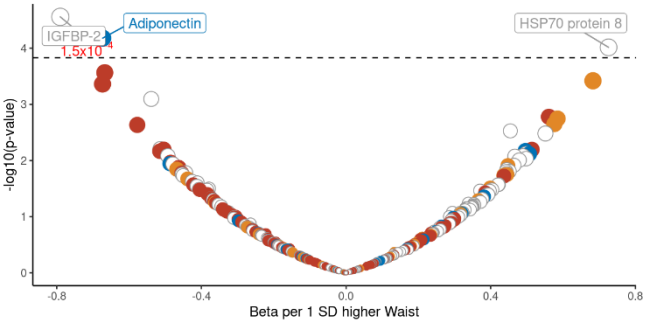
Effect size ○ 0.3 ○ 0.6 ○ 0.9 ○ 1.2



Hip
Significant proteins: Total = 1 (0%); Coloc with CHD = 0 (0%)
Non-significant Proteins: Total = 338 (100%); Coloc with CHD = 181 (100%)

a No coloc a Coloc (non-pleiotropic) a Coloc (mixed) a Coloc (pleiotropic)

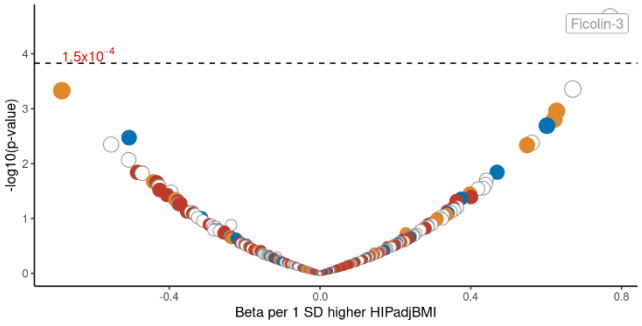
Effect size ○ 0.25 ○ 0.50 ○ 0.75



Waist
Significant proteins: Total = 3 (1%); Coloc with CHD = 1 (1%)
Non-significant Proteins: Total = 336 (99%); Coloc with CHD = 180 (99%)

a No coloc a Coloc (non-pleiotropic) a Coloc (mixed) a Coloc (pleiotropic)

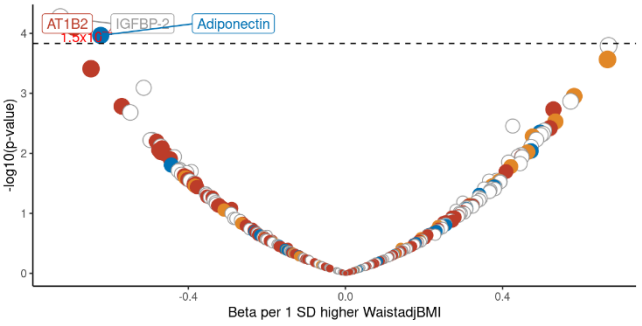
Effect size ○ 0.2 ○ 0.4 ○ 0.6



HIPadjBMI
Significant proteins: Total = 1 (0%); Coloc with CHD = 0 (0%)
Non-significant Proteins: Total = 338 (100%); Coloc with CHD = 181 (100%)

a No coloc a Coloc (non-pleiotropic) a Coloc (mixed) a Coloc (pleiotropic)

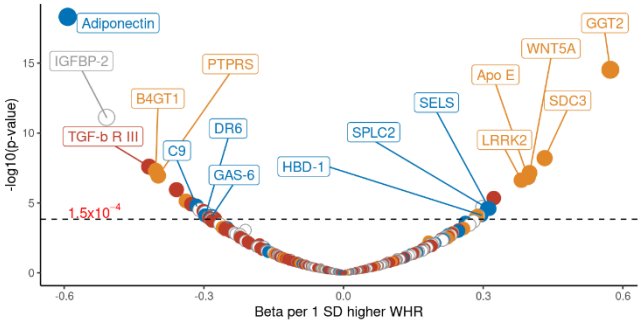
Effect size ○ 0.2 ○ 0.4 ○ 0.6



WaistadjBMI
Significant proteins: Total = 3 (1%); Coloc with CHD = 2 (1%)
Non-significant Proteins: Total = 336 (99%); Coloc with CHD = 179 (99%)

a No coloc a Coloc (non-pleiotropic) a Coloc (mixed) a Coloc (pleiotropic)

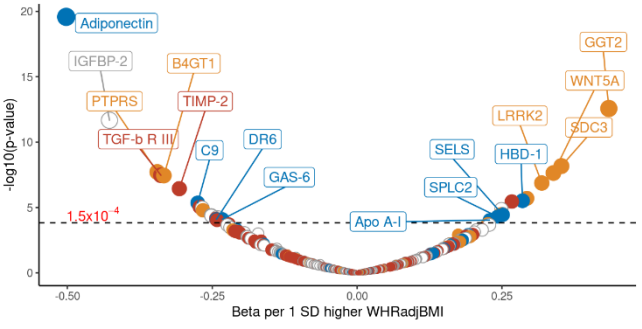
Effect size ○ 0.2 ○ 0.4 ○ 0.6



WHR
Significant proteins: Total = 31 (9%); Coloc with CHD = 24 (13%)
Non-significant Proteins: Total = 308 (91%); Coloc with CHD = 157 (87%)

a No coloc a Coloc (non-pleiotropic) a Coloc (mixed) a Coloc (pleiotropic)

Effect size ○ 0.1 ○ 0.2 ○ 0.3 ○ 0.4 ○ 0.5



WHRadjBMI
Significant proteins: Total = 31 (9%); Coloc with CHD = 23 (13%)
Non-significant Proteins: Total = 308 (91%); Coloc with CHD = 158 (87%)

a No coloc a Coloc (non-pleiotropic) a Coloc (mixed) a Coloc (pleiotropic)

Effect size ○ 0.1 ○ 0.2 ○ 0.3 ○ 0.4 ○ 0.5

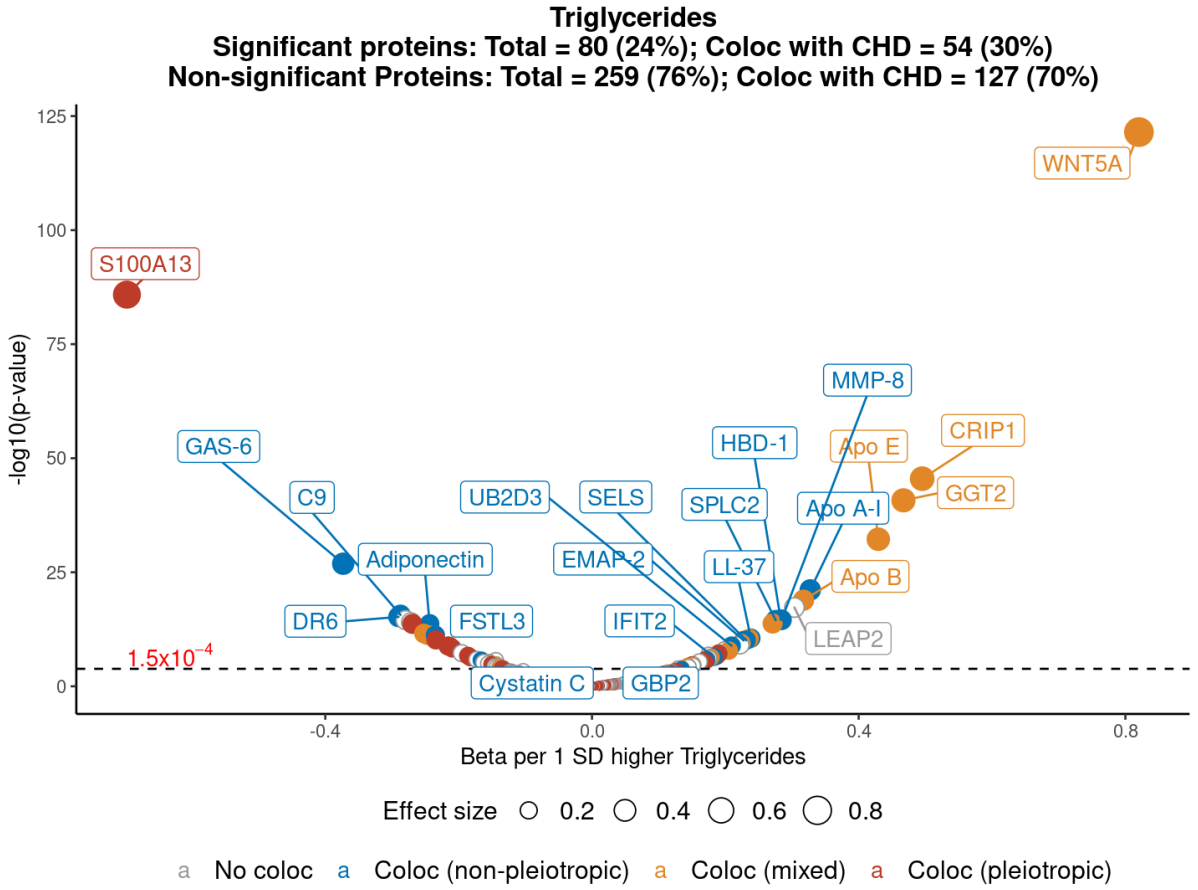


Figure 5.5: Volcano plots illustrating the association of polygenic scores for 13 cardiometabolic traits with levels of the 335 inflammatory proteins and four cytokines. The number of proteins significantly associated ($P<1.5x10^{-4}$) with each polygenic score is shown. Markers for each protein are sized according to their association with the respective polygenic score and coloured according to their colocalisation with CHD. Proteins that do not colocalise with CHD are shown as open grey circles, those that colocalise with CHD at non-pleiotropic loci are shown in blue, those that colocalise with CHD at both pleiotropic and non-pleiotropic loci are deemed "mixed" and shown in gold and those that colocalise with CHD at pleiotropic loci (*APOE*, *SH2B3*, *TM6SF2* or *ABO*) are shown in red. The top 10 proteins significantly associated with each score are labelled as well as any proteins significantly associated with the polygenic score that also colocalised with CHD at a non-pleiotropic locus.
Abbreviations: Coloc, Colocalisation; SD, Standard deviation; CHD, Coronary heart disease

Gene-set enrichment analysis

Proteins that colocalised with T2D were significantly enriched in 22 gene-sets across 8 categories (**Figure 5.6** and **Table S5.3**). The most significant of these was blood protein levels in GWAS catalog (overlap = 19%; adjusted $P = 1.5x10^{-5}$). This significant enrichment was also shared by proteins which colocalised with CHD with a greater degree of overlap (overlap = 23%; adjusted $P = 7.3x10^{-8}$; **Table S5.4**). Proteins that colocalised with T2D were significantly enriched in gene-sets related to liver development and function (Cairo liver development DN: overlap = 38%; adjusted P = 0.002; HSIAO Liver specific genes: overlap = 33%; adjusted P = 0.01) were significantly enriched. Interestingly, proteins that colocalised with both diseases were significantly enriched in proteins known to function in the complement cascade. Both

T2D and CHD showed significant enrichment in the human complement system gene-set from Wikipathways (**Table S5.3 and S5.4**), overlapping the gene-set by 30% and 36% respectively (T2D adjusted P = 0.03; CHD adjusted P = 0.002). Proteins colocalising with T2D were also significantly enriched in the KEGG complement and coagulation cascade gene-set (overlap = 39%; adjusted P = 0.01) whereas CHD proteins were enriched in the Hallmark complement gene-set (overlap = 25%; adjusted P = 0.05).

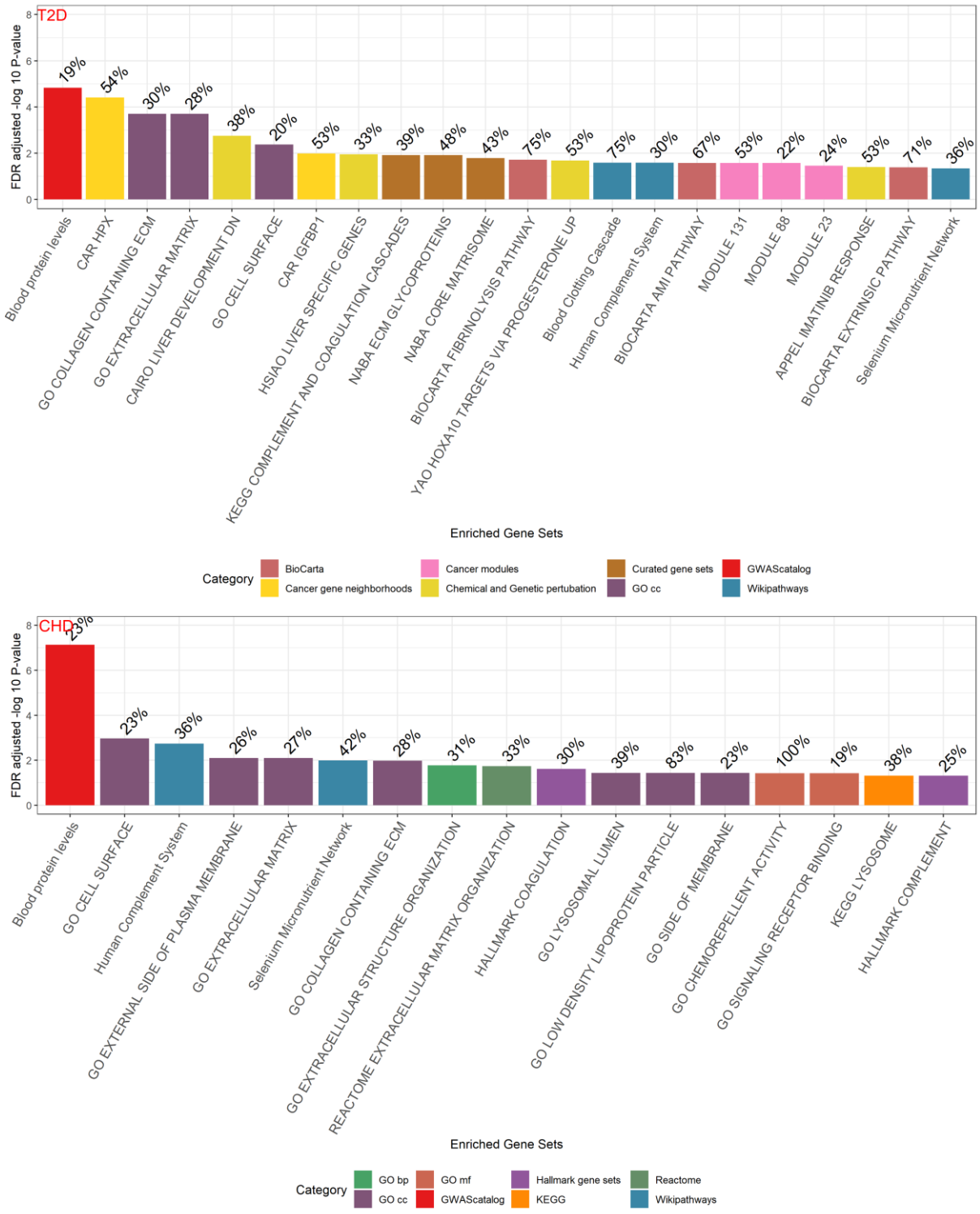


Figure 5.6: Bar plot showing the gene-sets that were significantly enriched in proteins that colocalised with either T2D or CHD. T2D results are shown in the top panel whereas CHD results are shown in the lower panel. The percentage overlap of the colocalised proteins with the total genes in each respective gene-set is shown above each bar. Colours correspond to the database each gene-set originated from.

Abbreviations: GO, Gene ontology terms; BP, Biological processes; CC, Cellular compartment; MF, Molecular function; KEGG, Kyoto Encyclopaedia of Genes and Genomes; ECM, Extracellular matrix; T2D, Type 2 diabetes; CHD, Coronary heart disease; FDR, False discovery rate

5.4: Discussion

To estimate whether there was evidence for a shared genetic basis underlying the inflammatory component of cardiometabolic diseases, this study used large-scale proteomic data for up to 1,358 inflammatory proteins within a Bayesian multi-trait colocalisation framework. The benefit of such an analysis is that trait colocalisation is fine mapped to a putative causal variant which may therefore tag an aetiologically relevant pathway[362]. Finally, the association of inflammatory proteins with polygenic scores for established cardiometabolic risk factors was estimated to determine whether risk factors were associated with levels of inflammatory proteins, thus mitigating the likelihood that inflammatory proteins levels are independent risk factors for cardiometabolic diseases.

The vast majority of loci with evidence for colocalisation between inflammatory proteins and cardiometabolic diseases were well-known pleiotropic loci, such as *APOE*, *ABO*, *TM6SF2*, *SH2B3* or *GCKR* and have been shown to associate with a wide of metabolic and other phenotypes, including anthropometric and lipid traits[370–373,377]. This pleiotropic effect is exemplified by the finding that the majority of proteins which colocalise with T2D and CHD did so at established pleiotropic loci.

Interestingly, the results of this study show that T2D and CHD colocalise with inflammatory proteins with robust evidence for colocalisation at rs429358 and rs7412, two independent missense variants (R^2 = 0.01) in apolipoprotein E (*APOE*). Allelic combinations of these two variants constitute the three main alleles of the *APOE* gene (ε2, ε3 and ε4), each of which have differential risk for diverse diseases[378]. The ε4 allele represents the C,C genotype for both variants and is associated with the highest APOE protein levels and concurrent 15-fold higher risk of Alzheimer's disease[378]. Phenome-wide analyses have demonstrated that the ε4 genotype is associated with higher hypercholesterolaemia and CHD risk, consistent with an effect of APOE on lipid levels[378]. In contrast, the same analysis demonstrated a significant protective effect of the ε4 genotype on both T2D and obesity[378]. This finding is interesting in the context of the results of this study as the two variants that make up the ε4 genotype may mediate disease risk via distinct pathways. It is likely that T2D risk is lowered due to the effect of ε4 on obesity risk while CHD risk is raised due to the effect of ε4 on lipid levels, however, more thorough investigations are required to fully assess this relationship. Taking this into consideration and as outlined in **Chapter** 1, the pleiotropic effects of these loci hinder the straight-forward interpretation of their effects on single traits as the locus may affect the phenotype of interest via many diverse pathways[62,73]. These findings do, however, suggest that a considerable proportion of the inflammatory response in cardiometabolic diseases may be being mediated by the effects of pleiotropic loci.

The findings of this study show that the majority of proteins colocalise with cardiometabolic diseases at pleiotropic loci and a proportion are also associated with established risk factors. Therefore, the emphasis from a drug target identification standpoint must be on those proteins which colocalise at non-pleiotropic loci and are not associated with established cardiometabolic risk factors. The results of this study show that only 6 and 10 proteins, 4% and 5% of all proteins which colocalised with T2D and CHD respectively, fell into this category. Of these, galectin-3, has been highlighted as a potential therapeutic target secreted by macrophages that links inflammation and insulin resistance that may be used to improve insulin sensitivity[379,380]. Galectin-3 colocalised with T2D and CHD at variants downstream of *HPR* and *TRIB1*, respectively. Studies have shown that haptoglobin-related protein, the protein product of *HPR* is a circulating ligand of galectin-3[381]. Galectin-3 levels have been shown to induce macrophage chemotaxis and are significantly higher in obese individuals when compared to lean individuals and are positively correlated with both BMI and HOMA-IR[379]. Consistent with an effect on insulin resistance, galectin-3 was shown to inhibit insulin-stimulated glucose transport in *in vitro* experiments in myocytes and adipocytes[379]. In addition, experiments in murine hepatocytes showed that galectin-3 blocked the actions of glucagon on hepatic glucose output[379]. Further *in vitro* experimentation showed that galectin-3 induces insulin desensitisation through insulin receptor binding, leading to abolished signalling. *In vitro* knockout experiments demonstrated lower pro-inflammatory gene expression in adipose tissue and improved glucose tolerance in obesity[379]. While this evidence suggests that galectin-3 may be a useful target to simultaneously lower inflammation and improve insulin resistance in obesity, it remains to be determined whether haptoglobin-related protein is the major ligand of galectin-3 and how this relationship may mediate T2D and CHD risk.

Of particular interest to this thesis, the missense variant, Asp358Ala (rs2228145), in *IL6R* was highlighted as a candidate variant where inflammatory proteins robustly colocalise with CHD. In line with previous results, Asp358Ala is associated with lower CHD risk[88] and, as presented in **Chapter 4**, is associated with HbA1c levels. In contrast to the findings in **Chapter 4**, T2D was not estimated to colocalise with IL-6 levels at Asp358Ala. However, closer examination of the magnitude of the meta-analysis effect estimate between Asp358Ala and T2D in **Chapter 4** shows that estimate (OR = 0.98) is comparatively small, thus requiring the very large sample size (260,614 cases and 1,350,640 controls) used to detect the association. As described in **Chapter 3**, $P \leq 1x10^{-4}$ is a commonly used prior probability that a variant is associated with trait of interest and is therefore sufficiently powered for use in colocalisation analyses. In the T2D dataset used in this study, the Asp358Ala variant was associated with T2D risk at a nominal significance threshold (OR = 0.99; 95% CI, 0.97, 1; P = 0.02). Therefore, the apparent lack of colocalisation is likely to be due to a lack of statistical power to detect the colocalisation rather

than a biological mechanism. This study also showed that CHD was estimated to colocalise with HDL levels at Asp358Ala. This is line with results of previous studies showing an association between Asp358Ala and higher HDL levels[330], an effect consistent with lower CHD risk[382]. Consistent with an effect on lower IL-6R mediated inflammation, Asp358Ala was associated with lower CRP, serum amyloid A (SAA and SAA2) [383,384], and fibrinogen[88,160] levels. SAA is secreted as part of the acute phase response where it replaces ApoA1 as the predominating apolipoprotein on HDL particles[385]. Inflammation lowers HDL levels and suppresses cholesterol efflux, the process by which cholesterol is removed from cells such as macrophages and excreted via the liver[386]. The presence of SAA is thought to contribute to this by lowering HDL cholesterol efflux capacity[386] thereby raising cardiovascular risk[387]. Overall, the inflammatory proteins which colocalise with CHD at Asp358Ala provide greater insight into the mechanisms underlying the role of IL-6 levels in CHD.

While considering the candidate variants where cardiometabolic diseases and inflammatory proteins are estimated to colocalise is valuable to identify pathways of interest, closer examination of the association between inflammatory proteins and established cardiometabolic risk factors can contextualise the inflammatory response within the aetiology of cardiometabolic diseases. The results of this study show that of the proteins that colocalise with T2D and CHD, 38% and 33% respectively were associated with at least one polygenic score for a cardiometabolic risk factor. Polygenic scores for triglycerides, WHR and WHRadjBMI were estimated to be associated with the levels of the greatest number of inflammatory proteins. Recent evidence from genetic studies has suggested that this is consistent with an insulin resistance phenotype[244,351,388,389] which is associated with higher adiposity, partly mediating the link between obesity and cardiometabolic diseases[244]. The hallmarks of an insulin resistant phenotype are higher fasting insulin adjusted for BMI[346], higher triglyceride levels and lower HDL levels[33,345,390]. In line with this, polygenic scores for insulin resistance and fasting insulin adjusted for BMI were associated with the levels of some inflammatory proteins, although not as many as the triglycerides and WHRadjBMI scores. These results therefore suggest that the levels of a proportion of the inflammatory proteins which colocalise with cardiometabolic diseases are governed by established cardiometabolic risk factors. As a result, these inflammatory proteins cannot be considered as independent risk factors for cardiometabolic diseases and are therefore not suitable to be used in an MR setting as they would invalidate the instrumental variable assumptions[62,73].

Overall, the results of gene-set enrichment analyses showed that inflammatory proteins which colocalised with cardiometabolic diseases were significantly enriched in proteins involved in the complement cascade. The human complement cascade is a central feature of the innate

immune system which is activated via three distinct pathways namely the classical, alternative and lectin pathways each of which lead to the formation of the membrane attack complex (MAC)[391]. Prospective observational epidemiological studies have shown that levels of complement component C3, the major component of the complement system, are associated with incident T2D and CHD, insulin resistance, abdominal obesity, lower HDL levels and higher fasting and 2-hour glucose levels in obesity[392–394]. As yet, the activation pathway mediating the higher levels of C3 in both T2D and CHD is unknown. Future investigation using genetic methods robust to reverse causality and residual confounding may help to elucidate the role of the complement cascade in cardiometabolic diseases.

In summary, several lines of evidence indicate that inflammatory proteins are unlikely to be independent causal risk factors for cardiometabolic diseases. Firstly, inflammatory proteins only colocalised with cardiometabolic diseases at loci which were already established T2D and CHD sentinels. Secondly, established cardiometabolic risk factors were associated with levels of approximately a third of all proteins which colocalised with T2D and CHD. Therefore, the levels of these proteins likely to be mediated by cardiometabolic risk factors instead of the proteins being independent risk factors for disease. Finally, the vast majority of inflammatory proteins colocalised with cardiometabolic diseases at established pleiotropic loci, thus making it difficult to pinpoint the pathways through-which these proteins act to modulate risk of cardiometabolic diseases.

This study has limitations. Firstly, the liberal selection criteria for the inclusion of proteins involved in “immune-related processes” likely led to the inclusion of proteins which function in the immune system but may not be involved in inflammation per se. This may have contributed noise to the analysis. Secondly, while only including cis-acting pQTLs for inflammatory proteins in the analysis may have aided the interpretability of the results, this may have come at the expense of missing potentially important loci where cardiometabolic diseases may have colocalised with inflammatory proteins. However, the reason for focussing the analysis on cis variants was based on proximal biology, i.e. having a high prior for the biological pathway involved. The interpretation of the biological pathways tagged by trans-pQTLs would have been much less clear, as they may affect protein levels and disease risk through different mechanisms. In addition, there are very few trans-pQTLs which are specific for protein levels that may have provided further insights, only those encoding known ligands, binding partners or receptors of inflammatory proteins may have provided further insights. In addition, the use of a permissive locus inclusion threshold of $P \leq 1x10\text{-}5$ suggests that only trans-pQTLs or loci which showed weak evidence ($P > 1x10\text{-}5$) for association with the disease of interest may have been missed. Thirdly, prior to estimating the genetic correlation between inflammatory proteins

and cardiometabolic traits, any variants with very large effect estimates were removed from the analysis. This was done to prevent variants such as these having a disproportionate influence on the LD score regression estimates[287], therefore biasing the result. However, as is the case with many molecular biomarkers, a single cis-acting variant (often in the protein encoding gene itself) may have a disproportionately large effect on levels of the protein. Exclusion of such variants may lead to underestimation of the trait heritability thereby also underestimating the true correlation between this protein and other traits. Fourth, the HyPrColoc method assumes a single causal variant in a given region[362], a limitation of most colocalisation methods[290]. While this study investigated comparatively small genomic regions, this assumption is more likely to be violated as the number of included traits increases. However, simulations using HyPrColoc have demonstrated that multiple independent variants in a region rarely invalidate the results, as long as the secondary variants do not explain more trait variation than the sentinel variant[362]. Future work could, however, make use of summary statistics conditioned on other independent signals within a locus to aid the estimation of the causal variant. Fifth, while HyPrColoc is able to consider an LD matrix to help resolve putative candidate variants in regions with complex LD, one was not provided. When using an LD matrix, a matrix of pairwise phenotypic correlation as well as a matrix of sample overlap must also be supplied[362]. As GWAS summary statistics from publicly available datasets were used, the estimation of pairwise phenotypic correlations was not possible. Finally, this analysis classified loci as pleiotropic based on published literature and was not based on a strict definition, as such a definition is still debated. Instead, only loci that are hallmark pleiotropic loci were considered as such, a potentially conservative approach. This does not rule out the possibility that loci such as TRIB1, which is associated with many lipid species and anthropometric traits[372,377,395–398], should not be considered as pleiotropic as well.

Chapter 6: Incretins and the incretin effect

Contributions and collaborations

The work in outlined in this chapter is all my own aside from the systematic literature search for genetic variants associated with incretin levels. This formed part of the work done by Robert Hansford, who's Wellcome Trust PhD rotation I helped supervise. However, all discussion of the results is my own.

6.1: The physiological roles of incretins and their alterations in obesity and T2D

The physiological roles of incretins

Until recently, our knowledge of glucose homeostasis was based upon the discoveries of insulin and glucagon. Insulin, secreted by β-cells in the pancreatic islet of Langerhans, has a hypoglycaemic effect on glucose levels after a meal[399,400]. Not long afterwards, glucagon secretion from the α-cells of the islet of Langerhans was discovered to reverse this effect, thereby raising glucose levels between meals[401]. Together, these hormones coordinate to maintain glucose homeostasis[402].

Following this, experimental evidence showed that plasma insulin levels were significantly higher upon the ingestion of glucose, compared to intravenous administration in humans[403,404]. This suggested that the source of this insulin potentiation was unlikely to be of pancreatic origin but rather secreted from the gut[405]. Glucose-dependent insulinotropic polypeptide (GIP) was the first of the incretins to be discovered[406–408]. Shortly afterwards a second incretin was discovered, glucagon-like peptide-1 (GLP-1), due to its structural similarities with glucagon[409,410]. The coordinated actions of GIP and GLP-1 on glucose-dependent insulin potentiation is now known as the "incretin effect", defined as the difference in pancreatic β-cell insulin secretion between oral and intravenously delivered glucose[411–413].

The absorption of glucose by sodium-dependent glucose transporter (SGLT) receptors on epithelial cells in the small intestine triggers GLP-1 and GIP secretion from L- and K-cells respectively (**Figure 6.1**)[402]. GLP-1 and GIP bind to their respective receptors, GLP-1R and GIPR, on pancreatic β-cells amplifying insulin secretion and contributing to the maintenance of glucose homeostasis[402]. GLP-1R expression is widely expressed in the pancreatic β-cells, central nervous system (CNS), kidneys liver, heart and gut[414]. Similarly, GIPR expression is much more ubiquitous, and is found in the CNS, stomach, heart and adipose tissue, as well as in the pancreas[415].

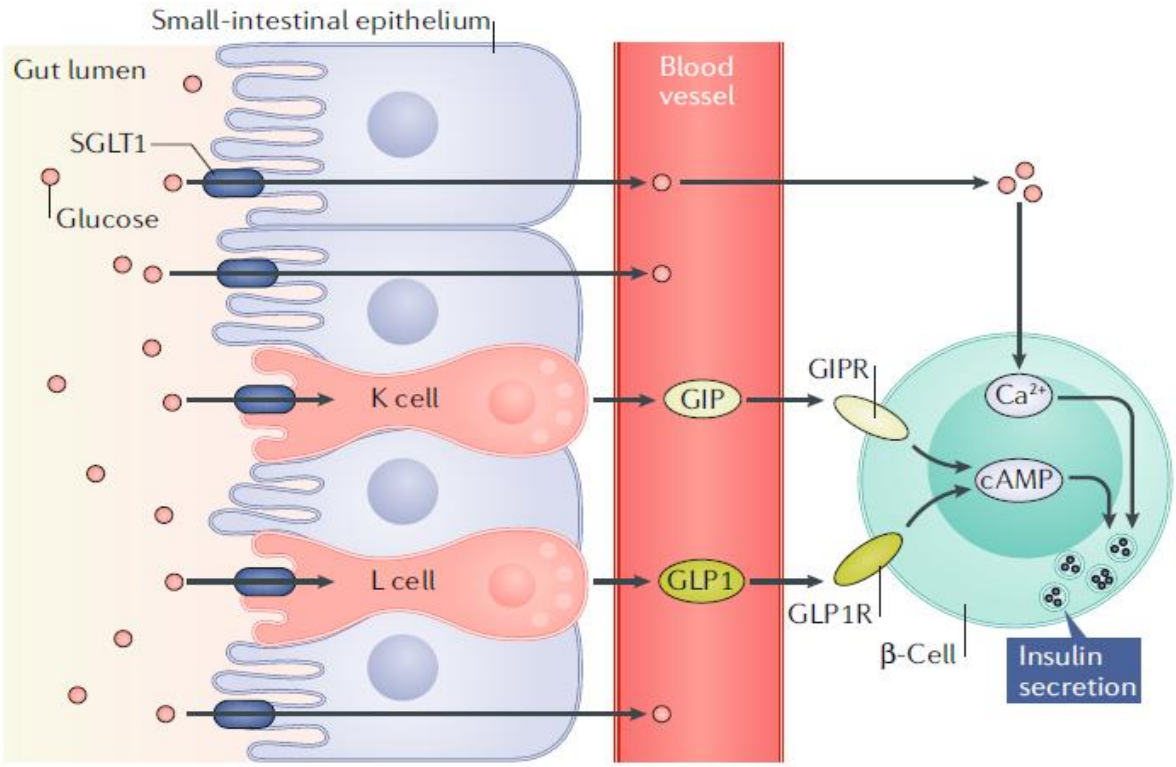


Figure 6.1: Overview of insulin potentiation by GIP and GLP-1. Glucose is absorbed by SGLT transporters on the surface of epithelial cells of the small intestine, resulting in elevated plasma glucose concentrations. This leads to higher Calcium^{2+} concentrations in pancreatic β-cells, triggering insulin secretion. Glucose detection and absorption by SGLT transporters on K cells or L cells triggers GIP and GLP-1 secretion respectively. Binding of GIP and GLP-1 to their respective receptors initiates a signalling cascade that leads to higher intracellular cAMP, triggering the amplification of the insulin secretion signal. Figure taken from[402].
Abbreviations: SGLT1, Sodium-dependent glucose transporter 1; GIP, Gastric inhibitory polypeptide; GLP-1, Glucagon-like peptide-1; GIPR, GIP receptor; GLP-1R, GLP-1 receptor; Ca^{2+}, Calcium^{2+}; cAMP, Cyclic adenosine monophosphate; β-cell, pancreatic β-cell

Incretin proteins are initially expressed as prohormones requiring enzymatic cleavage to liberate their active forms[414]. Proglucagon cleavage by prohormone convertase-1 liberates the more abundant GLP-$1_{7\text{-}36}$ amide isoform as well as the equipotent GLP-$1_{7\text{-}37}$ isoform[416]. Pro-GIP cleavage by the same enzyme liberates active $GIP_{1\text{-}42}$[417]. Both GLP-1 and GIP contain an alanine at position 2 in the peptide, making them ideal substrates for are rapid degradation by dipeptidyl peptidase IV (DPP4), resulting in a half-life of less than two minutes for GLP-1 and approximately seven minutes for GIP[414,418]. This rapid degradation facilitates fine-scale control over insulin-potentiating action and therefore glucose homeostasis. However, due to the physiological mechanism of incretins, their fasting levels are relatively low – making them challenging to measure, often requiring measurement during OGTT – this, coupled with their short half-lives, reduces the number of large-scale studies which have reliable incretin measures.

The incretin effect in health and T2D

In healthy individuals, the magnitude of the incretin effect is glucose-dependent[419–421] and estimated to be between 50-70% of postprandial insulin response[412,422]. GIP levels have been shown to mediate a substantially greater proportion of the incretin effect in individuals with normal glucose tolerance[423,424], whereas GLP-1 mediates approximately 32-55%[425–427]. However, the incretin effect has been shown to be diminished in patients with T2D, with around half the C-peptide production (150mM) per mM higher plasma glucose observed[419,420,428,429]. GLP-1 levels have been reported to be reduced in T2D[430–433], whereas postprandial GIP levels have been reported to be either normal or higher in T2D[433–437]. Later experiments showed that the insulinotropic effect of GIP was dampened in T2D, whereas the effects of GLP-1 were largely unaffected[433,438,439]. Contrastingly in obese participants without T2D, the insulinotropic effect of GIP was unaltered[440]. Later studies established that the diminished effect of GIP relative to GLP-1 was not exclusive to T2D but was common to monogenic diabetes, diabetic chronic pancreatitis and latent autoimmune diabetes patients[441,442]. On the basis of these findings, it was posited that impaired β-cell response to glucose, rather than defects in GIP signalling, was the likely cause of the diminished incretin effect in diabetes[441,442], suggesting that impairment of the incretin effect occurs in the early stages of diabetes pathogenesis[443].

Critical to the maintenance of glucose homeostasis, is the effect of incretins on glucagon secretion, as this may facilitate improved glycaemic control. In healthy individuals, studies have demonstrated that while GLP-1 stimulates insulin levels, it has a dose-dependent, suppressive effect on glucagon levels[444–447]. Conversely, GIP levels have been shown glucose-dependent glucagon secretion in healthy participants and T2D patients[421,429,448,449].

The role of GLP-1 in satiety, food consumption and weight loss

Aside from their effects on insulin secretion, both GLP-1 and GIP have diverse physiological actions on metabolically active tissues[422]. In clinical studies, higher GLP-1 levels have been shown to reduce appetite and as a result, food consumption[450]. The mechanism behind this is unclear, however, it is thought that the presence of the GLP-1R in the CNS[414,451] may be central to this effect[452]. Linked with this, GLP-1 levels have also been shown to decrease the rate of gastric emptying to the small intestine through the inhibition of gastrointestinal motility, an effect linked to satiety[451], in both healthy participants[425,453] and T2D patients[454]. Of specific relevance to cardiometabolic diseases, GLP-1 levels have been demonstrated to modulate body weight in both healthy participants and T2D patients[455–457].

GLP-1 mediated weight loss has been demonstrated in a RCT using liraglutide, a once-daily GLP-1 analogue in obese participants[458]. Participants who lost more than 5% of their baseline

weight during an initial low-calorie diet were randomised to receive either 3mg liraglutide or placebo for 56 weeks. The study aimed to assess whether GLP-1 maintained weight loss through caloric restriction and induced further weight loss[458]. Treatment with liraglutide led to further weight loss from the mean of 6.2% lost during caloric restriction, amounting to a total of 12.2% loss over the study duration. In addition, 81.4% of participants in the liraglutide group maintained the weight lost during caloric restriction, indicating that GLP-1 levels aided both weight maintenance and loss in obese participants[458]. The capacity to induce weight loss using GLP-1 agonists makes them attractive therapeutic targets for the treatment of obesity, coupled with their insulinotropic properties, GLP-1 agonists represent a viable therapeutic for T2D as well.

The role of GIP in obesity, dyslipidaemia and risk of T2D

Aside from the established role for GIP in modulating insulin and glucagon levels, studies conducted in human subcutaneous adipose tissue and *in vitro* studies of adipocytes have suggested a role in dietary fat metabolism and lipid storage in adipose tissue[459]. *In vitro* studies have suggested that GIP stimulates lipoprotein lipase (LPL) thereby promoting triacylglycerol (TAG) uptake into cells[459]. Findings from animal models have been varied[459–473] and highly dependent on experimental design[415]. Studies in obese humans have concluded that GIP does not influence food intake[474]. Findings from an observational cross-sectional study in 1,405 individuals across the spectrum of T2D risk, showed that doubled fasting GIP levels were associated with lower subcutaneous fat but higher visceral fat, waist-to-hip ratio, waist circumference and BMI in men[475]. Interestingly, in women the opposite was found, as high GIP levels were associated with 1.2% lower body fat percentage as well as lower BMI and waist circumferences[475]. In summary, interpretation of the available data from clinical studies is challenging due to the difficulty of estimating the effects of GIP independently of its insulinotropic effects. Taken together, the available evidence supports a role for GIP in rodent lipid metabolism, whereas its role in human lipid metabolism may be sex-specific and influenced by insulin levels, sensitivity, and BMI, thus making its role unclear[415,476].

As a result of the challenges surrounding incretin measurement in large epidemiological cohorts (outlined in the previous section), a considerable proportion of our knowledge of incretin physiology is based on experiments conducted in animal and *in vitro* models or from clinical studies in small cohorts. While there are many clinical studies investigating this, the findings are often in opposition or inconclusive. Despite this, the findings outlined in this section constitute the evidence base, upon which, both GIPR agonists and antagonists are being developed as putative treatment options against obesity and T2D[469,477,478].

6.2: The role of incretin levels in cardiovascular disease

Obesity and T2D, as outlined above, have been associated with raised CVD risk[15] as well as earlier onset[16] and greater severity[17] of CVD events. Recent evidence has suggested that the role of GIP in dietary lipid metabolism and adipose blood flow, coupled with GIPR expression in both adipocytes and cardiomyocytes[415], may implicate GIP in CVD pathogenesis[479,480]. Recent evidence from a meta-analysis of two large population-based cohort studies totalling 8,044 participants estimated the associations between fasting and post-challenge GIP levels with both total and cardiovascular mortality[481]. Higher fasting GIP levels were associated with higher hazard for both total and cardiovascular mortality (Hazard ratio [HR] per SD GIP, 1.22; 95% CI, 1.11, 1.35; P =4.5×10^{-5}) and (HR, 1.30; 95% CI, 1.11, 1.52; P = 0.001) in models extensively adjusted for age, sex and relevant cardiovascular risk factors[481]. Post-challenge GIP levels were not associated with either outcome[481]. Interestingly, neither fasting or post-challenge GLP-1 were associated with either outcome, suggesting that only GIP-mediated pathways may be implicated in raising cardiovascular risk[481].

These findings were somewhat reinforced by 2-sample MR estimates using a missense variant in *GIPR*, rs1800437 which encodes a substitution of glutamic acid for glutamine at position 354 of the GIPR peptide, as an IV for fasting GIP levels[481]. Estimates showed that genetically-predicted fasting GIP levels were associated with higher odds of CHD in both CARDIoGRAMplusC4D and UK Biobank (OR per copy of rs1800437, 1.67; 95% CI, 1.34, 1.99; P = 0.002) and (OR, 1.52; 95% CI, 1.27, 1.77; P = 0.001) respectively, as well as myocardial infarction (OR, 1.58; 95% CI, 1.22, 1.95; P = 0.01)[481]. In line with this, estimates in the reverse direction showed no significant effect of CHD on fasting GIP levels, suggesting that fasting GIP levels raise the risk of both CHD and myocardial infarction[481]. These estimates should be interpreted with caution, however, as (1) they represent the effect of a single variant on cardiovascular risk and do not model the effects of other variants in the region which may dampen or modulate this effect, (2) these estimates do not take into account that the association with CHD may be the result of LD between this variant and the true causative variant driving the association with CHD.

Similarly, genetic studies[482] using a low frequency missense variant in *GLP1R* (Ala316Thr; rs10305492) replicated the observational findings for GLP-1[481]. This variant was associated with lower fasting glucose (beta in SD per copy of rs10305492, -0.51; 95% CI, -0.20, -0.11; P=2.6×10^{-10}) T2D risk (OR, 0.83; 95% CI, 0.76, 0.91; P=9.4×10^{-5}) and CHD risk (OR, 0.93; 95% CI, 0.87, 0.98; P=9.2×10^{-3})[482].

Evidence from smaller clinical studies have demonstrated raised heart rate and differences in blood pressure during GIP infusions[448,483], however, while GIP's effects on heart rate appeared to be independent of its insulinotropic capacity, its effects on blood pressure were dependent on prevailing glycaemia. These effects should be interpreted with caution, however, as neither heart rate or blood pressure were predefined outcomes of the study, plus these effects are likely to be transient, as trends towards normalisation after 90 minutes of GIP infusion were noted[479]. This also does not rule out the possibility that postprandial GIP levels lead to spikes in heart rate that rapidly normalise following the degradation of active GIP by DPP-4. Studies using *in vitro* cellular models have suggested that these effects on blood pressure may be mediated by secretion of endothelin-1 (ET-1), a vasoconstrictor stimulated by signalling cascades upon binding of GIP to GIPR[484]. ET-1 has been shown to stimulate secretion of osteopontin (OPN), an inflammatory protein that facilitates leukocyte cell adhesion and migration that acts in atherogenic pathways, eventually leading to vascular remodelling and stiffening[485,486]. While these findings are in line with evidence suggesting that GIP raised cardiovascular risk, further evidence in larger epidemiological cohorts is required to confirm this[455,487].

6.3: Incretin-based therapies for cardiometabolic diseases

The discovery of GIP and the elucidation of its physiological role led to great enthusiasm about its potential as a possible therapeutic for T2D[488]. However, evidence of a reduced insulinotropic capacity in T2D[419,420,428,429], coupled with raised adipose deposition[459] and glucagon secretory responses[421,448] in non-diabetic participants dampened this interest. Added to this, a suitable treatment strategy for GIP was unclear due to evidence from animal models suggesting that both agonism and antagonism of GIPR could protect against dietary-induced obesity[460,466,489]. It has been suggested that chronically raised GIP levels, mimicking receptor agonism, may lead to reductions in GIPR expression, thereby leading to desensitisation to GIP's effects[490]. Other studies have suggested that naturally occurring end-products of endogenous GIP degradation may antagonise the GIPR[491]. Finally, human and murine GIPR share only 81% amino acid sequence homology[492]. These subtle differences coupled with differential responses to antagonists[493,494] have complicated the extrapolation of findings from animal studies into clinically meaningful evidence in humans.

Attention was soon diverted from GIP with the discovery of GLP-1 and evidence of its metabolically favourable effects on satiety[450] and glucagon suppression[444]. As postprandial GIP levels are consistent between healthy participants and T2D patients[433–436] and GLP-1 levels were lower in T2D[430–433], this presented an opportunity to raise GLP-1 levels in T2D

using GLP-1R agonism as a possible treatment. The effects of one such agonist, liraglutide, were assessed in a 20 week double-blind, placebo-controlled trial in 564 obese participants using orlistat, a lipase inhibitor, as a comparator[495]. Participants were randomised to one of four liraglutide doses, orlistat or placebo and were subjected to a 500 kcal per day restricted diet and physical activity regimen throughout the trial[495]. Participants in the liraglutide group lost significantly more weight than either comparator, mean weight loss for the four liraglutide doses estimated to be 4.8 kg, 5.5 kg, 6.3 kg, and 7.2 kg respectively, compared to 2.8 kg for placebo and 4.1 kg for orlistat. Of note, 76% of participants in the highest liraglutide group lost more than 5% of their baseline weight, compared to 30% and 44% on placebo and orlistat respectively[495]. Additionally, liraglutide was found to lower blood pressure and reduce the prevalence of pre-diabetes at most doses[495]. However, the most common adverse effects of liraglutide treatment were transient nausea and vomiting[495]. Overall, liraglutide treatment had a metabolically favourable, dose-dependent effect on weight as well as a protective effect against pre-diabetes, however, adverse effects such as nausea may impact upon its use as a chronic therapeutic[495].

In a 30 week trial of a second GLP-1R agonist, exenatide, in 295 T2D patients, once weekly administration of exenatide significantly lowered HbA1c and led to a greater proportion of patients achieving target HbA1c levels of 7% or less[496]. In addition, exenatide treatment was not shown to increase the risk of hypoglycaemia and replicated the beneficial effects on weight loss, however, nausea was again reported as a significant adverse event[496]. As mentioned previously, cardiovascular disease prevalence is higher in T2D patients[15]. Considering the metabolically favourable effects of GLP-1R agonists, investigations have assessed the cardioprotective efficacy of both lixisenatide[497] and exenatide[498]. Both studies included T2D patients with previous cardiovascular events, however, no significant differences were observed in the rate of cardiovascular events between GLP-1R agonist treatment and placebo[497,498].

Recently, evidence suggesting that the insulinotropic effect of GIP is raised in T2D patients receiving insulin treatment when near-normal glycaemic levels are achieved, has reignited interest in the therapeutic potential of GIPR agonism[421,428,499]. Studies in these patients have noted that GIP enhanced the glucose stimulated insulin response and that glucagon secretion was only stimulated when glucose levels were low[421,428,499]. This is beneficial, as glucagon secretion is regulated effectively, thus buffering against hypoglycaemia through hepatic glucose production[421]. However, the requirement of re-establishing near normo-glycaemia suggests that GIP needs to be administered in tandem with an additional glucose-lowering therapy[489,500]. To that end, recent pharmaceutical research has led to the development of GLP-

1R-GIPR co-agonists[428,501–503] which exert additive metabolically favourable effects as well as normalising adverse effects of single agonism such as nausea[489,500]. The first of these, NNC0090-2746, displays balanced agonistic effects at both receptors[504]. In a 12-week randomised, double-blind, placebo-controlled phase 2a trial, 108 T2D patients with poorly controlled glucose tolerance received daily 1.8mg doses of the dual agonist or placebo[504]. Escalating doses of liraglutide to a maximum of 1.8mg were used as an open-label comparator[504]. The dual agonist significantly lowered HbA1c levels by up to 1.04% and fasting plasma glucose compared to placebo over the 12 week period[504]. In line with GLP-1R agonism, significant weight loss up to 1.87% from baseline were noted, with continued reductions in body weight over the study duration[504]. Additionally, total cholesterol and leptin levels were significantly lowered compared to placebo and reduced the occurrence of adverse events such as nausea[504].

The efficacy of a second dual agonist, tirzepatide, was assessed in a double-blinded, randomised phase 2 trial where 316 T2D patients with poorly controlled glycaemia were randomised to one of four tirzepatide doses or placebo for 26 weeks[503]. Tirzepatide displayed unbalanced agonism at the two receptors, displaying greater affinity for the GIPR[503]. Treatment group assignment was stratified by baseline HbA1c, metformin use and BMI[503]. A dose-dependent effect of tirzepatide on HbA1c was observed, with mean reductions from baseline 1.06%, 1.73%, 1·89% and 1.94% respectively[503]. At the end of the trial, up to 90% of patients in the tirzepatide group achieved HbA1c levels of less than 7%[503]. In line with findings for other dual agonists, similar reductions in fasting plasma glucose, weight and total cholesterol were observed[503].

Taken together, these results suggest that while single GLP-1R or GIPR agonists may display metabolically beneficial effects, the use of dual agonists lead to dose-dependent additive effects on weight loss, glycaemic control and total cholesterol levels while lowering the incidence of adverse effects such as nausea. It should be noted that triple agonists targeting the GIPR, GLP-1R, and glucagon receptor are in development, however, their efficacy in large-scale human trials are yet to be assessed[489,500]. In addition, the cardiovascular risks of dual agonists have not yet been assessed. While pharmaceutical development has focussed on agonism as the favourable treatment strategy, further research into the clinical safety and efficacy of dual agonists is required to facilitate their establishment as a putative treatment for T2D.

6.4: Overview of the genetic architecture underpinning incretin levels

Introduction

As outlined above, incretins have become the focus of significant attention as putative targets for pharmacological modulation in T2D[415,489,505]. However, considering the often-contradictory clinical findings in the published literature it is unclear whether agonism or antagonism of their effects is the preferable treatment modality. As outlined previously, genetic analyses are robust to residual confounding and are therefore useful tools to predict the effect of pharmacological manipulation of this pathway in disease[56,57,506]. I therefore aimed to survey the current knowledge of genetic variants associated with incretin levels in healthy participants from large population-based epidemiological cohorts to identify variants of interest for further characterisation.

Methods

The published literature was systematically searched using the PubMed (MEDLINE) database to identify studies describing associations between genetic variants and circulating levels of incretins. The search strategy was focussed on human studies conducted in population-based cohorts investigating genetic associations with at least one incretin and was not limited to GWAS. My search included for all clinical trials, journal articles and meta-analyses published in English between June 2007 (the year the WTCCC GWAS[228] was published, considered as the inception of large-scale GWAS analyses) and 4 November 2019. The search strategy is outlined in **Table S6.1** and led to the inclusion of 2 full text articles in the review after using the exclusion criteria detailed in **Figure 6.2**.

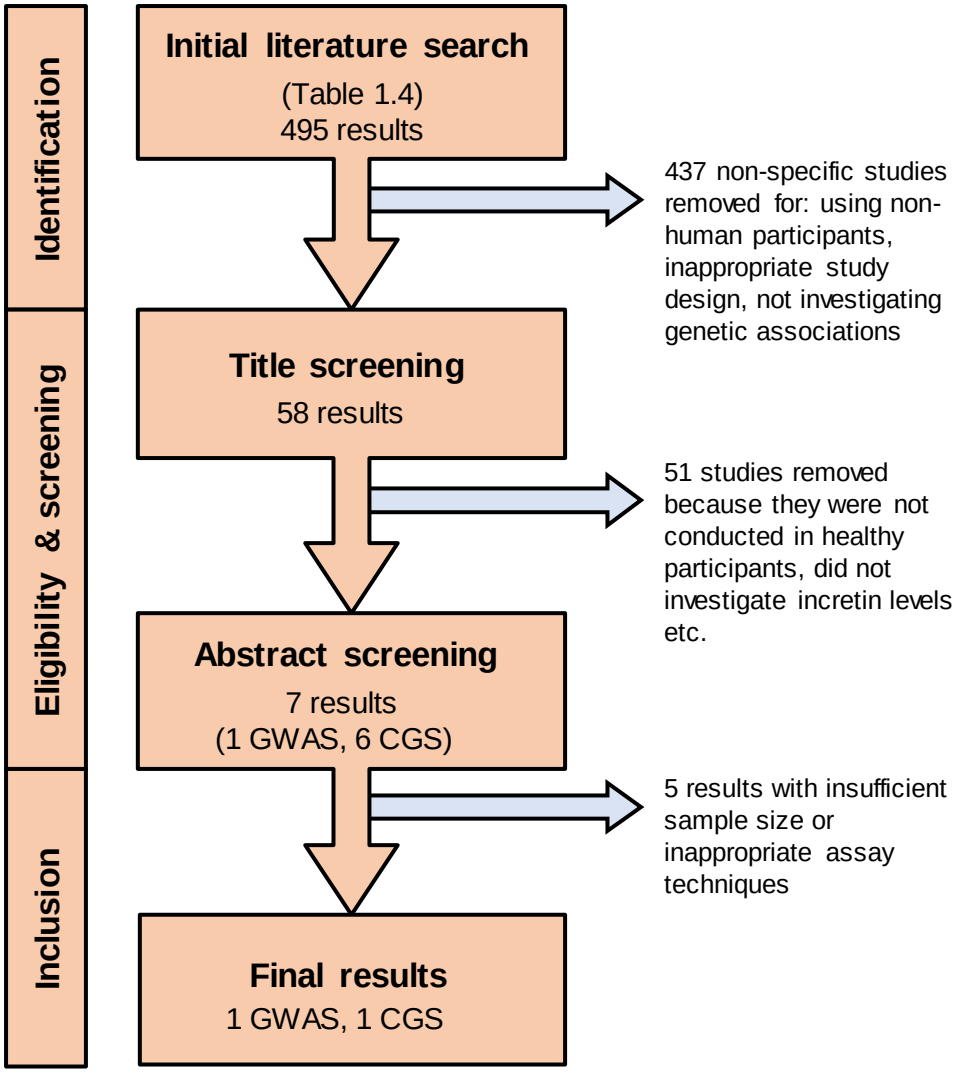


Figure 6.2: Schematic of the systematic literature search. The exclusion criteria and the number of articles excluded at each stage are detailed.
Abbreviations: GWAS, Genome-wide association study; CGS, Candidate-gene study

Results and discussion

The systematic search identified a total of 495 articles that were then prioritised to 2 full text articles[507,508] (**Table 6.1**) for review after applying the exclusion criteria summarised in **Figure 6.2**. The GWAS was conducted using participants derived from two prospective Scandinavian population-based cohorts – the Malmö Diet Cancer Cardiovascular Cohort (MDC) in Malmö, Sweden (*n*=3,344) and the Prevalence, Prediction and Prevention of Diabetes Botnia (PPP-Botnia) cohort, a population-based cohort in Western Finland (*n*=5,208)[508]. The CGS was conducted on 171 nondiabetic individuals[509]. All studies followed a standard 75g OGTT protocol in fasted participants and measured levels of GIP and GLP-1 at fasting and 2 hours post-OGTT. The CGS also measured incretin levels at 30 minutes post-OGTT.

Table 6.1: Details of the studies included in the systematic review					
Study	**Participating cohorts**	**Participant health and population**	**Number of participants**	**GLP-1 assay**	**GIP assay**
GWAS[508]	MDC	Healthy Europeans	3,344	In-house radioimmunoassay[510]	Human GIP Total ELISA kit EZHGIP-54K (EMD Millipore)
	PPP-Botnia		5,208	GLP1T-36HK radioimmunoassay (EMD Millipore)	
CGS[507]	Tübingen Family Study for type 2 diabetes	Healthy Europeans	171	In-house radioimmunoassay[511]	In-house radioimmunoassay[512]

Abbreviations: GWAS, Genome-wide association study; CGS, Candidate-gene study; MDC, Malmö diet and Cancer; PPP-Botnia, Prevalence prediction and prevention of diabetes in Botnia; GLP-1, Glucagon-like peptide-1; GIP, Gastric inhibitory polypeptide

Together, the two studies identified 12 variants associated with circulating levels of GLP-1 and GIP (**Table 6.2**). No associations with other incretins were reported in suitably powered studies. This not only highlights a deficit in our knowledge of the genetic underpinnings of other incretin levels but also illustrates how few studies have investigated the genetics of incretins. The variant identified by the CGS[507] was associated with 30-minute measures of both GLP-1 and GIP but was not associated with fasting or 2-hour measures of either incretin. Furthermore, the association of this variant with levels of GLP-1 and GIP was not replicated by the GWAS, thus its relevance is currently unknown.

Of the variants identified by the GWAS[508], seven were missense variants and some, such as rs1800437 in the GIP receptor gene (*GIPR*), were associated with incretin measures at multiple timepoints. Several of the reported variants were in high LD ($R^2 > 0.8$) with one another, such as those at the *GIPR, ABO* and *SLC5A1* loci, effectively representing the same association signal. A total of four variants at the pleiotropic *ABO* locus, all in high LD ($R^2 = 0.83$) with one another, were associated with levels of GIP and GLP-1. This is unsurprising as this locus has previously been shown to be associated with levels of cytokines[232,234,235] (as described in **Chapter 1**), blood proteins[513], cholesterol[514], HbA1c[347], T2D[279], and risk of CHD and large artery stroke[515]. Given that the *ABO* locus was associated with lower 2-hour and not fasting GIP levels[508], it may be speculated that the *ABO* locus may impact upon levels of DPP-4 which in-turn induce GIP degradation[418], leading to lower GIP levels.

Table 6.2: Genetic variants associated with levels of GIP and GLP-1 in published studies									
Outcomes	**variant**	**Gene**	**Position**	**LD (r^2) [a]**	**EA/OA**	**Variant type**	**N**	**Beta (SE) in log(pmol/l)**	**P-value**
Fasting GIP	rs17681684	*GLP2R*	chr17:9792768	-	A/G	Missense	7,826	0.055 (0.009)	1.4×10^{-9}
	rs1800437	*GIPR*	chr19:46181392	0.89	G/C	Missense	7,825	0.076 (0.010)	4.1×10^{-15}
	rs2287019	*GIPR*	chr19:46202172		T/C	Intron	7,828	-0.066 (0.010)	4.0×10^{-11}
30 min GIP	rs151290[b]	*KCNQ1*	chr11:2821615	-	A/C	Intron	1,578	3[c]	0.004
2hr GIP	rs1800437	*GIPR*	chr19:46181392	0.89	G/C	Missense	7,448	0.072 (0.009)	1.6×10^{-17}
	rs2287019	*GIPR*	chr19:46202172		T/C	Intron	7,450	-0.066 (0.009)	4.0×10^{-14}
	rs150112597	*HOXD1*	chr2:177054181	-	C/G	Missense	7,443	-0.280 (0.049)	1.6×10^{-8}
	rs927332	*F13A1*	chr6:6331873	-	T/C	Intergenic	7,442	0.040 (0.007)	4.1×10^{-8}
	rs635634	*ABO*	chr9:136155000	0.83	T/C	Upstream	7,446	-0.070 (0.009)	7.4×10^{-14}
	rs507666	*ABO*	chr9:136149399		A/G	Intron	7,450	-0.069 (0.009)	8.9×10^{-14}
	rs651007	*ABO*	chr9:136153875		T/C	Upstream	7,449	-0.063 (0.009)	2.9×10^{-12}
	rs579459	*ABO*	chr9:136154168		C/T	Upstream	7,450	-0.062 (0.009)	3.4×10^{-12}
	rs17683430	*SLC5A1*	chr22:32487700	1	A/G	Missense	7,450	0.120 (0.014)	3.0×10^{-18}
	rs17683011	*SLC5A1*	chr22:32445946		G/A	Missense	7,450	0.120 (0.014)	3.2×10^{-18}
30 min GLP-1	rs151290[b]	*KCNQ1*	chr11:2821615	-	A/C	Intron	1,578	5[c]	0.02
2hr GLP-1	rs17683011	*SLC5A1*	chr22:32445946	-	G/A	Missense	3,849	0.091 (0.017)	4.2×10^{-8}

Abbreviations: variant, Single nucleotide polymorphism; LD, Linkage disequilibrium; EA, Effect allele; OA, Other allele; N, Number of participants; SE, Standard error; GWAS, Genome-wide association study; GIP, Gastric inhibitory polypeptide; GLP-1, Glucagon-like polypeptide-1; Min, Minutes; Chr, Chromosome

a. LD is only reported for variants on the same chromosome
b. Variant was reported in the candidate gene study[507]
c. Variant effects were reported as fold changes relative to wild-type alleles

The results of the GWAS[508] also showed an association of *GIPR* variants with both fasting and two-hour GIP levels. The LD between these two variants, rs1800437 and rs2287019, is high (R^2 = 0.89; **Table 6.2**). Interestingly, an intronic variant, rs10423928, in perfect LD with the *GIPR* missense variant rs1800437 was initially discovered in a GWAS[344] of 2-hour glucose levels in up to 15,234 non-diabetic participants. The association of this variant with both 2-hour glucose levels[344] as well as fasting and 2-hour GIP levels[508] fits well with the role of GIP in insulin potentiation and glucose homeostasis. In line with results from observational studies[475], previous studies have shown that the GIP decreasing alleles of the two variants at the *GIPR* locus, rs1800437 and rs2287019, are associated with lower BMI[516] and waist-to-hip ratio[517]. Interestingly, lower rs1800437-mediated GIP levels have also been shown to be associated with higher T2D risk (OR, 1.06; 95% CI, 1.04, 1.08; $P = 3\times10^{-14}$) when adjusted for BMI in 74,124 T2D European cases and 824,006 controls[22]. This suggests that this raised T2D risk is not mediated by obesity-related pathways but is instead resultant of the diminished insulinotropic and glucose-lowering actions of GIP in T2D[419,420,428,429].

Of interest to this thesis was rs1800437, a missense variant in *GIPR*. It was chosen for further characterisation because (i) rs1800437 has translational consequences inducing the substitution of glutamic acid for glutamine at position 354 of the GIPR peptide (E354Q)[518]. (ii) the G allele (E354; frequency ~0.74) is associated with both higher fasting and 2-hour GIP levels of 0.076 and 0.072 log (pmol/L) higher fasting GIP levels per copy of E354 respectively (**Table 6.2**). (iii) this general increase in GIP signalling[508] mimics the effect of GIPR agonism therefore the functional consequences of E354 are of particular interest for the development of GIPR-targeting therapies for diabetes, as previously outlined.

Chapter 7: Genetically predicted glucose-dependent insulinotropic polypeptide (GIP) levels and cardiovascular disease risk are driven by independent mechanisms at the GIPR

Contributions and collaborations

I designed the study, analysed the data and wrote the first draft of the manuscript. Claudia Langenberg, Nicholas J. Wareham, Pallav Bhatnagar, Matthew Coghlan, Fiona Gribble and Frank Reimann supervised the study. I performed the 2SMR analyses and analysed the data with the help of Robert Hansford. Christopher N. Foley provided statistical support. Victoria Au Yeung conducted the partial correlation analysis and created the Gaussian Graphical Model. Stephen Burgess, Aslan M. Erzurumluoglu, Isobel D. Stewart, Eleanor Wheeler and Maik Pietzner assisted me with data collation and contributed to the data analysis. All authors contributed to the interpretation of results and writing or revision of the manuscript.

Abstract

Background: There is significant interest in pharmacological agonism of the glucose-dependent insulinotropic polypeptide (GIP) receptor as a therapeutic for type 2 diabetes (T2D) and obesity. Evidence from *in vitro* and animal models as well as acute clinical studies has implicated higher fasting GIP levels in raising cardiovascular disease risk, a safety concern threatening the use of these therapies. Recent genetic epidemiological evidence also suggests that higher GIPR-mediated GIP levels are associated with elevated risk for cardiovascular disease.

Methods: I integrated large-scale genomic data for 24 different cardiometabolic diseases, as well as anthropometric, glycaemic, lipid and ~6,000 'omics biomarkers (metabolomic and aptamer-based plasma proteomic measures) to systematically characterise the clinical consequences of the *GIPR* missense variant rs1800437 (G allele; E354 henceforth). Bayesian multi-trait colocalisation was applied to define distinct clusters of traits driven by a shared causal variant. Conditional analyses were used to identify variants contributing to associations between the *GIPR* locus and traits of interest.

Findings: I identified E354 as the putatively causal variant shared between fasting GIP levels (higher), diabetes-related (protective) and adiposity-related (adverse) traits (posterior probability for colocalisation, PP_{Coloc}>0.97; $PP_{explained}$=0.99; PP_{causal}=0.96). In unconditional analyses, each copy of E354 was significantly associated with an increased risk of CHD, albeit with an odds ratio (OR) of 1.03 (95% CI, 1.02,1.05; $P=2x10^{-6}$), in line with associations observed for lipids, anthropometric traits and other known cardiovascular risk factors. I identified robust evidence of colocalisation at a distinct signal for associations with CHD and lipids (LDL, HDL, total cholesterol, lipoprotein A and ApoB) driven by a known independent missense variant (rs7412; distance to E354 ~770Kb) in *APOE* (PP_{coloc}>0.99; $PP_{explained}$=1; PP_{causal}=0.99). However, the low LD between rs7412 and E354 (R^2=0.004) and results from conditional analyses showed that this signal was not responsible for the association between E354 and CHD (OR for E354 after adjustment for rs7412: 1.03 (95% CI, 1.02,1.04; P=0.003). However, E354's effect on CHD was completely attenuated when conditioned on an additional established CHD signal, rs1964272, (R^2 with E354=0.27), an intronic variant in *SNRPD2* (OR for E354 after adjustment for rs1964272: 1.01; 95% CI, 0.99, 1.03; P=0.06).

Interpretation: I demonstrate that associations of glycaemic and adiposity traits versus CHD and lipid traits in the *GIPR* region are driven by distinct genetic signals. This study provides evidence that inclusion of GIPR agonists in dual agonists could potentiate the protective effect of GLP-1 agonists on diabetes without undue cardiovascular risk.

7.1: Introduction

As outlined in **Chapter 6**, the incretin hormones glucose-dependent insulinotropic polypeptide (GIP) and glucagon-like peptide-1 (GLP-1) are well known for their insulinotropic activity[412,422], which is diminished in type 2 diabetes (T2D)[419,420,428,429]. This has prompted significant therapeutic interest in the agonism of their respective receptors, GIPR and GLP-1R, to rescue their insulinotropic effects[489,500]. Moreover, preclinical and clinical data demonstrate that dual agonism of the GIPR and GLP-1R deliver superior glycaemic and weight control efficacy compared to selective GLP-1R agonism[503,519,520]. Clinical proof-of-concept for the superiority of tirzepatide, a dual GIP/GLP-1R agonist, versus GLP-1R agonism alone was established in a 6-month dose range finding Phase 2b trial in subjects with type 2 diabetes[503]. Post hoc analysis of the tirzepatide phase 2b trial reported a beneficial impact on cardiovascular risk biomarkers compared to the blinded GLP-1R agonist included in the trial[521].

There exists little direct preclinical experimental physiologic or pharmacologic evidence for GIPR agonism contributing to cardiovascular risk[480,522]. GIP exhibits largely anti-atherogenic effects on vascular endothelial cells[523] with the exception that it has been reported to stimulate expression of osteopontin in the vasculature in an endothelin-1 dependent manner[484]. However, recent evidence from a meta-analysis[481] of two large population-based cohort studies suggests that higher fasting but not post-challenge GIP levels were associated with higher hazard for cardiovascular mortality in models adjusted for age, sex and relevant cardiovascular risk factors (HR, 1.30; 95% CI, 1.11, 1.52; P=0.001). Neither fasting nor post-challenge GLP-1 were associated with cardiovascular mortality, suggesting that only GIP-mediated pathways are implicated in raising cardiovascular risk[481], a finding corroborated by both clinical trial data[524–527] and genetic evidence[482].

Genetic studies using two-sample Mendelian randomisation (2SMR) have reinforced suggestions that higher GIP levels raise cardiovascular risk[481]. A missense variant in *GIPR*, rs1800437 (E354Q), encoding a substitution of glutamic acid for glutamine at position 354 of the GIPR peptide, was used as an instrumental variable for fasting GIP levels[481]. This variant has previously been associated with higher 2-hour glucose[528], BMI[529] and fasting and 2-hour GIP levels[508]. Estimates showed that genetically-predicted higher fasting GIP levels were associated with higher odds of both coronary heart disease (CHD) and myocardial infarction (MI)[481]. In line with a predicted causal direction from fasting GIP levels to CHD risk, estimates in the reverse direction showed no significant effect of CHD on fasting GIP levels[481]. These

genetic estimates should be interpreted with caution, however, as (1) they represent the effect of a single variant on cardiovascular risk and do not model the effects of other variants in the region which may dampen or modulate this effect, and (2) they do not take into account that the association between E354 and CHD may be entirely synthetic due to LD between this variant and the true CHD causal variant.

Considering the pharmacological interest in modulating this pathway as a potential T2D therapeutic, increases in cardiovascular risk would represent a major concern with regard to the safety and continued development of these therapies. To quantitatively assess the GIPR-mediated cardiovascular risk, I used a three-stage study design. Firstly, I integrated large-scale genomic data for cardiometabolic diseases as well as anthropometric, glycaemic, lipid, metabolomic and proteomic traits to characterise the association of E354 with clinical biomarkers. Secondly, I used a Bayesian multi-trait colocalisation method to formally test whether the genetic signal at the *GIPR* locus mediating a protective effect on T2D is the same signal driving cardiovascular disease risk. Lastly, results from statistical colocalisation were confirmed using conditional analysis to identify independent variants for each trait and account for the LD between these variants and E354.

7.2: Methods

Study design

Three sets of genetic analyses were used to investigate the relationships between the GIPR and cardiovascular risk. Firstly, the association of a missense variant (E354Q) in the GIP receptor gene (*GIPR*) with CHD and 23 different cardiometabolic diseases, as well as anthropometric, glycaemic, lipid and ~6,000 'omics biomarkers was estimated using univariate 2SMR (**Table S7.1**). Secondly, using the traits significantly associated with E354, Bayesian multi-trait colocalisation was applied in the *GIPR* region to define distinct clusters of traits driven by a shared causal variant. Finally, conditional analyses were used to identify independent variants contributing to associations between the *GIPR* locus and traits of interest. All studies were approved by local institutional review boards and ethics committees and participants gave written informed consent.

Study participants

EPIC-Norfolk[241] (**Table 7.1**) is a population-based prospective cohort of individuals aged between 40-79 years and living in Norfolk (a county of the United Kingdom) at the time of recruitment from primary-care outpatient clinics in the city of Norwich and surrounding areas. The study was approved by the Norfolk Research Ethics Committee (ref. 05/Q0101/191) and all participants gave their written consent before entering the study.

Fenland[244] (**Table 7.1**) is a population-based cohort study of individuals without diabetes who were born between the years of 1950 and 1975 and recruited through population-based general practice registers in Cambridge, Ely and Wisbech (Cambridgeshire county, United Kingdom). Ethical approval for the study was given by the Cambridge Local Ethics committee (ref. 04/Q0108/19) and all participants gave their written consent prior to entering the study.

UK Biobank[66] (**Table 7.1**) is a population-based cohort study of individuals recruited from 22 rural and urban recruitment centres in the United Kingdom. European ancestry participants with available genome-wide genotyping and phenotypic data were included in this study. Ethical approval for the UK Biobank study was given by the North West - Haydock Research Ethics Committee (16/NW/0274). This research was conducted using application 44448. Participants gave their electronic consent to use their anonymised data and samples for health-related research, to be re-contacted for further sub-studies, and for access to their health-related records.

Table 7.1: Study participants			
Study	**Fenland**	**EPIC-Norfolk**	**UK Biobank**
Participants, N	10,708	11,539	452,197
Age at baseline, mean years (SD)	49 (7)	60 (9)	57 (8)
Women, N (%)	5,714 (53)	6,198 (54)	245,277 (54)
Men, N (%)	4,994 (47)	5,341 (46)	206,883 (46)
BMI in kg/m2, mean (SD)	26.9 (4.9)	26.2 (3.7)	27.4 (4.8)
Waist-to-hip ratio, mean (SD)	0.74 (0.08)	0.86 (0.09)	0.87 (0.09)
Systolic blood pressure in mmHg, mean (SD)	123 (15)	136 (18)	138 (19)
Diastolic blood pressure in mmHg, mean (SD)	74 (10)	82 (11)	82 (10)
Fasting glucose in log-pg/mL, median (IQR) [a]	1.57 (1.50, 1.63)	N/A	N/A
2-hr glucose in log-pg/mL, median (IQR) [a]	1.63 (1.44, 1.79)	N/A	N/A
Fasting insulin in log-pg/mL, median (IQR) [a]	3.66 (3.29, 4.06)	N/A	N/A

Abbreviations: N/A, not available; N, number of participants; SD, standard deviation; BMI, body mass index; mmHg; Millimetres Mercury; pg; Picograms; mL, Millilitres; IQR, Interquartile range

a. Glycaemic measures from Epic-Norfolk and UK Biobank were not used in this study

Genotyping and imputation

Genome-wide genotyping in the Fenland cohort was performed in 3 sub-cohorts using either the Affymetrix genome-wide Human variant Array 5.0, the Affymetrix UK Biobank Axiom Array or the Illumina CoreExome-24 v1 chip, with imputation to the Haplotype reference consortium v1.1[271], the 1000 genomes project[272] and the UK10K[273] reference panels. Samples from EPIC-Norfolk and UK Biobank were genotyped using the Affymetrix UK Biobank Axiom Array and imputed to the same reference panels.

Profiling of the plasma proteome

Fasted EDTA plasma samples from 12,084 participants from the Fenland[244] study were subjected to proteomic profiling by SomaLogic Inc. (Boulder, US) using an aptamer-based technology (SOMAscan v4). The relative abundances of 4,775 human proteins were measured using 4,979 SOMAmers, as previously described[364]. To account for within run hybridisation variability, control probes were used to generate a scaling factor for each sample. Differences in total signal between samples as a result of variation in overall protein concentration or technical variability such as reagent concentration, pipetting or assay timing, were accounted for using the ratio between each SOMAmer's measured value and a reference value. The median of these ratios was computed for each dilution set (40%, 1% and 0.005%) and applied to each dilution set. Samples were removed if they failed SomaLogic QC measures or did not meet the acceptance criteria of between 0.25-4 for all scaling factors. A total of 10,078 samples had available genotype data and were used in this study. Aptamer target annotations and mapping to UniProt accession numbers as well as gene identifiers were provided by SomaLogic.

Plasma metabolomic profiling

The EPIC-Norfolk population-based cohort[241] (described previously) consists of two sub-cohorts, a T2D case-cohort and a quasi-random selection of participants from the larger EPIC[242,530] study. The levels of up to 1,504 metabolites were measured in three batches using the Metabolon DiscoveryHD4 platform[531] (Metabolon, Inc., Durham, USA), in citrate plasma samples collected at baseline. Measurements were made in approximately 12,000 samples, in two sets of approximately 6000 quasi-randomly selected samples, which were preceded by measurements in an incident T2D case-cohort (N= 1503; 857 in the sub-cohort).

Briefly, raw data were extracted, peaks were identified and assessed for quality by Metabolon. Metabolite identification was done by comparing measures to a curated library containing the retention time, mass to charge ratio and chromatographic data of known metabolites. Each metabolite was then quantified using an area-under-the-curve method and the data were normalised to correct for instrument tuning variations across run-days. Data normalisation for each run-day set the median value for each metabolite to 1, normalising each measurement proportionately. Metabolite annotations and pathway classifications are as reported by Metabolon, Inc.

Statistical analysis

GWAS of plasma proteins

Prior to GWAS, relative abundances of each SOMAmer were transformed within each genotyping subset using the rank-based inverse normal transformation. These were then regressed on age, sex, sample collection site and 10 principal components to generate residuals. GWAS was then performed using BGENIE (v1.3), assuming an additive model. Variants with minor allele frequency < 0.001, imputation quality < 0.4 or Hardy Weinberg Equilibrium $p < 1x10^{-7}$ in any of the genotyping subsets were excluded from further analyses. Results from the three genotyping subsets were later combined in a fixed-effects meta-analysis using METAL. Only variants present in the largest genotyping subset were taken forward for further analysis, variants specific to either of the smaller subsets were not considered.

Two SOMAmers were used to target circulating GIP, namely 16292-288 and 5755-29. SOMAmer 16292-288 was selected against amino acids 1-93 of the precursor protein, corresponding to the signal peptide, first propeptide and GIP peptide in Uniprot[532]. Overlapping this region, 5755-29 targeted amino acids 22-153 which corresponded to the first propeptide, GIP peptide and the second propeptide in UniProt[532]. In order to ascertain whether the

underlying genetics at *GIPR* were comparable between the relative abundance SOMAmers and absolute quantitation techniques such as enzyme-linked immunosorbent assays (ELISA), I performed pairwise genetic colocalisation analyses between GIP measures and cardiometabolic traits (**Methods S7.1**).

GWAS of plasma metabolites

Metabolite quantitation was performed for a total of 1,008 metabolites across 8 metabolite classes and as yet unidentified metabolites, measured in up to 11,539 individuals. GWAS was performed in 2 sets, for all metabolites present in at least 100 individuals in both sets. The first set consisted of up to 5,841 individuals from both the T2D sub-cohort and the first batch of quasi-randomly selected samples. The second set consisted of up to 5,698 individuals from the second batch of quasi-randomly selected samples. Metabolite levels were then natural log-transformed and winsorised to five SDs. Residuals were calculated, adjusting for age, sex (and measurement batch where appropriate) and standardised (mean =0, SD =1). Genotyping and imputation were performed as described above. Following imputation, variants with imputation quality INFO < 0.4 or minor allele account (MAC) of <= 2 within EPIC-Norfolk were excluded.

GWAS were performed using mixed linear models in BOLT-LMM[533] v2.2. In instances where BOLT-LMM[533] failed, related individuals were excluded (identity by descent > 0.185) and linear regression models were run using variant SNPTEST[274] v2.4.1, while also adjusting for the top 4 principal components. Variants with evidence of deviation from Hardy Weinberg equilibrium ($p<10^{-6}$), as well as associations with effect estimates with absolute values above 10 or standard errors less than 0 or above 10 were removed. Variants with MAF <0.0001 or imputation quality scores lower than 0.3 were excluded. An inverse variance weighted meta-analysis was performed to pool associations from the 2 GWAS sets, using METAL[276]. This included a minor allele count filter of 10 for each individual contributing GWAS, so that these associations did not contribute to the meta-analysis.

Association between E354, cardiometabolic and molecular traits

This work leveraged GWAS summary statistics from in-house studies and publicly available data. All phenotypes included in this study, along with details of their respective studies can be found in **Table S7.1**. Only self-reported, white European participants were included.

The exposure variable for this study was the genotype of rs1800437 (HUGO Gene Nomenclature Committee gene name, *GIPR*; transcript change, NCBI transcript identifier NM_000164.4 c.1060G>C; protein change, E354). The E345 variant is encoded by the G

allele and has an allele frequency in European ancestry participants in the 1000 Genomes Project of 79%. Fasting and 2-hour GIP levels used as outcomes in this study were taken from a recent GWAS meta-analysis including up to 7,828 Scandinavian participants[508].

The disease outcomes investigated in this study included GWAS summary statistics of T2D[22], T2D adjusted for BMI[22] and CHD[289] (a meta-analysis of UK Biobank[66] and CARDIoGRAMplusC4D[534]). In addition to these broader cardiometabolic disease outcomes, summary statistics from cardiovascular disease and stroke sub-types were used. I included summary statistics from 16 cardiovascular disease subtypes defined in UK Biobank[535] and a total of five stroke subtypes from the MEGASTROKE[536] consortium.

I gathered genome-wide summary statistics for the following glycaemic traits from different resources. Firstly, published measures of fasting glucose, 2-hour glucose, fasting insulin and corrected insulin response from the MAGIC[344,346,537] consortium were used. GWAS summary statistics for fasting glucose, 2-hour glucose and fasting insulin were all adjusted for BMI[344,346]. Measures of HbA1c and non-fasted plasma glucose levels will be described along with the other biomarkers measured in UK Biobank[66] and InterAct[538].

Summary statistics for each of the 19 biomarkers included in this study (**Table S7.1**) were taken from a meta-analysis of UK Biobank[66] and InterAct[538] data. All biomarkers in InterAct, except HbA1c, were measured using a Cobas® (Roche Diagnostics, Mannheim, Germany) assay on a Roche Hitachi Modular P analyser. HbA1c was measured on erythrocyte samples using a Tosoh (HLC-723G8) assay on a Tosoh G8 analyser. All traits were then regressed on age, age^2, sex, and centre using multivariable linear regression to generate residuals. These residuals were then rank-based inverse normal transformed. The top 10 genetic principal components were included in the final linear regression model. Traits measured in UK Biobank were also rank-based inverse normal transformed within each respective aliquot. In the final model, all traits were adjusted for age, age^2, sex, aliquot, genotyping chip, lipid lowering medication and the top 40 principal components using a multivariable linear mixed model. GWAS were performed using a linear mixed model implemented in BOLT-LMM[533] to account for cryptic population structure and relatedness. Results from both GWAS analyses were then meta-analysed using METAL[276].

The anthropometric outcomes included in this study were estimated in a meta-analysis of Genetic Investigation of Anthropometric Traits (GIANT)[343,539] and UK Biobank data[66]. In UK Biobank[66], weight was measured using a calibrated electronic scale (TANITA model BC-418 MA; Tanita, Tokyo, Japan). Height was measured with a wall-mounted stadiometer (SECA

202; Seca, Birmingham, United Kingdom). BMI (in kg/m2) was calculated as weight divided by height squared. Waist and hip circumferences were measured with a non-stretchable sprung tape measure (Wessex tape, London, United Kingdom). WHR was the ratio between the waist and hip circumferences. Residuals were then generated by regressing each outcome against age, age^2, study-specific covariates and BMI (if applicable). Residuals were estimated for each sex independently and transformed using a rank-based inverse normal transformation prior to association testing using multivariable linear regression. Results from UK Biobank GWAS analyses were then meta-analysed with results from GIANT[343,539] using METAL.

Bio-impedance measurements were performed in UK Biobank[66] using the Tanita BC418MA body composition analyser (Amsterdam, The Netherlands). All bio-impedance traits were natural log transformed and adjusted for age (and total fat mass or height^2 – if adjusted) in each sex separately to generate residuals. As before, the residuals were rank-based inverse normal transformed. Each trait was then tested for association with genetic variants using a linear mixed model in BOLT-LMM[533], adjusting for age, sex, genotyping chip, and the top 40 principal components.

I performed two-sample univariate MR analyses using the Wald ratio method[540] to estimate the potential causal effect of fasting GIP levels on various cardiometabolic and molecular traits (**Table S7.1**). Genetically predicted fasting GIP levels were used as the exposure and E354 was used as the instrumental variable. Associations of E354 with each of the cardiometabolic and molecular traits listed above were extracted from GWAS summary statistics and used as outcomes. All summary statistics were aligned to the fasting GIP raising allele (G) of E354. Estimates of the association of E354 with disease outcomes were expressed as odds ratios per copy of E354. Associations with all other traits were expressed as SD unit increase in trait (beta) per copy of E354. A nominal significance threshold of $P<0.05$ was used to ascertain statistical significance for all outcomes aside from cardiovascular disease subtypes, protein and metabolite levels, where Bonferroni significance thresholds accounting for the number of disease subtypes, proteins or metabolites were used. All data analysis was performed using R version 3.6.3.

Multi-trait colocalisation across cardiometabolic traits

I used multi-trait colocalisation (HyPrColoc)[362] at the *GIPR* locus to 1) identify cardiometabolic traits that share a common causal variant, and 2) identify distinct clusters of cardiometabolic traits driven by distinct causal variants. The same prior configurations and thresholds as described in **Chapter 5** were used.

All variants within 1Mb either side of E354 were extracted from GWAS summary statistics for 26 cardiometabolic traits of interest. The GIP measures considered were fasting GIP as measured by SOMAmers X16292_288 and 5755-29, as well as fasting and 2-hr GIP measures from the Malmö Diet and Cancer (MDC) sub-cohort of Almgren *et al.* 2017[508]. MDC measures were preferred to those from either PPP-Botnia or the meta-analysis of the two cohorts as the MDC genotyping platform had better variant coverage than PPP-Botnia, despite PPP-Botnia having a larger sample size. In addition, the variant coverage of the meta-analysis was limited to variants present in both genotyping subsets, effectively limiting the variant coverage to that of PPP-Botnia. The anthropometric traits included were BMI, WHR, and hip and waist circumferences. Anthropometric traits adjusted and unadjusted for BMI were included, where applicable. T2D and CHD were included as disease outcomes of interest. Glycaemic measures included non-fasted glucose, HbA1c, 2-hr glucose adjusted for BMI, fasting glucose adjusted for BMI and fasting insulin adjusted for BMI. Finally, lipid traits of interest were LDL, HDL, total cholesterol, triglycerides, lipoprotein A, apolipoprotein A1 and apolipoprotein B. Summary statistics used for fasting and 2-hour glucose as well as fasting insulin were not the same as those used to characterise the association of E354 with cardiometabolic traits. Despite these summary statistics being well-powered to detect associations at E354, their imputation to older reference panels led to poor variant coverage, which made them unsuitable for use in a colocalisation setting. To address this, I carried out separate GWASs for these three traits in the densely imputed Fenland cohort (**Methods S7.2**) and used these summary statistics for HyPrColoc.

Sensitivity analyses were run as described in **Chapter 5.** The MDC cohort was genotyped using an exome-wide platform (Illumina Infinium OmniExpressExome v1.0), thus limiting the variant coverage especially in intergenic regions. To address this and provide greater genomic context, I ran a sensitivity analysis using the same configuration and sensitivity assessments as above, while excluding the GIP traits measured in MDC.

Finally, heatmaps based on similarity matrices estimating how often trait pairs were clustered together across all algorithm parameter choices were drawn. In addition, stacked regional association plots were drawn for each cluster of traits using the gassocplot R package. LD data used to plot the correlation between the candidate variant and other variants in the locus were estimated using Plink v1.9[541] in EPIC-Norfolk. All data analysis was performed using R version 3.6.3.

Conditional analysis at the *GIPR* locus

To determine whether the association between E354 and CHD was due to LD between E354 and other CHD lead variants in the *GIPR* region, I performed conditional analysis using GCTA[542] v1.93.1. Using GWAS summary statistics for CHD[289], I implemented a step-wise selection based on regional LD to identify independent variants associated with CHD on chromosome 19. Selection was performed using a selection threshold of $P< 1x10^{-5}$, a threshold for collinearity between variants of 0.05 and a minor allele frequency threshold of 1%. An LD reference panel from the EPIC-Norfolk cohort imputed as described above was used for selection. Using the identified independent variants, the association between E354 and CHD was conditioned on each respective variant to estimate whether the association was attenuated. Attenuation of the association would imply that the association between E354 and CHD was due to the residual LD between E354 and one of the independent variants, rather than being driven by a separate mechanism. This procedure was repeated for all traits associated with E354. If E354 (or a proxy variant in complete LD with E354) was identified as one of the independent variants, conditional analysis was not performed. Following this, regional association plots were generated to visualise the association of E354 with CHD, before and after conditioning using LocusZoom v1.2. I also extracted the estimates for other variants previously found to be associated with fasting GIP levels[508] from the CHD summary statistics[289]. This was done to ascertain whether any other loci showed overlapping associations with both fasting GIP levels and CHD.

7.3: Results

Characterisation of a missense variant E354Q (rs1800437) in the *GIPR*

T2D and related glycaemic outcomes

Among the cardiometabolic disease outcomes examined, higher E354-mediated fasting GIP levels were associated with lower T2D risk (OR per copy of E354, 0.97; 95% CI, 0.96, 0.99; $P=7\times10^{-5}$; **Figure 7.1A**), an effect which strengthened following BMI adjustment (0.93; 95% CI, 0.91, 0.95; $P=3\times10^{-14}$). In line with this, lower 2-hour glucose levels were observed (2-hour glucose in mmol/L per copy of E354, -0.09; 95% CI, -0.11, -0.07; $P=2\times10^{-15}$; **Figure 7.1B**). Additionally, HbA1c levels were shown to be 0.01 SD units lower per copy of E354. I observed no association with either of the insulin outcomes.

Cardiovascular outcomes and related risk factors

In contrast to the T2D association, E354 was associated with higher CHD risk (OR per copy of E354, 1.03; 95% CI, 1.02, 1.05; $P=2x10^{-6}$; **Figure 7.1A**). This was coupled with higher levels of several cardiovascular lipid risk factors including apolipoprotein A1, apolipoprotein B, HDL, low-density lipoprotein and total cholesterol (**Figure 7.1B**). In contrast to the other lipid traits, E354 was associated with lower triglyceride levels. Across stroke sub-types, no significant associations between E354 and stroke were observed aside from a lower risk of small vessel stroke (OR per copy of E354, 0.93; 95% CI, 0.87, 0.98; P=0.009; **Figure 7.1A**). E354 was not significantly associated with other cardiovascular disease subtypes in UKBB, however, a suggestive association with abdominal aortic aneurysm was observed (**Figure S7.1**).

Anthropometric outcomes

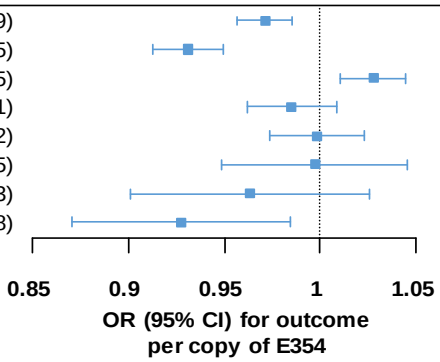
Each copy of E354 was associated with 0.03 SD higher BMI (95% CI, 0.03, 0.04; $P=3\times10^{-59}$; **Figure 7.1B**). This was coupled with similar associations between E354 and higher regional anthropometric measures such as hip and waist circumferences and waist-to-hip ratio. In line with this, significant associations were found with all regional adiposity measures from a large GWAS based on bio-impedance data (**Figure S7.2**).

Biomarker outcomes

Of the 19 biomarkers investigated, E354 was significantly associated with lower levels of only two, namely albumin and creatinine (beta in SD units per copy of E354, -0.01; 95% CI, -0.02, -0.01; $P=6\times10^{-6}$; and -0.02; 95% CI, -0.02, -0.01; $P=1\times10^{-11}$, respectively; **Figure 7.1B**).

A.

Outcome	Cases	Controls	OR (95% CI)	P-value
Type 2 diabetes	74,124	842,006	0.97 (0.96, 0.99)	$3x10^{-4}$
Type 2 diabetes (BMI adj.)	74,124	842,006	0.93 (0.91, 0.95)	$3x10^{-14}$
Coronary heart disease	34,541	261,984	1.03 (1.02, 1.05)	$2x10^{-6}$
Any Stroke	40,585	406,111	0.99 (0.96, 1.01)	0.2
Any Ischemic Stroke	34,217	406,111	1.00 (0.97, 1.02)	0.89
Cardioembolic Stroke	7,193	406,111	1.00 (0.95, 1.05)	0.91
Large Artery Stroke	4,373	406,111	0.96 (0.90, 1.03)	0.24
Small Vessel Stroke	5,386	406,111	0.93 (0.87, 0.98)	0.009



B.

Outcome	N participants	Beta (95% CI)	P-value
Fasting glucose (BMI adj.)	51,750	0.01 ($2x10^{-4}$, 0.02)	0.05
Non-fasted plasma glucose	413,905	0.02 (0.01, 0.02)	$3x10^{-8}$
2-hr glucose (BMI adj.)	41,888	-0.09 (-0.11, -0.07)	$2x10^{-15}$
Fasting insulin (BMI adj.)	51,750	-0.003 (-0.01, -0.004)	0.04
Corrected insulin response	5,318	0.05 (-0.01, 0.11)	0.1
HbA1c	451,782	-0.01 (-0.02, -0.01)	$7x10^{-8}$
ApoA1	412,328	0.01 (0.01, 0.02)	$3x10^{-6}$
HDL	450,957	0.02 (0.01, 0.02)	$7x10^{-9}$
ApoB	448,859	0.02 (0.01, 0.02)	$5x10^{-13}$
LDL	375,774	0.02 (0.02, 0.03)	$2x10^{-16}$
Lipoprotein A	406,825	0.00 (-0.00, 0.01)	0.12
Total cholesterol	377,031	0.02 (0.02, 0.03)	$3x10^{-15}$
Triglycerides	450,625	-0.01 (-0.02, -0.01)	$2x10^{-5}$
C-reactive protein	465,067	-0.01 (-0.01, -0.00)	0.02
BMI	738,628	0.03 (0.03, 0.04)	$3x10^{-59}$
Hip circumference	568,765	0.03 (0.02, 0.03)	$8x10^{-28}$
Hip circumference (BMI adj.)	633,860	-0.01 (-0.01, -0.00)	0.01
Waist circumference	654,577	0.03 (0.03, 0.03)	$4x10^{-42}$
Waist circumference (BMI adj.)	654,253	0.00 (-0.00, 0.00)	0.96
Waist-to-hip ratio	636,672	0.02 (0.02, 0.03)	$2x10^{-22}$
Waist-to-hip ratio (BMI adj.)	636,282	0.00 (-0.00, 0.01)	0.17
Albumin	415,714	-0.01 (-0.02, -0.01)	$6x10^{-6}$
Alkaline phosphatase	450,743	0.01 (0.00, 0.01)	0.006
Alanine aminotransferase	452,291	-0.00 (-0.01, 0.00)	0.77
Aspartate transaminase	450,594	-0.00 (-0.01, 0.00)	0.24
Bilirubin	448,652	0.00 (-0.00, 0.01)	0.1
Calcium	414,173	-0.01 (-0.01, 0.00)	0.07
Creatinine	451,942	-0.02 (-0.02, -0.01)	$1x10^{-11}$
γ-glutamyl transpeptidase	450,745	0.00 (-0.00, 0.01)	0.98
Urate	451,665	-0.00 (-0.01, 0.00)	0.08

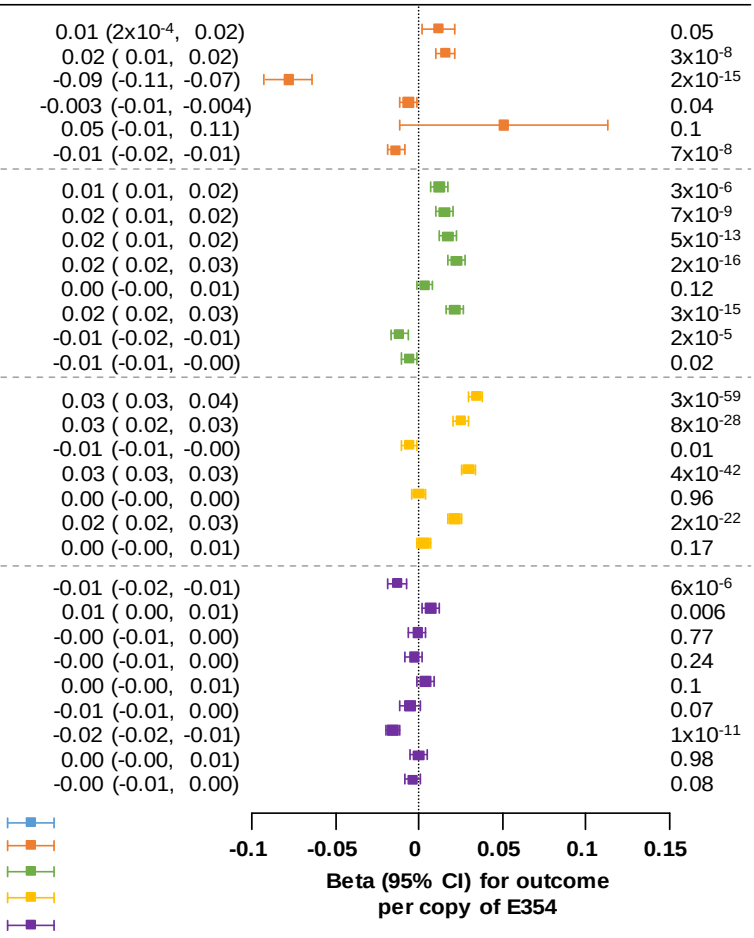


Figure 7.1: Associations between E354 (rs1800437) and cardiometabolic diseases, glycaemic traits, cardiovascular risk factors and lipids, anthropometric traits and biomarkers. Panel A. Associations with cardiometabolic diseases are shown in blue and are represented as odds ratios (95% CI) for each disease per copy of rs1800437. **Panel B.** Associations with glycaemic traits are shown in orange, cardiovascular and lipid traits in green, anthropometric traits and biomarkers are shown in yellow and purple respectively. Estimates are represented as beta (95% CI) for each outcome per copy of rs1800437. All traits are in SD units aside from fasting and 2-hour glucose which are in mmol/L, fasting insulin in log (pmol/L) and HbA1c in mmol/mol. Fold change insulin represents the fold change in insulin levels between fasting to 2-hour measures.

Abbreviations: OR, Odds ratio; CI, Confidence interval; N, Number; BMI, Body mass index; adj., Adjusted; HbA1c, Glycated haemoglobin; ApoA1, Apolipoprotein A1; ApoB, Apolipoprotein B; HDL, High-density lipoprotein; LDL, Low-density lipoprotein; γ, Gamma.

Associations of E354 with plasma protein and metabolite levels

Next, I estimated the association of E354 with the fasting levels 4,979 human proteins from the SOMAscan® v4 system. Significant associations with the levels of three proteins were found (**Figure S7.3**), one of these being 0.08 SD higher fasting GIP levels (95% CI, 0.05, 0.11; $P=4\times10^{-6}$) as measured by SOMAmer 16292-288. Interestingly, this analysis did not find a significant association between the other GIP SOMAmer, 5755-29, and E354. Lower levels of secretoglobin family 3A member 1 (SCGB3A1) and glutaminyl-peptide cyclotransferase-like protein (QPCTL) were also found to be associated with E354. In contrast with a previous report[484], no association between E354 and osteopontin was found.

Lower levels of a single metabolite were found to be significantly associated with E354, an unidentified metabolite named X-12283 (beta in SD units per copy of E354, -0.08; 95% CI, -0.12, -0.05; $P=2\times10^{-5}$; **Figure S7.4**) which was detectable in 8,278 participants. To elucidate the metabolite class and putative function of X-12283, partial correlations were estimated in 11,966 participants between X-12283 and 883 metabolites with less than 50% missingness (**Methods S7.3**). A total of 11 metabolites were significantly correlated with X-12283, of these, five showed a correlation estimate with X-12283 greater than 0.1 (**Figure S7.5**). In addition to significant correlations with unknown metabolites, X-12283 was most significantly correlated with the known amino acid indolepropionate (correlation estimate = 0.23; $P= 2.14\text{x}10^{-51}$; **Figure S7.5**).

Multi-trait colocalisation across cardiometabolic traits

Multi-trait colocalisation between several cardiometabolic risk factors and disease outcomes was run at the *GIPR* locus in the 1Mb regions either side of E354. A total of 425 variants were included in the main analysis, which was limited due to the inclusion of fasting and 2-hour GIP measures from MDC[508], whereas 5,016 were included in the sensitivity analysis (**Table 7.2**). Using a default prior and threshold configuration, 5 distinct trait clusters were identified, 3 of which were shared by both the main and sensitivity analyses (**Table 7.2**). Cluster similarity across all prior and threshold permutations for the two analyses are summarised in heatmaps **Figure 7.2**. Results for all permutations for both the main and sensitivity analyses can be found in **Tables S7.2 and S7.3** respectively.

Of the clusters identified, two distinct clusters were of interest. The first, driven by rs7412 a missense variant in the apolipoprotein E gene (*APOE*), contained CHD and lipid traits – many of which are established cardiovascular risk factors. Both the PP_{coloc} and PP_{causal} were estimated to be 1 in the two analyses, demonstrating robust evidence for colocalisation (**Table 7.2 and Figure S7.6**). This robustness is further emphasised as the same cluster of traits was

identified when using more stringent values for the prior probability of colocalisation (**Figure 7.2, Tables S7.2 and S7.3**).A second cluster consisting of GIP traits, anthropometric traits and 2-hour glucose was driven by rs11672660, which is in perfect LD with rs1800437 (E354; R^2 = 1) (**Table 7.2 and Figure S7.7**). The PP_{coloc} for both analyses showed robust evidence for colocalisation (PP_{coloc}>0.9; $PP_{explained}$=0.99; PP_{causal}=0.96), making it highly likely that rs11672660 was the causal variant. Due to the LD between rs1800437 and rs11672660, the cluster of BMI and waist circumference identified in the sensitivity analysis form part of the cluster with the fasting GIP levels. This is evidenced by the clear association signal across all traits led by rs11672660 (**Figure S7.7**).

Critically, these results replicate my findings using pairwise-trait colocalisation at this locus, showing that fasting GIP levels and cardiovascular risk are driven by independent variants (R^2 between rs11672660 and rs7412 = 0.004) (**Table 7.2; Figures S7.6-S7.8; Figure 7.2**). Additionally, both colocalisation analyses demonstrate that the underlying genetics at *GIPR* are comparable between the relative abundance GIP levels measured by SOMAmer 16292-288 and the absolute quantitation methods of previous analyses[508]. Together these results robustly demonstrate that the GIP-raising and cardiovascular risk increasing effects at this locus are distinct (**Tables S7.2 and S7.3**).

A third cluster including a mixture of glycaemic, anthropometric traits and ApoA1 levels were estimated to colocalise at rs4420638 which was in LD with rs429358 (R^2 = 0.69), a missense variant in *APOE*. In the sensitivity analysis, rs429358 was identified as the candidate variant and T2D was estimated to colocalise with the traits mentioned above. As the sensitivity analysis included more variants and therefore had greater genomic context, rs429358 is likely to be the candidate variant at which these traits colocalise. Supporting this, rs429358 was not present in the main analysis, as this was limited to variants present in all datasets, including fasting GIP measures from MDC which included only exome-wide variants. At default settings, non-fasted glucose was estimated to be part of the cluster, however, at more stringent prior and threshold configurations it was removed, thus replicating the cluster identified in the sensitivity analysis (**Figure 7.2A and Table S7.2**). Again, the high PP_{coloc} demonstrated robust evidence for colocalisation between these traits at rs429358. ApoA1 was estimated to colocalise at rs429358, an independent variant in *APOE*, unlike the other lipid traits which colocalised with CHD at rs7412 (R^2 = 0.69) (**Table 7.2**).

Finally, a cluster between T2D and hip circumference adjusted for BMI was identified in the main analysis, however, was not replicated in the sensitivity (**Table 7.2**). Instead, a cluster between triglycerides and hip circumference adjusted for BMI was identified in the sensitivity

analysis, driven by an independent variant rs5117 (R^2 with rs10408179 = 0.001) (**Table 7.2**). This discrepancy is likely to be resultant of the low number of variants present in the main analysis. When provided with more genomic context, triglycerides were estimated to colocalise with hip circumference adjusted for BMI instead of T2D. Supporting this, rs5117 was not present in the main analysis, as this was limited to variants present in the summary statistics from MDC for fasting GIP as before.

		Table 7.2: Clusters of colocalised traits identified by the main and sensitivity analyses at default settings									
		Main Analysis					Sensitivity Analysis				
Locus	**Candidate LD (R^2)[d]**	**Colocalised Traits**	**PP Coloc[a]**	**Candidate variant**	**PP explained**	**N variants**	**Colocalised Traits**	**PP Coloc[a]**	**Candidate variant**	**PP explained**	**N variants**
GIPR	1	LDL, CHD, HDL, Total Cholesterol, Lipoprotein A, ApoB	1	rs7412	1	425	LDL, CHD, HDL, Total Cholesterol, Lipoprotein A, ApoB	1	rs7412	1	5,016
GIPR	0.69	Glucose, HbA1c, ApoA1, WHRadjBMI, Waist circumference adjBMI, WHR	0.70	rs4420638	1	425	HbA1c, ApoA1, WHRadjBMI, Waist circumference adjBMI, WHR, T2D	0.85	rs429358	1	5,016
GIPR	1	GIP SOMAmer 16292_288, Fasting GIP, 2hr GIP, BMI, Hip circumference, Waist circumference, 2hr Glucose adjBMI	0.97	rs11672660	0.99	425	GIP SOMAmer 16292_288, Hip circumference, 2hr Glucose adjBMI	0.90	rs11672660	0.99	5,016
GIPR	1[b]						BMI, Waist circumference[b]	0.99	rs1800437	0.68	5,016
GIPR	NA	T2D, Hip circumference adjBMI	0.97	rs10408179	1	425					
GIPR	NA						Triglycerides, Hip circumference adjBMI	0.98	rs5117	0.93	5,016

Abbreviations: GIPR, Glucose-dependent insulinotropic polypeptide receptor; LD, Linkage disequilibrium; PP, Posterior probability; coloc, Colocalisation; N, Number; variants, Single nucleotide polymorphisms; LDL, Low-density lipoprotein; CHD, Coronary heart disease; HDL, High-density lipoprotein; ApoB, Apolipoprotein B; Glucose, Non-fasted glucose; ApoA1, Apolipoprotein A1; adj., Adjusted for; WHR, Waist-to-hip ratio; BMI, Body mass index; T2D, Type 2 diabetes

a. Traits are reported at default settings for Hyprcoloc: Prior 2 = 0.02; regional and alignment thresholds = 0.5; thus, the posterior probabilities will vary dependent on the prior configuration
b. The rs1800437 and rs11672660 variants are in perfect LD ($R^2 = 1$) meaning that the BMI + WC cluster at the *GIPR* locus, reported separately in the sensitivity analysis, forms part of the larger cluster with the fasting GIP traits as in the main analysis. The BMI + WC cluster has therefore not been counted as an independent cluster in results
c. Blank rows for either analysis indicate a cluster not identified in the respective analysis
d. The LD in R^2 between the candidate variants for the main and sensitivity analyses respectively

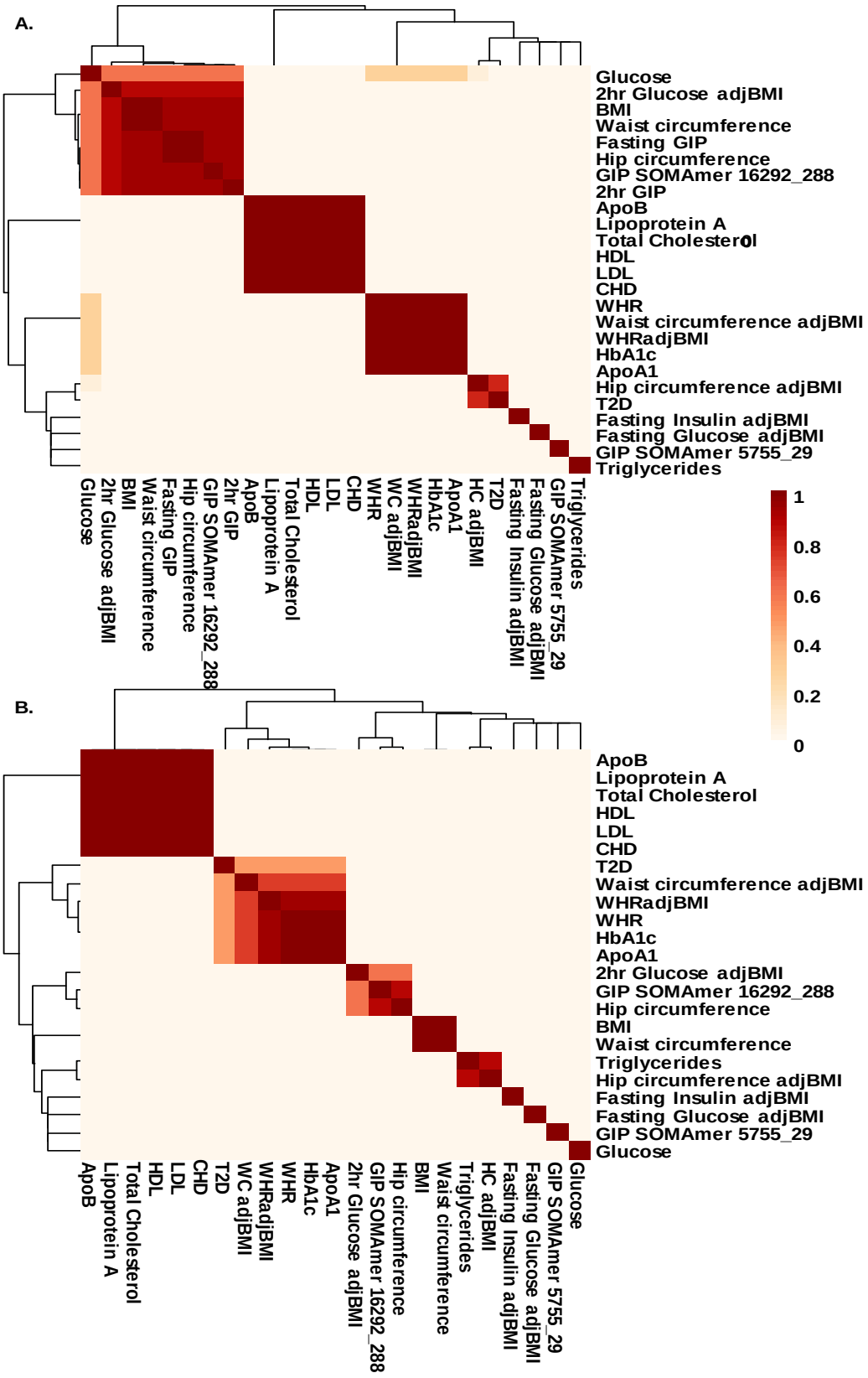


Figure 7.2: Similarity heatmap for each cluster at the *GIPR* locus across prior and threshold permutations. Traits that were estimated to colocalise are clustered together. Darker colours represent traits which were estimated to colocalise more often across prior and threshold permutations. **Panel A**. Main analysis. **Panel B**. Sensitivity analysis

Abbreviations: LDL, Low-density lipoprotein; CHD, Coronary heart disease; HDL, High-density lipoprotein; ApoB, Apolipoprotein B; Glucose, Non-fasted glucose; ApoA1, Apolipoprotein A1; adj., Adjusted for; WHR, Waist-to-hip ratio; BMI, Body mass index; T2D, Type 2 diabetes; WC, Waist circumference; HC, Hip circumference

Conditional analysis at the *GIPR* locus

The univariate two-sample MR results showed that E354 was associated with a total of 20 traits at a nominal significance threshold (**Figure 7.1**). Independent signal selection showed that E354, or proxy variants in high LD ($R^2 > 0.9$) with E354, were identified as independent signals for fasting GIP, 2-hour glucose, total cholesterol levels, BMI and X-12283 levels. Independent signal selection identified a total of 24 variants independently associated with CHD on chromosome 19, four of which were in the 1Mb regions either side of E354 at the *GIPR* locus (**Table 7.3**). Conditioning the association between E354 and CHD on the residual LD between E354 and rs7412, the variant estimated to drive the cluster with CHD, showed a slight attenuation of this association (OR per copy of E354 after adjustment 1.03; 95% CI, 1.02, 1.04; P=0.003), however, the association remained significant. Of the independent variants identified, rs1964272 an intronic variant in small nuclear ribonucleoprotein D2 polypeptide (*SNRPD2*), was estimated to be in the strongest LD with E354 (R^2=0.27) (**Figure 7.3 and Figure S7.9**). The association between E354 and CHD risk was attenuated when conditioned on rs1964272 (OR per copy of E354 after adjustment, 1.01; 95% CI, 0.99, 1.03; P = 0.06) (**Table S7.4**). In line with this, the association between E354 and small vessel stroke was also attenuated when conditioning on rs1964272 (**Table S7.4**), however, the association of E354 with non-fasted glucose remained nominally significant (P=0.05). None of the other loci previously reported as associated with fasting GIP levels were found to be associated with CHD (**Table S7.5**). Interestingly, rs1964272 was also shown to be associated with levels of the proteins QPCTL and SCGB3A1 indicating confounding by LD for the proteomics data as well (**Figure S7.10**). In accordance with this, the association between E354 and levels of QPCTL was also attenuated to non-significance when conditioned on rs1964272 (beta QPCTL per copy of E354 after adjustment, 0.01; 95% CI, -0.02, 0.04; P=0.48). The association between E354 and SCGB3A1 levels was attenuated when conditioned on rs61703905 but remained significant (**Table S7.4**).

Table 7.3: Independent CHD variants identified using approximate conditional analysis.

Variant [a]	Chr:pos	EA	EAF	Marginal Beta (SE) [b]	Marginal P-value [b]	Conditional Beta (SE) [c]	Conditional P-value [c]	N	R^2 with rs1800437
rs429358	19:45411941	T	0.85	-0.09 (0.008)	$2.86x10^{-27}$	-0.08 (0.008)	$5.87x10^{-23}$	286,423	0.001
rs7412	19:45412079	T	0.08	-0.14 (0.011)	$1.66x10^{-35}$	-0.12 (0.011)	$1.58x10^{-28}$	275,803	0.004
rs11673093	19:45742094	A	0.26	0.04 (0.007)	$4.11x10^{-11}$	0.04 (0.007)	$3.09x10^{-10}$	300,789	0
rs1964272	19:46190268	A	0.48	-0.03 (0.006)	$9.65x10^{-9}$	-0.03 (0.006)	$1.87x10^{-7}$	299,519	0.27

Abbreviations: Chr, Chromosome; pos, Position; EA, Effect allele; EAF, Effect allele frequency; SE, Standard error; N, Number of participants; R^2, Linkage disequilibrium estimate

a. The independent CHD variants in the 1Mb region either side of E354 are shown
b. Log odds ratios from the original GWAS summary statistics
c. Log odds ratios from the joint model fitted by GCTA

Much like the CHD association, conditioning the associations between E354 and lipid traits on rs7412 did show partial attenuation for ApoB and LDL, however, these associations remained significant (**Table S7.4**). In line with this, conditioning the association of E354 with triglycerides on rs4803936, an intronic variant in protein phosphatase 5 catalytic subunit (*PPP5C*), also remained significant, suggesting that E354 has independent effects on lipid metabolism. Interestingly, the same was not true for the associations of E354 with ApoA1 and HDL when conditioning on rs2238689, an intronic variant in *GIPR* (R^2 with E354=0.36), as these associations were totally attenuated (**Table S7.4**).

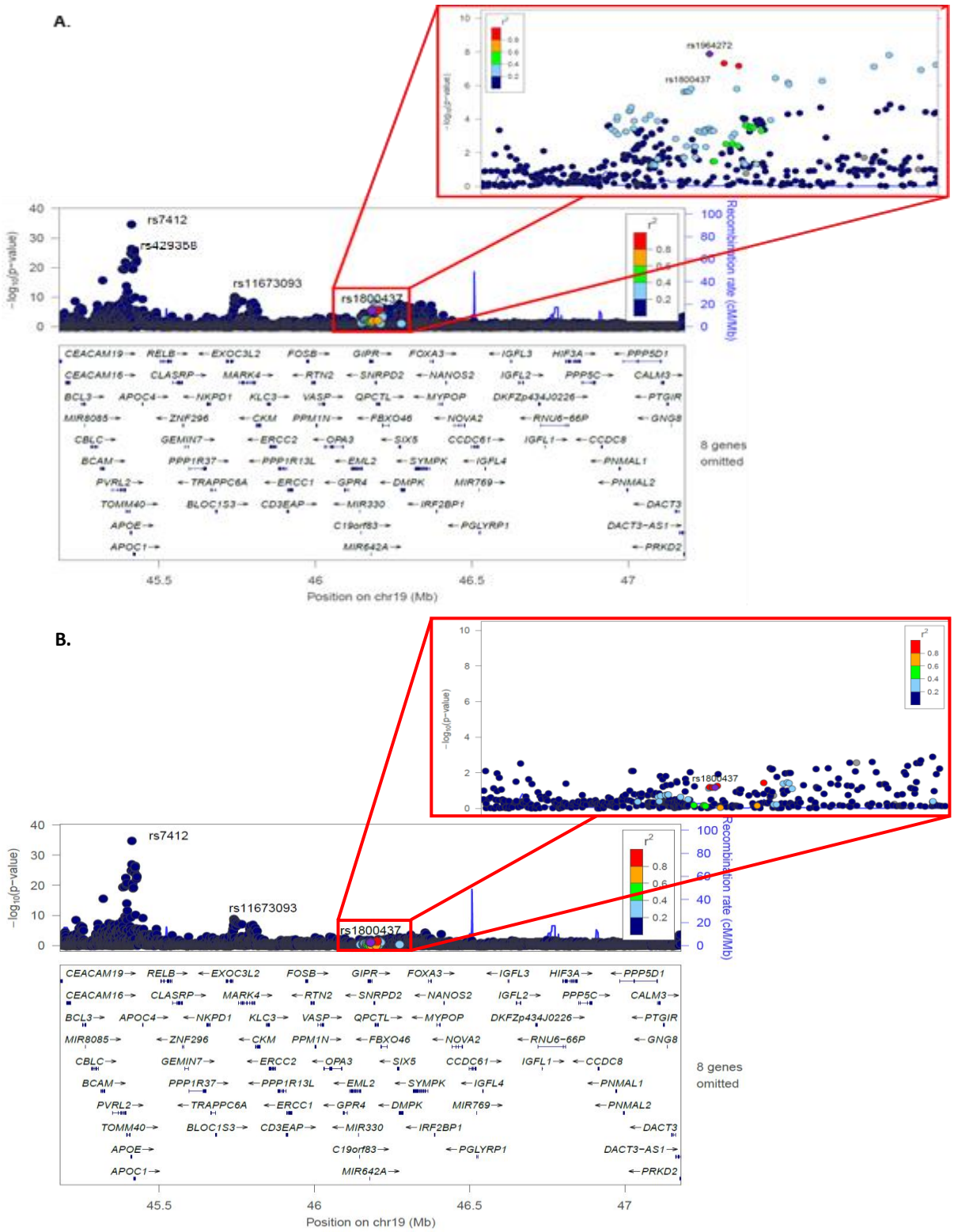


Figure 7.3: Regional association plots depicting CHD lead variants in the *GIPR* region. Panel A. The independent CHD lead variants in the *GIPR* region are labelled and their respective associations with CHD are shown before conditional analysis. The region around rs1800437 (E354) is expanded in the red insert to show the LD and proximity of rs1964272 to rs1800437. **Panel B.** The associations of variants in the *GIPR* region after conditioning on rs1964272. The region around rs1800437 (E354) is expanded in the red insert to show the attenuation of the E354 signal when conditioned on rs1964272.

7.4: Discussion

Recently, evidence linking higher fasting GIP levels with increased CVD risk[481,484] has led to safety concerns surrounding GIPR agonism via either selective GIPR agonists or dual-GIPR and GLP-1R agonists[481]. As new T2D therapies cannot confer clinically significant increases in CVD risk[543], these findings may be considered detrimental to therapeutics targeting this pathway.

In this study, I applied Bayesian multi-trait colocalisation and conditional analysis to gain greater resolution into the underlying genetic architecture of CHD and its relation to fasting GIP levels at the *GIPR* locus. Results from multi-trait colocalisation robustly partitioned the colocalisation of CHD and lipid traits to a cluster at *APOE* that was independent from a cluster of fasting and 2-hour GIP, glycaemic and anthropometric traits driven by E354. Several factors highlight the robust partitioning of these two distinct clusters of colocalised metabolic traits at the *GIPR* locus, (1) colocalisation analyses were consistent between the main and sensitivity analyses, (2) the same variant was identified as putatively causal in each analysis, and (3) traits included in each cluster were stable at increasingly stringent prior and threshold configurations. The clustering of these traits at independent variants, is likely to represent independent biological mechanisms. APOE has a central role in lipid metabolism and transport in blood[544] and *APOE* is an established gene associated with both plasma lipid levels[396] and CHD risk[534]. In contrast, the traits clustering at *GIPR* have all been associated with E354 previously and are in line with the established functions of GIP[508,528,529]. Results from conditional analysis demonstrated a robust attenuation of the E354 association with CHD, small vessel stroke and QPCTL levels when adjusted for rs1964272, an intronic variant in *SNRPD2*. This, together with the multi-trait colocalisation results, demonstrates that these association signals for CVD at *GIPR* are not mediated by an independent effect of *GIPR* variants on CVD risk but are instead the result of low-level LD between E354 and rs1964272. This variant lies within an established CHD risk locus[289,534] and functions as part of a spliceosomal complex mediating pre-mRNA splicing[534], however, its role in CHD pathophysiology is unclear.

The conditional analysis results also highlighted the effects of E354 on lipid metabolism and BMI. E354 is an established variant associated with higher BMI[529] and was identified as an independent variant in this analysis. Studies in animal models have posited a role for fasting GIP in lipid metabolism, suggesting that the effects of GIP on lipid deposition[479,480] may contribute towards weight gain and higher BMI[481]. The conditional analysis results demonstrate that while the association between E354 and some lipid traits is attenuated when

conditioning on rs7412 in *APOE*, the association remains significant. This suggests that *GIPR*-mediated higher fasting GIP levels have an independent effect on some lipid levels.

Taken together, these findings demonstrate the specificity of E354's effects on fasting GIP levels and crucially demonstrate that higher E354-mediated fasting GIP levels are not associated with CVD risk. I therefore provide evidence that the inclusion of GIPR agonists in dual agonists simultaneously targeting GIPR and GLP-1R could potentiate the protective effect of GLP-1 agonists on diabetes without undue cardiovascular risk, an aspect which has yet to be assessed in clinical trials. In addition, results from clinical trials of dual agonists have shown that the effects of GIP on weight gain are counteracted by GLP-1, as significant weight loss was demonstrated, while potentiating improved glycaemic control and protection against T2D[503,504]. Many studies have shown that GLP-1R agonism achieved through chronic pharmacologic therapy, or genetic gain of function, is associated with improved cardiovascular outcomes[482,524–527]. Hence, the totality of available evidence suggests that dual agonism of these receptors may exploit the metabolically favourable combined pharmacology of these incretins without undue CVD risk. However, this proposition requires formal assessment in clinical trials such as the recently initiated SURPASS cardiovascular outcomes trial of the novel GIP/GLP-1R dual agonist tirzepatide (clinicaltrials.gov registered trials number NCT04255433).

This study has potential limitations. Firstly, this analysis focuses on a single locus associated with both fasting GIP levels and CHD. In doing so, this assumes that the *GIPR* locus is a suitable proxy for fasting GIP levels within which to partition the associations of these two complex traits. Considering that the association with 2-hour glucose is statistically robust and in line with the established function of GIP, this is a reasonable assumption. In addition, no other locus has been reported to be associated with both fasting GIP and CHD, therefore, this is likely to be the only locus where these traits overlap. Examining the association of other variants found to be associated with fasting GIP levels[508] showed no association of any of these variants with CHD[289]. This does not, however, exclude the possibility that other variants that have not yet been shown to be associated with GIP levels, do not contribute to CVD risk. Secondly, this analysis made use of fasting GIP measures from aptamer-based technology, which used two SOMAmers to target overlapping portions of the GIP precursor peptide which are present in the active GIP peptide. To enable fine-scale control of insulin potentiation, active GIP is rapidly degraded by dipeptidyl peptidase 4 (DPP-4), cleaving the first two amino acids from active GIP(1-42) to create GIP(3-42), which has been shown to antagonise GIPR[418,489]. Due to the nature of the technology it is unable to discriminate between active and degraded forms of the GIP peptide, a limitation common to previously used GIP quantitation

techniques[508]. Thirdly, protein abundances were measured in plasma samples, where protein function and abundance may differ from that measured intracellularly and is likely to include purposefully secreted as well as leaked proteins.

This study highlights the specificity of E354 for fasting GIP levels and robustly demonstrates that CVD risk and fasting GIP levels are driven by independent causal variants in the region surrounding *GIPR*. I also demonstrate that *GIPR*-mediated cardiovascular risk is in fact the result of low-level LD between E354 and rs1964272 and not the result of independent effects of *GIPR* variants on CVD risk. Taken together, these findings are promising for the development of therapeutics designed to pharmacologically modulate this pathway in T2D patients, as *GIPR*-mediated effects are highly specific and do not confer undue CVD risk.

Chapter 8: Concluding Discussion

8.1: Summary of the findings

This PhD thesis set out to identify the aetiological pathways underpinning the roles of inflammatory biomarkers and incretins, two areas of considerable pharmacological interest, in cardiometabolic diseases. This was based on the integration of large-scale genetic data in deeply phenotyped cohorts with multiple 'omics datasets to further our understanding of the molecular underpinnings of T2D and CHD.

Studies have shown that obesity is accompanied with an inflammatory response[109,138,223,356] which has been shown to be associated with insulin resistance[138,223,270], diabetes[109,223,334,356] and atherosclerosis[4,90,155,160,383]. In Chapter 2, the observational associations of four proinflammatory cytokines IFNγ, IL-6, IL-8 and TNFα with cardiometabolic diseases and related risk factors were estimated. Levels of IFNγ, IL-6 and TNFα were associated with all measures of adiposity that were considered and most lipid levels, whereas IL-8 levels were only associated with WHR and HbA1c. In line with the hypothesis that higher BMI is a major driver of inflammation in cardiometabolic diseases[94,112,123,223,545], adjustment for BMI attenuated the majority of these associations. IL-6, TNFα and to a lesser extent IFNγ, remained associated with lower HDL but higher triglycerides, fasting insulin, HbA1c and WHR, hallmarks of an insulin resistance phenotype[33,345,346,390]. Similarly, IL-6 and TNFα were shown to be associated with both incident T2D and CHD, despite extensive adjustment for established cardiometabolic risk factors.

Next, to gain a better understanding of the genetics underpinning the levels of these four cytokines, I conducted a GWAS in approximately 17,000 individuals, the largest cytokine GWAS to date. A total of 22 variant-cytokine associations were identified in 19 loci, 16 of which were novel and outside the *MHC* gene region. Six of these were associated with IFNγ levels, three with IL-6 levels, five with IL-8 levels and six with TNFα levels. Missense variants in the receptors for the respective cytokines were identified: *DARC* (IL-8)*, IL6R* (IL-6)*, CXCR1/2* (IL-8) and *IFNGR1* (IFNγ), making them ideal candidates for follow-up. While these variants explained the greatest proportion of variance in levels of the respective cytokines, only the *IL6R* region showed strong evidence of a shared signal underlying cytokine levels and cardiometabolic diseases, warranting further follow-up of only this region.

In a meta-analysis of 15 prospective studies higher IL-6 levels were associated with incident T2D risk. Closer examination of the Asp358Ala variant in a large trans-ethnic meta-analysis showed an association with lower odds of T2D. This was conditionally independent of a nearby association peak for T2D, consistent across ethnic groups, and unlikely to be influenced by potential diagnostic misclassification. In line with the observational findings, the 358Ala allele was associated with lower HbA1c levels and showed a significant interaction with higher BMI on HbA1c levels. Mediation analyses showed that IL-6 levels mediated up to 5% of the association between higher BMI and T2D, this was 3-4 times smaller than for CHD. This, together with rescaled projections of IL-6R antagonism in T2D of below 10% risk reduction, suggested that while IL-6 mediated inflammation plays a role in the aetiology of T2D, the impact of this pathway on disease risk in the general population may be small.

Next, I integrated large-scale genomic and aptamer-based proteomics data in a Bayesian multi-trait colocalisation framework to systematically identify inflammatory proteins and pathways associated with cardiometabolic diseases. Considering *cis* genetic association data for 335 inflammatory proteins, a total of 164 and 181 showed strong evidence of a shared signal with T2D and CHD, respectively. The candidate variant at 95% and 85% of the regions shared between inflammatory proteins and either T2D or CHD, respectively, was an established pleiotropic locus. In addition, 42% and 36% of inflammatory proteins which shared a signal with T2D and CHD, respectively, were estimated to be driven by established risk factors for cardiometabolic diseases. Consistent with an insulin resistance phenotype, the triglycerides and WHRadjBMI scores were associated with levels of the greatest number of proteins for both diseases.

Finally, recent observational and genetic epidemiological studies using a missense variant in *GIPR*, E354, suggested an association between higher GIPR-mediated fasting GIP levels and CHD risk[481], a concern for therapeutic development targeting GIPR agonism in T2D. Using a Bayesian multi-trait colocalisation framework in the region around *GIPR*, I showed that GIP, adiposity and glycaemic traits shared a signal at E354, a missense variant in *GIPR*, whereas CHD and lipid traits shared an independent signal at rs7412, a missense variant in *APOE.* However, the low LD between rs7412 and E354 was estimated not to be driving the association of E354 with CHD. Instead, the association of E354 with CHD was completely attenuated when conditioned on rs1964272, an established CHD signal in an intronic region in *SNRPD2*.

In this final chapter, I will discuss the strengths and limitations of the research presented in this thesis and how this may have impacted the findings. Next, I will consider the implications of my findings and outline future directions for research in this field.

8.2: Strengths

To my knowledge, the work presented in this thesis represents the most comprehensive and systematic investigation of the role of inflammation in the aetiology of T2D and CHD. This work leveraged large-scale genomic data in deeply phenotyped cohorts such as UK Biobank and integrated this with large, publically available datasets to ensure that all analyses were suitably powered. This work examined the role of inflammation on both a locus and global perspective to identify and characterise the inflammatory proteins and pathways involved in the aetiology of cardiometabolic diseases. My work on cytokine levels used a systematic evidence-based approach resulting in the largest GWAS of cytokine levels to date, further elucidating the genetic determinants of IL-6 and TNFα but most importantly, providing evidence of the first genome-wide evidence of variants associated with IL-8 and IFNγ levels. I also conducted the most comprehensive epidemiological and human genetics study to date on the association between IL-6 mediated inflammation and the risk of T2D. This work integrated trans-ethnic genetic data from UK Biobank and publically available datasets to demonstrate a consistent effect of this pathway across ancestries that is not due to diagnostic misclassification with type 1 diabetes. I provide the first human genetics evidence supporting the long-standing hypothesis of a role for chronic inflammation in T2D and used this evidence to project the effects of IL-6 antagonism in T2D, an area not yet addressed by RCTs.

Later, this was expanded to identify inflammatory proteins and pathways involved in the aetiology of cardiometabolic diseases by integrating genomic data with large-scale aptamer-based proteomics data. This work used contemporary genetic methods to address several important aetiological questions. Firstly, this work identified inflammatory proteins which may be involved in T2D and CHD risk and highlighted loci underpinning inflammation in cardiometabolic disease risk. Secondly, using polygenic scores, an approach robust to reverse causality, I demonstrated that inflammation in cardiometabolic diseases is likely to be driven by both pleiotropy and established cardiometabolic risk factors.

The work on GIP presented in this thesis illustrates how genetics may be used to partition the effects of variants to assess the safety of a therapeutic target. Genomic, metabolomic and proteomic data were integrated to investigate the relationship between GIPR signalling and CHD risk, using E354 as an instrument for GIPR agonism. Integrating these biological layers helped to characterise the molecular effects of GIPR agonism on the proteomic and metabolomic profile, highlighting the specificity of therapeutically targeting this pathway in T2D. This work had several advantages over previous efforts. Firstly, the entire genomic region surrounding *GIPR* was considered facilitating a comprehensive investigation into the effects of other variants which may dampen or modulate the effects of E354 on CHD risk.

Secondly, this facilitated an in-depth assessment of the LD in the region to determine whether the association between E354 and CHD risk was an entirely synthetic effect. This work has promising implications for pharmacological modulation of this pathway. I showed that GIP, anthropometric and glycaemic traits are driven by a variant that is distinct from a variant driving CHD risk in the *GIPR* region. Finally, using conditional analysis I demonstrated that cardiovascular risk was not estimated to be mediated via GIPR, suggesting that therapeutics targeting this pathway would not unduly raise cardiovascular risk.

8.3: Limitations

Approximation of pharmacological agonism or antagonism using genetic variants

In **Chapter 4** and **Chapter 6** of this thesis, missense variants, Asp358Ala and E354Q, were used as genetic instruments for pharmacological antagonism and agonism of the IL-6R and GIPR, respectively. Experimental evidence[310] has demonstrated that the 358Ala allele is a major determinant of soluble IL-6R levels, leading to a reduction in IL-6R cell surface expression and impaired cellular responsiveness to IL-6. Similar functional evidence[490] has indicated that the *GIPR* missense variant 354Q results in an exaggerated reduction in GIPR expression on the plasma membrane following GIP stimulation, coupled with a delayed recovery in GIP sensitivity after GIP stimulation, mimicking pharmacological antagonism. The work presented in **Chapter 7** used the other allele of this variant, E354 as an instrument for GIPR agonism. However, there are several key differences between pharmacological modulation of these receptors compared to the effects of naturally occurring genetic variants which mean that genetic variants may not accurately approximate the effects of pharmacological modulation.

Firstly, unlike pharmacological antagonism or agonism which may be complete or incomplete, the alleles of these common variants represent partial loss- and gain-of-function respectively. In addition, genetic variants are usually associated with small differences in the activity of the target gene, while drugs typically have larger effects on target activity[222,353]. It should therefore be noted that the effects of these variants are likely to differ with respect to the effects of pharmacological antagonism or agonism of these receptors. Using these variants as proxies for pharmacological modulation of these receptors is likely to have resulted in an underestimation of the potential pharmacological effects of targeting these receptors. This may mean that smaller, potentially detrimental effects were overlooked, thus this approach cannot be used to definitively rule out safety concerns of therapeutically targeting a pathway.

Second, these genetic variants represent a lifelong chronic exposure to antagonism and agonism respectively, which is not reflective of the comparatively shorter, higher dosage acute administrations of pharmacological therapy. As shown in **Chapter 4**, this limitation can be accounted for by re-scaling genetic effect estimates to be comparable to the effects of pharmacological antagonism of IL-6R estimated in clinical trials[222,353], where these are available, thus estimating the potential effect of IL-6R antagonism in T2D. However, the fact remains that 358Ala is representative of partial loss-of-function, therefore, the projections of potential pharmacological efficacy of IL-6R antagonism in T2D are likely to be underestimated. With respect to the E354 *GIPR* variant, re-scaling the estimate to predict the potential therapeutic benefit of GIPR agonism was not possible as clinical trial data for GIPR mono-

agonists are not available, as therapeutic development has focussed instead on GLP-1R mono agonists[458,495,498] or dual GLP-1R/GIPR agonism[503,504,520,521]. Overall, this approach enables the estimation of an association between pharmacologically modulating a pathway and a beneficial effect on an outcome of interest, in a framework that is robust to non-genetic confounding. This is therefore not an accurate estimate of the potential effect size of a pharmacological intervention.

Finally, the projections of therapeutic efficacy in **Chapter 4** assume additive risk of the IL-6 pathway on type 2 diabetes risk, which is in line with assumptions of genetic association studies of common variants and previously used Mendelian randomization approaches in this and other settings. However, the therapeutic efficacy may be non-linear and vary with time[62], an effect which was not tested for. Projections of therapeutic efficacy in scenarios where the association between the exposure and outcome is non-linear and/or time-dependent may lead to inaccurate estimates.

Limitations to causal inference

In **Chapter 5** of this thesis, I used Bayesian multi-trait colocalisation to systematically assess *cis*-acting genetic regulation shared between inflammatory proteins and cardiometabolic diseases. The aim being to identify inflammatory proteins that are potentially involved in cardiometabolic diseases and thereby explore the inflammatory pathways which participate in the aetiology of cardiometabolic diseases. In this case using a multivariable MR framework would have provided an empirical estimate of the causal relevance of each identified protein with disease, adjusted for established cardiometabolic risk factors[74], however, this was not possible for several reasons. Firstly, comparatively few variants were found to be associated with levels of the inflammatory proteins considered in the study, an issue common to the cytokine analyses in **Chapter 3**. In practice, MR analyses require sufficient numbers of variants to reliably estimate the causal relevance of the exposure for the outcome, which was not possible for these analyses. Secondly, a single *cis*-acting variant, either in the protein encoding gene or in the protein's receptor, often has a disproportionately large effect on protein levels. This disproportionally large effect may lead to biased causal estimates, particularly when few variants are associated with levels of the protein[61]. An example of this was shown in **Chapter 3** where a single *cis*-acting variant in the receptor for the respective cytokine explained the greatest proportion of variance in cytokine levels. Inclusion of these variants in an MR model made lead to biased estimates, purely driven by these variants of large effect. Third, the analysis showed that some of the variants underlying protein levels were pleiotropic. As described in **Chapter 1**, horizontal pleiotropy invalidates the instrumental variable assumptions for MR[62], these variants are associated with many diverse traits and

would therefore need to be removed in an MR analysis[74]. Considering the low numbers of variants associated with inflammatory protein levels, this would decrease the number of variants further. Fourth, variants in the MHC gene region are often removed from MR analyses due to the complex LD in the region. This particularly impacts inflammatory proteins as often variants in the MHC region are associated with their levels. Despite this, in **Chapter 5**, I was able to estimate the association of polygenic scores for cardiometabolic risk factors with inflammatory protein levels, the reverse causal direction. This analysis showed that the levels of a proportion of inflammatory protein levels are associated with established cardiometabolic risk factors, effectively ruling them out as independent risk factors for cardiometabolic diseases.

Colocalisation analyses assume a single causal variant

In this thesis, Bayesian pairwise[284] and multi-trait colocalisation[362] analyses were used in **Chapters 3**, **5 and 6**. While these methods enable the identification of shared genetic signals between traits or trait-sets, both methods share a common limitation where only a single independent causal variant is assumed to be present in a region of interest[284,362]. This study used relatively small regions of interest (1Mb-2Mb) to reduce the likelihood of this assumption being violated, however, it is possible that this may have been violated despite taking this into consideration. The likelihood of violating this assumption increases with an increasing number of traits[362]. Considering that in **Chapter 5**, a total of 341 traits (2 disease traits and 339 inflammatory proteins) were included in the multi-trait colocalisation model for each of the 85 loci considered, the likelihood of violating this assumption for at least one of the traits is considerable. For this reason, sensitivity analyses were conducted where the colocalisation model was run over increasingly stringent prior and threshold configurations. In addition, simulations using HyPrColoc have demonstrated that multiple independent variants in a region rarely invalidate the results, so long as secondary variants in the same region do not explain more trait variance than the sentinel variant[362]. To that end, the proportion of variance explained was estimated for both sentinels and secondary signals, only the *ABO* locus was deemed to have potentially violated this assumption, thus estimates need to be interpreted with caution.

Associations between genetic variants and phenotypes may be false positives, biased or study-specific

Genetic variants which reach genome-wide significance for an association with a phenotype of interest may not be truly associated with the trait due to chance, study-specific effects, confounding by LD or bias. One of the most challenging aspects of GWAS is refining association signals to truly causal variants[91]. While statistical methods such as fine mapping

were applied in the cytokine GWAS presented in **Chapter 3**, there were some loci where it was challenging to resolve the association signal to a single variant. Therefore, it may be that some of the results are confounded by LD. However, the approach used in **Chapter 7** overcame the limitations of previous analyses[481] by using conditional analyses to conclude that CHD risk was not mediated via GIPR. Throughout this study, every effort has been made to mitigate the possibility of LD confounding.

Many GWAS studies adopt a widely used a Bonferroni genome-wide significance threshold of $P \leq 5x10^{-8}$. This threshold was first used in 2005, the advent of early GWAS studies[546], and is based on the estimated number of common (MAF > 5%), independent variants that distinguish haplotypes throughout the genome[547]. This threshold has subsequently been used to limit the chance of false positive findings (Type 1 error) to a tolerable proportion of 5%. However, with the development of denser imputation reference panels that are widely adopted in contemporary GWAS, several studies[547] have advocated for a revision of this threshold to ensure that the false positive rate remains constrained at 5%. The converse is also true, meaning that variants which are truly associated with a trait, may be deemed to be non-significant (Type 2 error). In order to limit the impact of Type 2 error, the largest sample sizes were used where possible. Throughout this study, Bonferroni thresholds have been used to ensure that the Type 1 error is limited to 5%. In **Chapter 3** of this thesis, the conventional genome-wide significance threshold of $P \leq 5x10^{-8}$ was used. Despite this some of the loci estimated to be associated with cytokine levels may have reached significance by chance and may therefore be false positives. To ensure that their findings are truly associated with the phenotype of interest, many GWAS studies typically use a two-stage discovery and replication study design[548,549]. This also ensures that the genetic variants' effects replicate across several studies, limiting the chances of identifying a study-specific effect. In **Chapter 3**, variants associated with cytokine levels were identified in EPIC-Norfolk and Fenland, however, as no studies with a sufficiently large sample size were available, replication of these findings was not possible. However, consistency in the effect size and direction was observed between the two studies, minimising the likelihood of these findings being study-specific or due to chance.

This thesis used participants from three population-based studies: UK Biobank[66], EPIC-Norfolk[241] and Fenland[244]. A commonly observed phenomenon in population-based studies is the healthy volunteer effect, whereby individuals who are generally healthier and of better socioeconomic status relative to the general population are more likely to participate in studies[550]. In the case of UK Biobank there is evidence for this effect, study participants were leaner, less likely to smoke and drink alcohol on a daily basis and were comparatively of better socioeconomic status than the general UK population[551]. No formal comparisons of either

Fenland or EPIC-Norfolk relative to the general population are currently available, however, it must be assumed that this is a likely limitation[243]. Where possible, publically available datasets from large consortium-based efforts were used to ensure the generalisability and specificity of results. These have the advantage of being highly powered due to large sample sizes and use strict disease definitions and inclusion criteria to ensure specificity.

The field of human genetics is heavily biased towards sampling participants of predominately European ancestry[552–554]. This bias has led to a systematic underrepresentation of participants from non-European ancestry, limiting the generalisability of scientific research findings[67]. For example, it is currently unknown whether IL-6R antagonism or GIPR agonism exhibit ethnicity- or sex-specific effects. Clinical trials of these therapies have included participants of different ethnicities[503,555–557], however, this was comparatively fewer than the number of white European participants included. The efficacy of these therapeutics was demonstrated in these trials, however, due to the low numbers of participants from other ethnicities included, the trial may not have been sufficiently powered to detect differences in drug efficacy between ethnicities. Given that drug metabolism varies with ethnicity[558] and sex[559], this is an important consideration for ongoing research. However, ethnicity and sex specific GWAS analyses are rarely conducted and were unavailable for this study. Therefore, this needs to be considered for future work to ensure the generalisability of these findings. Genetic research stands to benefit greatly from the greater inclusion of participants from diverse ancestral backgrounds for several key reasons. Firstly, GWAS studies have identified many common variants which are associated with disease; however, these explain a fraction of disease heritability. Therefore, the effects of rare variants are believed to be increasingly important[67,553] and these variants are likely to be specific to particular populations. In addition, variant allele frequencies vary according to ancestry, thus inclusion of participants from diverse ancestral backgrounds provides greater statistical power to discover variants associated with phenotypes of interest[552]. Secondly, as a result of homologous recombination events, the size of LD blocks also vary with ancestry[560]. This property is extremely beneficial for the fine-mapping of association signals to identify causal variants if they are shared between ancestries[67].

From an ethical standpoint the lack of ancestral diversity in genetic studies is a barrier to the generalisability to therapeutic development and efficacy[561]. Developing therapeutics on the basis of research conducted solely in Europeans does not guarantee therapeutic efficacy when administered in individuals of other ancestries[552]. This form of sampling bias also hinders our understanding of the aetiology of cardiometabolic diseases and risk factors, as these are well established for European populations but less so for individuals of African or South-Asian ancestry, despite these populations bearing greater burden of cardiometabolic diseases[1].

Therefore, expanding the ancestral diversity of genetic and biomedical research in general is key to facilitating a better understanding of the aetiology of cardiometabolic diseases.

Statistical uncertainty

Statistical error in epidemiology refers to the random, uncorrelated error which may arise during phenotypic measurement within and between studies. In large cohort studies such as UK Biobank[66], every effort is made to adhere to standardised, rigorous protocols for phenotypic assessment to minimise sources and magnitude of error. Error in epidemiology is inversely related to sample size, therefore, increasing sample sizes tends to minimise error, thereby also increasing statistical power of the study[562]. In this thesis, meta-analysis was performed using inverse-variance weighting, ensuring that cohorts were weighted according to their sample size and therefore minimising the contribution of random error to the final standard error reported for each effect estimate. Where applicable, formal heterogeneity assessments were undertaken to ensure that estimates were directly comparable across studies or cohorts. Finally, results from the largest available dataset for each phenotype of interest to this study were used in each case.

8.4: Directions for future research and clinical implications

The need for larger GWAS to facilitate the discovery of further variants associated with biomarkers

Biomarkers such as inflammatory proteins and incretins are quantifiable intermediate traits which often lie on the causal pathway between established risk factors and disease outcomes and may therefore mediate some of the risk between exposures and disease[563]. The results presented in **Chapter 5** suggest that this is likely to be the case for a proportion of inflammatory proteins which colocalise with T2D and CHD, levels of which were shown to be affected by established cardiometabolic risk factors. However, in order to gain a better understanding of the genetics underpinning biomarker levels, larger and better powered GWAS studies are required. Statistical power in GWAS is directly related to the sample size used[564], therefore larger sample sizes are required to enable discovery of additional variants associated with biomarker levels. One of the biggest challenges faced throughout the work outlined in this thesis was the comparatively few genetic variants that have been found to be associated with biomarker levels. This, coupled with the finding that a considerable proportion of the identified loci were pleiotropic, hindered the capacity to empirically estimate their causal relevance for cardiometabolic diseases using an MR framework.

The cytokine GWAS presented in **Chapter 3** had a sample size of approximately 17,000 participants across the four cytokines, the largest sample size used to date. Despite this, a maximum of six loci were estimated to be associated with levels of IFNγ and TNFα. The same is true for variants associated with incretin levels, the largest study to date had a sample size of up to 7,828 participants and found a total of two variants associated with fasting GIP levels, five associated with 2-hr GIP levels and one associated with 2-hr GLP-1 levels. It is conceivable that cytokine and incretin levels may be governed by relatively few variants, some of which (e.g. missense variants in cytokine receptors) may have large effects on the trait, but it is also likely that studies simply do not have sufficient power to discover all variants associated with their levels. The contrast in participant numbers is particularly stark when considering the number of individuals included in GWAS studies of anthropometric traits present in large biobanks such as UK Biobank[66] for example BMI[565] or WHRadjBMI[30]. However, this discrepancy can be easily explained when considering the ease, speed and negligible cost of measuring anthropometric traits compared to biomarker measurements which require trained staff for blood extraction, laboratory equipment for blood processing and storage as well as expensive assays for quantification.

Ideally measurement of these biomarkers in large multinational, trans-ethic biobanks using gold standard assay techniques would facilitate large trans-ethnic GWAS efforts to discover

further variants associated with biomarker levels, however, in reality this would be extremely expensive and time consuming. It should also be considered whether plasma is the best tissue to study (discussed in detail below). Another option would be to leverage large-scale proteomic profiling methods that simultaneously quantify the levels of thousands of proteins, such as that used in this study, in large trans-ethnic biobanks. This would afford greater statistical power to discover variants associated with biomarker levels. However, this would require the normalisation of protein levels to an internal standard, therefore protein levels represent relative abundances which are incomparable across studies. In addition, questions surrounding the specificity of the technology in comparison to gold-standard assays such as ELISA need to be considered. Nevertheless, biomarker quantification in large sample sizes is necessary to facilitate thorough investigations of the genetics governing their levels and characterisation of the molecular pathways through which biomarkers act to mediate cardiometabolic risk.

Expanding the scope to include different tissues

Obesity is thought to be one of the main drivers for inflammation in cardiometabolic diseases[109,138,223,356]. Inflammation, is therefore hypothesised to partly mediate obesity-related risk in cardiometabolic diseases, as shown in **Chapter 4**. Logically and of direct relevance to cardiometabolic diseases, many epidemiological studies have focussed their attention on inflammatory proteins produced by adipose tissue such as IL-6[90,155,175,181,186,189] and TNFα[168,174,175,189,190]. While the liver has previously been considered to be a non-immunological organ, accumulating evidence suggest that inflammatory and metabolic processes converge and interact here, influencing the regulation of many inflammatory protein levels[566]. It is currently unclear how hepatic cross-talk may affect metabolic and inflammatory pathways and how this interaction may relate to risk of cardiometabolic diseases.

Most studies have investigated the role of the liver in cardiometabolic diseases within the context of non-alcoholic fatty liver disease (NAFLD) and how hepatic lipid accumulation may influence this phenotype[567–574]. It would be of interest to investigate the influence of NAFLD and other metabolic dysregulations on inflammatory protein levels and estimate the causal relevance of NAFLD for levels of inflammatory proteins. This work could leverage the release of liver MRI data in UK Biobank[66], combined with ultrasonography data in Fenland to create a score of liver fat for use in GWAS. Using variants associated with this phenotype, one could then leverage the availability of large-scale aptamer based proteomic data from SomaLogic in Fenland to estimate the causal relevance of hepatic lipid accumulation for inflammatory protein levels. Additionally, one could estimate the colocalisation between inflammatory protein levels and NAFLD associated variants, providing an insight into any possible shared aetiology

between NAFLD and inflammatory proteins. However, while expanding this work to other metabolically relevant tissues will help to contextualise the role of inflammation in cardiometabolic diseases, investigations into the role of biomarkers in other diseases should also be conducted.

Consideration of other diseases

One of the most advantageous aspects of investigating biomarkers as a phenotype is that they represent intermediate traits for many diseases and are readily druggable[55]. It is this property which makes biomarkers attractive drug targets for a range of diseases and presents interesting drug re-purposing opportunities outside the scope of the original drug indication[575]. Relevant examples include the re-positioning of IL-1Ra, IL-6R and TNFα receptor monoclonal antibody antagonists for use in rheumatoid arthritis patients[555,576,577]. Recently the CANTOS trial has investigated the use of canakinumab, an IL-1β antagonist, in both CHD[226] and T2D[221]. Many diseases are deemed to have a chronic inflammation component to their aetiology, for example type 1 diabetes, asthma and inflammatory bowel disease[94]. It would be possible to apply similar multi-trait colocalisation methods to those used in **Chapter 5** to estimate the colocalisation between inflammatory proteins and autoimmune diseases. This work could also leverage summary statistics conditioned on other independent signals within a locus to aid the estimation of causal variants in colocalisation analyses. Therefore, expanding this research to autoimmune conditions may further elucidate the underpinnings of chronic inflammation, highlight whether the genetic loci identified are different from those identified to play a role in cardiometabolic diseases. Furthermore, the inclusion of type 1 diabetes in colocalisation analyses with T2D would elucidate inflammatory pathways common to both diseases while also highlighting distinct aetiological pathways. Such cross-disease approaches may be conducted efficiently using genetics and may highlight aetiological pathways shared between cardiometabolic and autoimmune conditions while also discriminating between pathways specific to each disease type.

Investigation of incretins within the context of metabolic disease as a whole

Incretins have been shown to play a role in glucose stimulated insulin potentiation[412,422] and are therefore a promising pharmacological target for T2D[489,500]. In **Chapter 6** of this thesis I investigated the association between *GIPR* variants and cardiovascular risk. Experimental evidence has pointed to a role for GIP in triglyceride storage and adipose expansion, suggesting that GIP has more diverse metabolic functions than just insulin potentiation[459,578,579]. In addition, the effects of both GIP and GLP-1 have been shown to be associated with reduced body weight[455–457,471,475,580]. On balance, and considering the results presented in this study, GIP may offer promising protective effects on a wide range of

metabolic phenotypes and may therefore be suitable targets for the generalised improvement of metabolic health. However, our results suggest that GIPR agonism may lead to higher BMI. On this basis, further characterisation of the effects of variants associated with higher GIP and GLP-1 levels respectively on diverse metabolic traits would be of interest, specifically their effects on lipid profile and BMI. Therapeutic development in this field has shifted from designing mono-agonists for GIPR and GLP1R towards dual GIPR/GLP1R agonism[503,504,520,521]. In light of this, using genetics to investigate the possible interaction between GIP and GLP-1 levels on lipid profile and BMI as well as T2D and CHD risk would be of substantial interest. This work may provide estimates of the potential efficacy of dual agonism in both diseases and help to characterise the molecular effects of dual agonism with respect to how these relate to disease risk.

A focus on the in-depth characterisation of genetic loci

Throughout the history of GWAS for complex traits, a significant focus has been placed on ever-expanding sample sizes to achieve greater statistical power to detect variants associated with phenotypes of interest[91]. While I have advocated for this in the sections above, GWAS has already led to the identification of many thousands of loci associated with diverse traits[91]. However, the biological mechanism underlying many of these variant-trait associations are still yet to be elucidated. Indeed, in the case of anthropometric traits such as obesity, it could be argued that increasingly larger GWAS efforts are likely to lead to marginal gains in biological understanding over and above what is already known. For example, increasing the sample size for a GWAS of protein levels mainly leads to an increase in the number of *trans*-acting variants discovered. Therefore, while increasing sample sizes is beneficial, this should be balanced with emphasis on characterising the molecular effects of established loci.

To that end, the cytokine GWAS presented in **Chapter 3** highlighted variants in genes such as *CXCR1-CXCR2* and *IFNGR1* that are associated with IL-8 and IFNγ levels respectively. Further characterisation of the rs55799208 variant, in particular, at the *CXCR1-CXCR2* locus would be of interest as this is a missense variant in a receptor specific for IL-8, making this an ideal candidate for further follow-up. In light of the observational results presented in **Chapter 2**, using this variant to further characterise the association between IL-8 and HbA1c levels would be of interest, particularly when considering that IL-8 levels was not found to be associated with any other glycaemic traits.

Clinical implications

The work presented in this thesis integrated large-scale genetic data on cardiometabolic traits with multiple 'omics datasets to investigate the role of inflammatory biomarkers and incretins in the aetiology of cardiometabolic diseases. The observational findings of this study highlight that IL-6 and TNFα levels are associated with a wide variety of risk factors for cardiometabolic diseases as well as incident T2D and CHD. Expanding on these findings using genetics to further investigate the role of IL-6R-mediated inflammation in T2D demonstrated that IL-6 levels were associated with prevalent T2D. However, while IL-6 mediated inflammation plays a role in the aetiology of T2D, several lines of evidence suggest that the impact of this pathway on disease risk in the general population may be small. Firstly, people in the top decile of IL-6 levels had a 2-fold higher risk of incident T2D than people in the bottom decile. This is approximately three times smaller than the relative risk associated with being in the top decile of BMI in our study. While the association between Asp358Ala and lower T2D risk was statistically robust, consistent across studies and ancestries and was unaffected by possible misclassification of type 1 diabetes cases, the odds ratio of 0.98 for disease risk was a marginal reduction in disease risk compared to non-carriers. In addition, rescaled projections of IL-6R antagonism in T2D estimate risk reductions of below 10% in most simulated scenarios. In comparison, relatively safe and inexpensive lifestyle interventions[28], metformin[28] or other hypoglycaemic drugs[357] have been associated with reductions in the incidence of diabetes ranging from 30% to 70% in previous primary prevention trials. Finally, the proportion of BMI-associated risk of diabetes mediated by IL-6 levels was below 5%, ~3-4 times smaller than for coronary disease. Taken together, these results suggest that targeting inflammation via this pathway may not substantially impact on diabetic risk in obesity but still represents a biologically significant effect. While these results suggest that IL-6R antagonism may have limited efficacy for the primary prevention of T2D, these findings do not exclude that IL-6R inhibition may yield greater benefits on glucose metabolism in groups of people at risk for T2D. Results from observational and genetic analyses demonstrated that IL-6 levels were associated with HbA1c levels with genetic analyses showing a significant interaction between IL-6 levels and higher BMI on HbA1c levels. In line with this, evidence from clinical trials[212,359] has shown a statistically-significant improvement in HbA1c in rheumatoid arthritis patients randomized to IL-6R antagonists as opposed to placebo. Therefore, it is possible that IL-6R inhibition may yield clinically meaningful changes in glycaemia in people with inflammatory or immune conditions which are linked with diabetes and that reductions in HbA1c may be more pronounced in individuals below the median of the BMI polygenic score.

The GIP pathway has been highlighted as a potential therapeutic for use in T2D and therapies targeting GIPR/GLP-1 dual agonism are in development[503,504,520]. Clinical evidence suggests

that these therapies offer promising metabolic benefits such as improved insulin potentiation, lipid profile and weight management in T2D patients[503,504,520]. However, recent epidemiological evidence has suggested that agonism of GIPR may raise CHD risk[481]. This work used a genetic framework to investigate the relationship between higher GIPR-mediated GIP levels and CHD risk. The results showed that fasting GIP levels and CHD risk were driven by independent variants. In addition, the association between GIP levels and CHD risk was found to be the result of residual LD between E354, a missense variant in *GIPR*, and rs1964272, an intronic variant in *SNRPD2*. As a result, these results have positive implications for the development of dual GIPR/GLP1R agonists targeting this pathway, as these are not estimated to unduly raise CHD risk.

8.5: Conclusion

The work presented in this PhD integrates large-scale genetic data with multiple 'omics datasets to investigate the role of inflammatory biomarkers and incretins in the aetiology of cardiometabolic diseases, two areas that are of considerable pharmacological interest. The role of inflammation in cardiometabolic diseases was examined on both a locus-level and global perspective. Overall, the results of this study suggest that chronic inflammation in cardiometabolic diseases is largely driven by a combination of pleiotropic effects and established risk factors for cardiometabolic diseases. This work closely examined the aetiological role of proinflammatory cytokines in cardiometabolic diseases. In the largest GWAS of cytokine levels to date, I provide the first genetic variants associated with IL-8 and IFNγ at a genome-wide significance level and expand the knowledge of the genetic determinants underpinning IL-6 and TNFα levels. Evidence from observational and genetic analyses suggest that IL-6 and TNFα levels are more relevant for both incident and prevalent cardiometabolic diseases compared to IFNγ or IL-8. Further examination of the role of IL-6 mediated inflammation in T2D provided the first human genetics evidence supporting the long-standing hypothesis of a role for chronic inflammation in T2D. However, the results suggested that the impact of this pathway on T2D risk in the general population is likely to be small.

In this PhD thesis I also used a genetic framework to investigate the role of GIPR agonism, a pathway of considerable pharmacological promise for T2D, in raising CHD risk. I provide evidence that GIP, anthropometric and glycaemic traits are driven by distinct variants from those raising CHD risk and lipids in the *GIPR* region. I also demonstrate that raised CHD risk in the *GIPR* region is not driven by *GIPR* but instead by *SNRPD2*, a nearby CHD risk locus. This result holds considerable promise for continued therapeutic research targeting this pathway for use in T2D and obesity.

References

1 International Diabetes Federation. IDF Diabetes Atlas 2019, 9th Edition. 2019 http://www.idf.org/about-diabetes/facts-figures.

2 World Health Organization. Global Report on Diabetes. 2016; **978**: 88.

3 Thomas H, Diamond J, Vieco A, *et al.* Global Atlas of Cardiovascular Disease 2000-2016: The Path to Prevention and Control. *Glob Heart* 2018; **13**: 143–63.

4 Libby P, Buring JE, Badimon L, *et al.* Atherosclerosis. *Nat Rev Dis Prim* 2019; **5**: 56.

5 Lotta LA, Stewart ID, Sharp SJ, *et al.* Association of Genetically Enhanced Lipoprotein Lipase-Mediated Lipolysis and Low-Density Lipoprotein Cholesterol-Lowering Alleles with Risk of Coronary Disease and Type 2 Diabetes. JAMA Cardiol. 2018; **3**: 957.

6 Hendel RC, Jabbar AY, Mahata I. Initial diagnostic evaluation of stable coronary artery disease: The need for a patient-centered strategy. J. Am. Heart Assoc. 2017; **6**. DOI:10.1161/JAHA.117.006863.

7 Grundy SM, Becker D, Clark LT, *et al.* Detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III). Circulation. 2002; **106**: 3143–421.

8 Grundy SM, Cleeman JI, Bairey Merz CN, *et al.* Implications of recent clinical trials for the National Cholesterol Education Program Adult Treatment Panel III guidelines. Circulation. 2004; **110**: 227–39.

9 Goldfine AB, Shoelson SE. Therapeutic approaches targeting inflammation for diabetes and associated cardiovascular risk. *J Clin Invest* 2017; **127**: 83–93.

10 Roden M, Shulman GI. The integrative biology of type 2 diabetes. *Nat 2019 5767785* 2019; **576**: 51–60.

11 Hex N, Bartlett C, Wright D, Taylor M, Varley D. Estimating the current and future costs of Type1 and Type2 diabetes in the UK, including direct health costs and indirect societal and productivity costs. *Diabet Med* 2012; **29**: 855–62.

12 Caro JJ, Ward AJ, O'Brien JA. Lifetime costs of complications resulting from type 2 diabetes in the U.S. *Diabetes Care* 2002; **25**: 476–81.

13 World Health Organization. Fact sheet No.317: Cardiovascular diseases (CVDs). Available from World Heal. Organ. site, http//www.who.int/mediacentre/factsheets/fs317/en/. 2013. https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) (accessed Jan 22, 2020).

14 Benjamin EJ, Blaha MJ, Chiuve SE, *et al.* Heart Disease and Stroke Statistics'2017 Update: A Report from the American Heart Association. Circulation. 2017; **135**: e146–603.

15 Shah AD, Langenberg C, Rapsomaniki E, *et al.* Type 2 diabetes and incidence of cardiovascular diseases: A cohort study in 1·9 million people. *Lancet Diabetes Endocrinol* 2015; **3**: 105–13.

16 Booth GL, Kapral MK, Fung K, Tu J V. Relation between age and cardiovascular disease in men and women with diabetes compared with non-diabetic people: a population-based retrospective cohort study. *Lancet* 2006; **368**: 29–36.

17 Beckman JA, Paneni F, Cosentino F, Creager MA. Diabetes and vascular disease: pathophysiology, clinical consequences, and medical therapy: part II. *Eur Heart J* 2013; **34**: 2444–52.

18 Stumvoll M, Goldstein BJ, Van Haeften TW. Type 2 diabetes: Principles of pathogenesis and therapy. In: Lancet. Elsevier, 2005: 1333–46.

19 Nolan CJ, Damm P, Prentki M. Type 2 diabetes across generations: From pathophysiology to prevention and management. In: The Lancet. Elsevier, 2011: 169–81.

20 Lowe WL. Principles of Molecular Medicine, 16th editi. New Jersey, NJ, USA: Humana Press, 2005.

21 Savage DB, Petersen KF, Shulman GI. Disordered lipid metabolism and the pathogenesis of insulin resistance. Physiol. Rev. 2007; **87**: 507–20.

22 Mahajan A, Taliun D, Thurner M, *et al.* Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet* 2018; **50**: 1505–13.
23 Shai I, Jiang R, Manson JAE, *et al.* Ethnicity, obesity, and risk of type 2 diabetes in women: A 20-year follow-up study. *Diabetes Care* 2006; **29**: 1585–90.
24 Admiraal WM, Holleman F, Snijder MB, *et al.* Ethnic disparities in the association of impaired fasting glucose with the 10-year cumulative incidence of type 2 diabetes. *Diabetes Res Clin Pract* 2014; **103**: 127–32.
25 Van Valkengoed IGM, Argmann C, Ghauharali-Van Der Vlugt K, *et al.* Ethnic differences in metabolite signatures and type 2 diabetes: A nested case-control analysis among people of South Asian, African and European origin. *Nutr Diabetes* 2017; **7**: 300.
26 Scott RA, Langenberg C, Sharp SJ, *et al.* The link between family history and risk of type 2 diabetes is not explained by anthropometric, lifestyle or genetic risk factors: The EPIC-InterAct study. *Diabetologia* 2013; **56**: 60–9.
27 NCD Risk Factor Collaboration (NCD-RisC) B, Lu Y, Hajifathalian K, *et al.* Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. *Lancet (London, England)* 2016; **387**: 1513–30.
28 Diabetes Prevention Program Research Group. Reduction in the Incidence of Type 2 Diabetes With Lifestyle Intervention or Metformin. *N Engl J Med* 2002; **346**: 393–403.
29 Wang Y, Rimm EB, Stampfer MJ, Willett WC, Hu FB. Comparison of abdominal adiposity and overall obesity in predicting risk of type 2 diabetes among men. *Am J Clin Nutr* 2005; **81**: 555–63.
30 Lotta LA, Wittemans LBL, Zuber V, *et al.* Association of Genetic Variants Related to Gluteofemoral vs Abdominal Fat Distribution With Type 2 Diabetes, Coronary Disease, and Cardiovascular Risk Factors. *JAMA* 2018; **320**: 2553.
31 White M. Population Approaches to Prevention of Type 2 Diabetes. *PLOS Med* 2016; **13**: e1002080.
32 Hu FB, Manson JE, Stampfer MJ, *et al.* Diet, Lifestyle, and the Risk of Type 2 Diabetes Mellitus in Women. *N Engl J Med* 2001; **345**: 790–7.
33 Semple RK, Savage DB, Cochran EK, Gorden P, O'Rahilly S. Genetic Syndromes of Severe Insulin Resistance. *Endocr Rev* 2011; **32**: 498–514.
34 Helmerhorst HJF, Wijndaele K, Brage S, Wareham NJ, Ekelund U. Objectively measured sedentary time may predict insulin resistance independent of moderate- and vigorous-intensity physical activity. *Diabetes* 2009; **58**: 1776–9.
35 Hu FB, Li TY, Colditz GA, Willett WC, Manson JE. Television Watching and Other Sedentary Behaviors in Relation to Risk of Obesity and Type 2 Diabetes Mellitus in Women. *JAMA* 2003; **289**: 1785.
36 Giugliano RP, Wiviott SD, Blazing MA, *et al.* Long-term Safety and Efficacy of Achieving Very Low Levels of Low-Density Lipoprotein Cholesterol. *JAMA Cardiol* 2017; **2**: 547.
37 Di Angelantonio E, Gao P, Pennells L, *et al.* Lipid-related markers and cardiovascular disease prediction. *JAMA - J Am Med Assoc* 2012; **307**: 2499–506.
38 Ference BA, Ginsberg HN, Graham I, *et al.* Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. Evidence from genetic, epidemiologic, and clinical studies. A consensus statement fromthe European Atherosclerosis Society Consensus Panel. *Eur Heart J* 2017; **38**: 2459–72.
39 Libby P. Inflammation in atherosclerosis. *Arterioscler Thromb Vasc Biol* 2012; **32**: 2045–51.
40 Gisterå A, Hansson GK. The immunology of atherosclerosis. Nat. Rev. Nephrol. 2017; **13**: 368–80.
41 Robbins CS, Hilgendorf I, Weber GF, *et al.* Local proliferation dominates lesional macrophage accumulation in atherosclerosis. *Nat Med* 2013; **19**: 1166–72.
42 Bennett MR, Sinha S, Owens GK. Vascular Smooth Muscle Cells in Atherosclerosis. *Circ Res* 2016; **118**: 692–702.
43 Bentzon JF, Otsuka F, Virmani R, Falk E. Mechanisms of plaque formation and rupture. *Circ Res* 2014; **114**: 1852–66.

44 Bachmann JM, Willis BL, Ayers CR, Khera A, Berry JD. Association between family history and coronary heart disease death across long-term follow-up in men: The cooper center longitudinal study. *Circulation* 2012; **125**: 3092–8.
45 Mochari-Greenberger H, Mosca L. Differential Outcomes by Race and Ethnicity in Patients with Coronary Heart Disease: A Contemporary Review. Curr. Cardiovasc. Risk Rep. 2015; **9**. DOI:10.1007/s12170-015-0447-4.
46 Mochari-Greenberger H, Liao M, Mosca L. Racial and ethnic differences in statin prescription and clinical outcomes among hospitalized patients with coronary heart disease. *Am J Cardiol* 2014; **113**: 413–7.
47 Jousilahti P, Vartiainen E, Tuomilehto J, Puska P. Sex, age, cardiovascular risk factors, and coronary heart disease: A prospective follow-up study of 14 786 middle-aged men and women in Finland. *Circulation* 1999; **99**: 1165–72.
48 Zhao M, Vaartjes I, Graham I, *et al.* Sex differences in risk factor management of coronary heart disease across three regions. *Heart* 2017; **103**: 1587–94.
49 Lu Y, Hajifathalian K, Ezzati M, *et al.* Metabolic mediators of the effects of body-mass index, overweight, and obesity on coronary heart disease and stroke: A pooled analysis of 97 prospective cohorts with 1·8 million participants. *Lancet* 2014; **383**: 970–83.
50 Huxley RR, Woodward M. Cigarette smoking as a risk factor for coronary heart disease in women compared with men: A systematic review and meta-analysis of prospective cohort studies. *Lancet* 2011; **378**: 1297–305.
51 Murtagh EM, Nichols L, Mohammed MA, Holder R, Nevill AM, Murphy MH. Walking to improve cardiovascular health: a meta-analysis of randomised control trials. *Lancet* 2014; **384**: S54.
52 Wood AM, Kaptoge S, Butterworth A, *et al.* Risk thresholds for alcohol consumption: combined analysis of individual-participant data for 599 912 current drinkers in 83 prospective studies. *Lancet* 2018; **391**: 1513–23.
53 Franklin SS, Larson MG, Khan SA, *et al.* Does the relation of blood pressure to coronary heart disease risk change with aging?: The Framingham Heart Study. *Circulation* 2001; **103**: 1245–9.
54 Goode GK, Miller JP, Heagerty AM. Hyperlipidaemia, hypertension, and coronary heart disease. Lancet. 1995; **345**: 362–4.
55 Finan C, Gaulton A, Kruger FA, *et al.* The druggable genome and support for target identification and validation in drug development. *Sci Transl Med* 2017; **9**: 1–16.
56 Plenge RM, Scolnick EM, Altshuler D. Validating therapeutic targets through human genetics. Nat. Rev. Drug Discov. 2013; **12**: 581–94.
57 Nelson MR, Tipney H, Painter JL, *et al.* The support of human genetic evidence for approved drug indications. *Nat Genet* 2015; **47**: 856–60.
58 Hingorani AD, Kuan V, Finan C, *et al.* Improving the odds of drug development success through human genomics: modelling study. *Sci Rep* 2019; **9**: 1–25.
59 Dowden H, Munro J. Trends in clinical success rates and therapeutic focus. *Nat Rev Drug Discov* 2019; **18**: 495–6.
60 Wittemans LBL, Lotta LA, Langenberg C. Prioritising Risk Factors for Type 2 Diabetes: Causal Inference through Genetic Approaches. Curr. Diab. Rep. 2018; **18**: 1–10.
61 Smith GD, Ebrahim S. 'Mendelian randomization': Can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003; **32**: 1–22.
62 Holmes M V., Ala-Korpela M, Smith GD. Mendelian randomization in cardiometabolic disease: Challenges in evaluating causality. Nat. Rev. Cardiol. 2017; **14**: 577–99.
63 Ference BA, Robinson JG, Brook RD, *et al.* Variation in PCSK9 and HMGCR and risk of cardiovascular disease and diabetes. *N Engl J Med* 2016; **375**: 2144–53.
64 King EA, Wade Davis J, Degner JF. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. *PLoS Genet* 2019; **15**: 1–20.
65 Szustakowski JD, Balasubramanian S, Sasson A, *et al.* Advancing Human Genetics Research and Drug Discovery through Exome Sequencing of the UK Biobank. *medRxiv*

2020; : 2020.11.02.20222232.
66 Sudlow C, Gallacher J, Green J, *et al.* UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Med* 2015; **12**: e1001779.
67 Langenberg C, Lotta LA. Genomic insights into the causes of type 2 diabetes. Lancet. 2018; **391**: 2463–74.
68 Karczewski KJ, Snyder MP. Integrative omics for health and disease. Nat. Rev. Genet. 2018; **19**: 299–310.
69 Smith GD, Hemani G. Mendelian randomization: Genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* 2014; **23**: 89–98.
70 Burgess S, Thompson DJ, Rees JMB, Day FR, Perry JR, Ong KK. Dissecting Causal Pathways Using Mendelian Randomization with summarized genetic data: application to ge at menarche and risk of breast cancer. *Genetics* 2017; **207**: 481–7.
71 Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* 2013; **37**: 658–65.
72 Pingault JB, O'Reilly PF, Schoeler T, Ploubidis GB, Rijsdijk F, Dudbridge F. Using genetic data to strengthen causal inference in observational research. Nat. Rev. Genet. 2018; **19**: 566–80.
73 Evans DM, Davey Smith G. Mendelian Randomization: New Applications in the Coming Age of Hypothesis-Free Causality. *Annu Rev Genomics Hum Genet* 2015; **16**: 327–50.
74 Burgess S, Thompson SG. Multivariable Mendelian randomization: The use of pleiotropic genetic variants to estimate causal effects. *Am J Epidemiol* 2015; **181**: 251–60.
75 Burgess S, Dudbridge F, Thompson SG. Combining information on multiple instrumental variables in Mendelian randomization: Comparison of allele score and summarized data methods. *Stat Med* 2016; **35**: 1880–906.
76 Cho Y, Haycock PC, Sanderson E, *et al.* Exploiting horizontal pleiotropy to search for causal pathways within a Mendelian randomization framework. *Nat Commun* 2020; **11**: 1010.
77 Bowden J, Spiller W, Del Greco FM, *et al.* Improving the visualization, interpretation and analysis of two-sample summary data Mendelian randomization via the Radial plot and Radial regression. *Int J Epidemiol* 2018; **47**: 1264–78.
78 Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet* 2018; **50**: 693–8.
79 Bowden J, Del Greco M F, Minelli C, Davey Smith G, Sheehan N, Thompson J. A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Stat Med* 2017; **36**: 1783–802.
80 Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet* 2017; **13**: 1–22.
81 Richardson TG, Sanderson E, Palmerid TM, *et al.* Evaluating the relationship between circulating lipoprotein lipids and apolipoproteins with risk of coronary heart disease: A multivariable Mendelian randomisation analysis. *PLoS Med* 2020; **17**: 1–22.
82 Taylor AE, Davies NM, Ware JJ, Vanderweele T, Smith GD, Munafò MR. Mendelian randomization in health research: Using appropriate genetic variants and avoiding biased estimates. *Econ Hum Biol* 2014; **13**: 99–106.
83 Smith GD, Hemani G. Mendelian randomization: Genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* 2014; **23**: R89–98.
84 Swerdlow DI, Preiss D, Kuchenbaecker KB, *et al.* HMG-coenzyme A reductase inhibition, type 2 diabetes, and bodyweight: Evidence from genetic analysis and randomised trials. *Lancet* 2015; **385**: 351–61.
85 Preiss D. Risk of Incident Diabetes With Intensive-Dose Compared With Moderate-Dose Statin Therapy. *JAMA* 2011; **305**: 2556.
86 Sattar N, Preiss D, Murray HM, *et al.* Statins and risk of incident diabetes: a collaborative meta-analysis of randomised statin trials. *Lancet* 2010; **375**: 735–42.

87 Pushpakom S, Iorio F, Eyers PA, *et al.* Drug repurposing: Progress, challenges and recommendations. *Nat Rev Drug Discov* 2018; **18**: 41–58.
88 Swerdlow DI, Holmes M V., Kuchenbaecker KB, *et al.* The interleukin-6 receptor as a target for prevention of coronary heart disease: A mendelian randomisation analysis. *Lancet* 2012; **379**: 1214–24.
89 Ridker PM, Rifai N, Stampfer MJ, Hennekens CH. Plasma Concentration of Interleukin-6 and the Risk of Future Myocardial Infarction Among Apparently Healthy Men. *Circulation* 2000; **101**: 1767–72.
90 Danesh J, Kaptoge S, Mann AG, *et al.* Long-term interleukin-6 levels and subsequent risk of coronary heart disease: Two new prospective studies and a systematic review. *PLoS Med* 2008; **5**: 0600–10.
91 Visscher PM, Wray NR, Zhang Q, *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet* 2017; **101**: 5–22.
92 Schaid DJ, Chen W, Larson NB. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet* 2018; : 1.
93 Lotta LA, Sharp SJ, Burgess S, *et al.* Association Between Low-Density Lipoprotein Cholesterol–Lowering Genetic Variants and Risk of Type 2 Diabetes. *Jama* 2016; **316**: 1383.
94 Furman D, Campisi J, Verdin E, *et al.* Chronic inflammation in the etiology of disease across the life span. *Nat Med* 2019; **25**: 1822–32.
95 Turvey SE, Broide DH. Innate immunity. *J Allergy Clin Immunol* 2010; **125**. DOI:10.1016/j.jaci.2009.07.016.
96 Netea MG, Balkwill F, Chonchol M, *et al.* A guiding map for inflammation. Nat. Immunol. 2017; **18**: 826–31.
97 Kotas ME, Medzhitov R. Homeostasis, Inflammation, and Disease Susceptibility. Cell. 2015; **160**: 816–27.
98 Fullerton JN, Gilroy DW. Resolution of inflammation: A new therapeutic frontier. *Nat Rev Drug Discov* 2016; **15**: 551–67.
99 Liston A, Masters SL. Homeostasis-altering molecular processes as mechanisms of inflammasome activation. *Nat Rev Immunol* 2017; **17**: 208–14.
100 Cani PD, Jordan BF. Gut microbiota-mediated inflammation in obesity: a link with gastrointestinal cancer. Nat. Rev. Gastroenterol. Hepatol. 2018; **15**: 671–82.
101 Le Chatelier E, Nielsen T, Qin J, *et al.* Richness of human gut microbiome correlates with metabolic markers. *Nature* 2013; **500**: 541–6.
102 Chassaing B, Van De Wiele T, De Bodt J, Marzorati M, Gewirtz AT. Dietary emulsifiers directly alter human microbiota composition and gene expression ex vivo potentiating intestinal inflammation. *Gut* 2017; **66**: 1414–27.
103 Richards JL, Yap YA, McLeod KH, Mackay CR, Mariño E. Dietary metabolites and the gut microbiota: an alternative approach to control inflammatory and autoimmune diseases. *Clin Transl Immunol* 2016; **5**: e82.
104 Slavich GM, Irwin MR. From stress to inflammation and major depressive disorder: A social signal transduction theory of depression. *Psychol Bull* 2014; **140**: 774–815.
105 Tobaldini E, Fiorelli EM, Solbiati M, Costantino G, Nobili L, Montano N. Short sleep duration and cardiometabolic risk: from pathophysiology to clinical evidence. Nat. Rev. Cardiol. 2019; **16**: 213–24.
106 Reutrakul S, Van Cauter E. Sleep influences on obesity, insulin resistance, and risk of type 2 diabetes. Metabolism. 2018; **84**: 56–66.
107 Booth FW, Roberts CK, Laye MJ. Lack of Exercise Is a Major Cause of Chronic Diseases. In: Comprehensive Physiology. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2012. DOI:10.1002/cphy.c110025.
108 Hayashino Y, Jackson JL, Hirata T, *et al.* Effects of exercise on C-reactive protein, inflammatory cytokine and adipokine in patients with type 2 diabetes: A meta-analysis of randomized controlled trials. *Metabolism* 2014; **63**: 431–40.
109 Ouchi N, Parker JL, Lugus JJ, Walsh K. Adipokines in inflammation and metabolic disease. Nat. Rev. Immunol. 2011; **11**: 85–97.

110 Lackey DE, Olefsky JM. Regulation of metabolism by the innate immune system. *Nat Rev Endocrinol* 2015; **12**: 15–28.
111 Guilherme A, Virbasius J V., Vishwajeet P, Czech MP, Puri V, Czech MP. Adipocyte dysfunctions linking obesity to insulin resistance and type 2 diabetes. *Nat Rev Mol Cell Biol* 2008; **9**: 367–77.
112 Hotamisligil GS. Inflammation, metaflammation and immunometabolic disorders. *Nature* 2017; **542**: 177–85.
113 Reilly SM, Saltiel AR. Adapting to obesity with adipose tissue inflammation. *Nat Rev Endocrinol* 2017; **13**: 633–43.
114 Arner P, Bernard S, Salehpour M, *et al.* Dynamics of human adipose lipid turnover in health and metabolic disease. *Nature* 2011; **478**: 110–3.
115 Lago F, Dieguez C, Gómez-Reino J, Gualillo O. Adipokines as emerging mediators of immune response and inflammation. *Nat Clin Pract Rheumatol* 2007; **3**: 716–24.
116 Lago F, Gómez R, Gómez-Reino JJ, Dieguez C, Gualillo O. Adipokines as novel modulators of lipid metabolism. Trends Biochem. Sci. 2009; **34**: 500–10.
117 Eguchi K, Nagai R. Islet inflammation in type 2 diabetes and physiology. J. Clin. Invest. 2017; **127**: 14–23.
118 Park EJ, Lee JH, Yu GY, *et al.* Dietary and Genetic Obesity Promote Liver Inflammation and Tumorigenesis by Enhancing IL-6 and TNF Expression. *Cell* 2010; **140**: 197–208.
119 Wu H, Ballantyne CM. Skeletal muscle inflammation and insulin resistance in obesity. *J Clin Invest* 2017; **127**: 43–54.
120 Brestoff JR, Artis D. Immune regulation of metabolic homeostasis in health and disease. Cell. 2015; **161**: 146–60.
121 Virtue S, Vidal-Puig A. Adipose tissue expandability, lipotoxicity and the Metabolic Syndrome - An allostatic perspective. *Biochim Biophys Acta - Mol Cell Biol Lipids* 2010; **1801**: 338–49.
122 Lee YS, Kim JW, Osborne O, *et al.* Increased adipocyte O2 consumption triggers HIF-1α, causing inflammation and insulin resistance in obesity. *Cell* 2014; **157**: 1339–52.
123 Donath MY, Shoelson SE. Type 2 diabetes as an inflammatory disease. *Nat Rev Immunol* 2011; **11**: 98–107.
124 Zmora N, Bashiardes S, Levy M, Elinav E. The Role of the Immune System in Metabolic Health and Disease. *Cell Metab* 2017; **25**: 506–21.
125 Saltiel AR, Olefsky JM. Inflammatory mechanisms linking obesity and metabolic disease. J. Clin. Invest. 2017; **127**: 1–4.
126 Crewe C, An YA, Scherer PE. The ominous triad of adipose tissue dysfunction: Inflammation, fibrosis, and impaired angiogenesis. J. Clin. Invest. 2017; **127**: 74–82.
127 Pasarica M, Rood J, Ravussin E, Schwarz J-M, Smith SR, Redman LM. Reduced Oxygenation in Human Obese Adipose Tissue Is Associated with Impaired Insulin Suppression of Lipolysis. *J Clin Endocrinol Metab* 2010; **95**: 4052–5.
128 Pasarica M, Sereda OR, Redman LM, *et al.* Reduced adipose tissue oxygenation in human obesity evidence for rarefaction, macrophage chemotaxis, and inflammation without an angiogenic response. *Diabetes* 2009; **58**: 718–25.
129 Burke B, Giannoudis A, Corke KP, *et al.* Hypoxia-induced gene expression in human macrophages: Implications for ischemic tissues and hypoxia-regulated gene therapy. *Am J Pathol* 2003; **163**: 1233–43.
130 Hara Y, Wakino S, Tanabe Y, *et al.* Rho and Rho-kinase activity in adipocytes contributes to a vicious cycle in obesity that may involve mechanical stretch. *Sci Signal* 2011; **4**: ra3–ra3.
131 Li Q, Hata A, Kosugi C, Kataoka N, Funaki M. The density of extracellular matrix proteins regulates inflammation and insulin signaling in adipocytes. *FEBS Lett* 2010; **584**: 4145–50.
132 Unger RH. Lipotoxicity in the pathogenesis of obesity-dependent NIDDM: Genetic and clinical implications. Diabetes. 1995; **44**: 863–70.
133 Unger RH. Lipotoxic Diseases. *Annu Rev Med* 2002; **53**: 319–36.
134 Shai I, Rimm EB, Hankinson SE, *et al.* Multivariate assessment of lipid parameters as

predictors of coronary heart disease among postmenopausal women: Potential implications for clinical guidelines. *Circulation* 2004; **110**: 2824–30.
135 Cinti S, Mitchell G, Barbatelli G, *et al.* Adipocyte death defines macrophage localization and function in adipose tissue of obese mice and humans. *J Lipid Res* 2005; **46**: 2347–55.
136 Strissel KJ, Stancheva Z, Miyoshi H, *et al.* Adipocyte death, adipose tissue remodeling, and obesity complications. *Diabetes* 2007; **56**: 2910–8.
137 Weisberg SP, Mccann D, Desai M, Rosenbaum M, Leibel RL, Ferrante AW. Obesity is associated with macrophage accumulation in adipose tissue. *J Clin Invest* 2003; **112**: 1796–808.
138 Xu H, Barnes GTGT, Yang Q, *et al.* Chronic inflammation in fat plays a crucial role in the development of obesity-related insulin resistance. *J Clin Invest* 2003; **112**: 1821–1830.
139 Hotamisligil GS, Shargill NS, Spiegelman BM. Adipose expression of tumor necrosis factor-α: Direct role in obesity-linked insulin resistance. *Science (80-)* 1993; **259**: 87–91.
140 Hotamisligil GS, Arner P, Caro JF, Atkinson RL, Spiegelman BM. Increased adipose tissue expression of tumor necrosis factor-α in human obesity and insulin resistance. *J Clin Invest* 1995; **95**: 2409–15.
141 Kern PA, Saghizadeh M, Ong JM, Bosch RJ, Deem R, Simsolo RB. The expression of tumor necrosis factor in human adipose tissue: Regulation by obesity, weight loss, and relationship to lipoprotein lipase. *J Clin Invest* 1995; **95**: 2111–9.
142 Sartipy P, Loskutoff DJ. Monocyte chemoattractant protein 1 in obesity and insulin resistance. *Proc Natl Acad Sci U S A* 2003; **100**: 7265–70.
143 Kern PA, Ranganathan S, Li C, Wood L, Ranganathan G. Adipose tissue tumor necrosis factor and interleukin-6 expression in human obesity and insulin resistance. *Am J Physiol - Endocrinol Metab* 2001; **280**: 745–51.
144 Han MS, White A, Perry RJ, *et al.* Regulation of adipose tissue inflammation by interleukin 6. *Proc Natl Acad Sci U S A* 2020; **117**: 2751–60.
145 Bruun JM, Pedersen SB, Richelsen B. Regulation of Interleukin 8 Production and Gene Expression in Human Adipose Tissue in Vitro. *J Clin Endocrinol Metab* 2001; **86**: 1267–73.
146 Rocha VZ, Folco EJ, Sukhova G, *et al.* Interferon-γ, a Th1 cytokine, regulates fat inflammation: A role for adaptive immunity in obesity. *Circ Res* 2008; **103**: 467–76.
147 Wentworth JM, Zhang JG, Bandala-Sanchez E, *et al.* Interferon-gamma released from omental adipose tissue of insulin-resistant humans alters adipocyte phenotype and impairs response to insulin and adiponectin release. *Int J Obes* 2017; **41**: 1782–9.
148 Trujillo ME, Sullivan S, Harten I, Schneider SH, Greenberg AS, Fried SK. Interleukin-6 Regulates Human Adipose Tissue Lipid Metabolism and Leptin Production *in Vitro*. *J Clin Endocrinol Metab* 2004; **89**: 5577–82.
149 Boutens L, Stienstra R. Adipose tissue macrophages: going off track during obesity. Diabetologia. 2016; **59**: 879–94.
150 Kamei N, Tobe K, Suzuki R, *et al.* Overexpression of monocyte chemoattractant protein-1 in adipose tissues causes macrophage recruitment and insulin resistance. *J Biol Chem* 2006; **281**: 26602–14.
151 Arner E, Mejhert N, Kulyté A, *et al.* Adipose tissue MicroRNAs as regulators of CCL2 production in human obesity. *Diabetes* 2012; **61**: 1986–93.
152 Lumeng CN, Delproposto JB, Westcott DJ, Saltiel AR. Phenotypic switching of adipose tissue macrophages with obesity is generated by spatiotemporal differences in macrophage subtypes. *Diabetes* 2008; **57**: 3239–46.
153 Kratz M, Coats BR, Hisert KB, *et al.* Metabolic dysfunction drives a mechanistically distinct proinflammatory phenotype in adipose tissue macrophages. *Cell Metab* 2014; **20**: 614–25.
154 Pradhan AD, Manson JE, Rifai N, Buring JE, Ridker PM. C-reactive protein, interleukin 6, and risk of developing type 2 diabetes mellitus. *J Am Med Assoc* 2001; **286**: 327–34.

155 Bertoni AG, Burke GL, Owusu JA, *et al.* Inflammation and the incidence of type 2 diabetes: The Multi-Ethnic Study of Atherosclerosis (MESA). *Diabetes Care* 2010; **33**: 804–10.
156 Bastard J-P, Jardel C, Bruckert E, *et al.* Elevated Levels of Interleukin 6 Are Reduced in Serum and Subcutaneous Adipose Tissue of Obese Women after Weight Loss*. *J Clin Endocrinol Metab* 2000; **85**: 3338–42.
157 Xydakis AM, Case CC, Jones PH, *et al.* Adiponectin, Inflammation, and the Expression of the Metabolic Syndrome in Obese Individuals: The Impact of Rapid Weight Loss through Caloric Restriction. *J Clin Endocrinol Metab* 2004; **89**: 2697–703.
158 Monzillo LU, Hamdy O, Horton ES, *et al.* Effect of lifestyle modification on adipokine levels in obese subjects with insulin resistance. *Obes Res* 2003; **11**: 1048–54.
159 The Diabetes Prevention Program Research Group. Intensive lifestyle intervention or metformin on inflammation and coagulation in participants with impaired glucose tolerance. *Diabetes* 2005; **54**: 1566–72.
160 Sarwar N, Butterworth AS, Freitag DF, *et al.* Interleukin-6 receptor pathways in coronary heart disease: A collaborative meta-analysis of 82 studies. *Lancet* 2012; **379**: 1205–13.
161 Koenig W, Sund M, Fröhlich M, *et al.* C-reactive protein, a sensitive marker of inflammation, predicts future risk of coronary heart disease in initially healthy middle-aged men: Results from the MONICA (monitoring trends and determinants in cardiovascular disease) Augsburg cohort study, 1984. *Circulation* 1999; **99**: 237–42.
162 Libby P. Inflammation in atherosclerosis. *Arterioscler Thromb Vasc Biol* 2012; **32**: 2045–51.
163 Schoenborn JR, Wilson CB. Regulation of Interferon-γ During Innate and Adaptive Immune Responses. Adv. Immunol. 2007; **96**: 41–101.
164 Breland UM, Halvorsen B, Hol J, *et al.* A potential role of the CXC chemokine GROα in atherosclerosis and plaque destabilization: Downregulatory effects of statins. *Arterioscler Thromb Vasc Biol* 2008; **28**: 1005–11.
165 Satterthwaite G, Francis SE, Suvarna K, *et al.* Differential gene expression in coronary arteries from patients presenting with ischemic heart disease: Further evidence for the inflammatory basis of atherosclerosis. *Am Heart J* 2005; **150**: 488–99.
166 Apostolopoulos J, Davenport P, Tipping PG. Interleukin-8 production by macrophages from atheromatous plaques. *Arterioscler Thromb Vasc Biol* 1996; **16**: 1007–12.
167 Mclaughlin T, Ackerman SE, Shen L, Engleman E. Role of innate and adaptive immunity in obesity-associated metabolic disease. J. Clin. Invest. 2017; **127**: 5–13.
168 Spranger J, Kroke A, Möhlig M, *et al.* Inflammatory cytokines and the risk to develop type 2 diabetes: Results of the prospective population-based European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam study. *Diabetes* 2003; **52**: 812–7.
169 Senn JJ, Klover PJ, Nowak IA, Mooney RA. Interleukin-6 induces cellular insulin resistance in hepatocytes. *Diabetes* 2002; **51**: 3391–9.
170 Klover PJ, Zimmers TA, Koniaris LG, Mooney RA. Chronic Exposure to Interleukin-6 Causes Hepatic Insulin Resistance in Mice. *Diabetes* 2003; **52**: 2784–9.
171 Uysal KT, Wiesbrock SM, Marino MW, Hotamisligil GS. Protection from obesity-induced insulin resistance in mice lacking TNF- α function. *Nature* 1997; **389**: 610–4.
172 Pedersen BK, Steensberg A, Schjerling P. Muscle-derived interleukin-6: Possible biological effects. J. Physiol. 2001; **536**: 329–37.
173 Olson NC, Callas PW, Hanley AJG, *et al.* Circulating levels of TNF-α are associated with impaired glucose tolerance, increased insulin resistance, and ethnicity: The insulin resistance atherosclerosis study. *J Clin Endocrinol Metab* 2012; **97**: 1032–40.
174 Herder C, Schneitler S, Rathmann W, *et al.* Low-Grade Inflammation, Obesity, and Insulin Resistance in Adolescents. *J Clin Endocrinol Metab* 2007; **92**: 4569–74.
175 Brahimaj A, Ligthart S, Ghanbari M, *et al.* Novel inflammatory markers for incident pre-diabetes and type 2 diabetes: the Rotterdam Study. *Eur J Epidemiol* 2017; **32**: 217–26.
176 Biggs ML, Mukamal KJ, Luchsinger JA, *et al.* Association between adiposity in midlife and older age and risk of diabetes in older adults. *Obstet Gynecol Surv* 2010; **65**: 708–

10.
177 Yusuf S, Hawken S, Ôunpuu S, *et al.* Obesity and the risk of myocardial infarction in 27 000 participants from 52 countries: A case-control study. *Lancet* 2005; **366**: 1640–9.
178 Stefan N, Häring HU, Hu FB, Schulze MB. Metabolically healthy obesity: Epidemiology, mechanisms, and clinical implications. *Lancet Diabetes Endocrinol* 2013; **1**: 152–62.
179 Langenberg C, Sharp SJ, Schulze MB, *et al.* Long-term risk of incident type 2 diabetes and measures of overall and regional obesity: The epic-interact case-cohort study. *PLoS Med* 2012; **9**: 17.
180 Wallace TM, Levy JC, Matthews DR. Use and Abuse of HOMA Modeling. *Diabetes Care* 2004; **27**: 1487–95.
181 Herder C, Færch K, Carstensen-Kirberg M, *et al.* Biomarkers of subclinical inflammation and increases in glycaemia, insulin resistance and beta-cell function in non-diabetic individuals: The Whitehall II study. *Eur J Endocrinol* 2016; **175**: 367–77.
182 McGillicuddy FC, Chiquoine EH, Hinkle CC, *et al.* Interferon ?? attenuates insulin signaling, lipid storage, and differentiation in human adipocytes via activation of the JAK/STAT pathway. *J Biol Chem* 2009; **284**: 31936–44.
183 Herder C, Baumert J, Thorand B, *et al.* Chemokines as risk factors for type 2 diabetes: Results from the MONICA/KORA Augsburg study, 1984-2002. *Diabetologia* 2006; **49**: 921–9.
184 Herder C, Baumert J, Zierer A, *et al.* Immunological and cardiometabolic risk factors in the prediction of type 2 diabetes and coronary events: MONICA/KORA Augsburg case-cohort study. *PLoS One* 2011; **6**: e19852.
185 Liu C, Feng X, Li Q, Wang Y, Li Q, Hua M. Adiponectin, TNF-α and inflammatory cytokines and risk of type 2 diabetes: A systematic review and meta-analysis. *Cytokine* 2016; **86**: 100–9.
186 Wang X, Bao W, Liu J, *et al.* Inflammatory markers and risk of type 2 diabetes: a systematic review and meta-analysis. *Diabetes Care* 2013; **36**: 166–75.
187 Herder C, Baumert J, Thorand B, *et al.* Chemokines and incident coronary heart disease: Results from the MONICA/KORA Augsburg case-cohort study, 1984-2002. *Arterioscler Thromb Vasc Biol* 2006; **26**: 2147–52.
188 Moreno Velásquez I, Gajulapuri A, Leander K, Berglund A, De Faire U, Gigante B. Serum IL8 is not associated with cardiovascular events but with all-cause mortality. *BMC Cardiovasc Disord* 2019; **19**: 34.
189 Kaptoge S, Seshasai SRK, Gao P, *et al.* Inflammatory cytokines and risk of coronary heart disease: New prospective study and updated meta-analysis. *Eur Heart J* 2014; **35**: 578–89.
190 Clarke R, Valdes-Marquez E, Hill M, *et al.* Plasma cytokines and risk of coronary heart disease in the PROCARDIS study. *Open Hear* 2018; **5**: e000807.
191 Crandall J, Fowler S, Goldberg R, *et al.* Intensive Lifestyle Intervention or Metformin on Inflammation and Coagulation in Participants With Impaired Glucose Tolerance committee consist-ing of Steven Haffner (Chair), Marinella Temprosa. *Diabetes* 2005; **54**. https://diabetes.diabetesjournals.org/content/diabetes/54/5/1566.full.pdf.
192 Rao SR. Inflammatory markers and bariatric surgery: A meta-analysis. *Inflamm Res* 2012; **61**: 789–807.
193 Morin-Papunen L, Rautio K, Ruokonen A, Hedberg P, Puukka M, Tapanainen JS. Metformin Reduces Serum C-Reactive Protein Levels in Women with Polycystic Ovary Syndrome. *J Clin Endocrinol Metab* 2003; **88**: 4649–54.
194 De Jager J, Kooy A, Lehert P, *et al.* Effects of short-term treatment with metformin on markers of endothelial function and inflammatory activity in type 2 diabetes mellitus: A randomized, placebo-controlled trial. *J Intern Med* 2005; **257**: 100–9.
195 Huang NL, Chiang SH, Hsueh CH, Liang YJ, Chen YJ, Lai LP. Metformin inhibits TNF-α-induced IκB kinase phosphorylation, IκB-α degradation and IL-6 production in endothelial cells through PI3K-dependent AMPK phosphorylation. *Int J Cardiol* 2009; **134**: 169–75.
196 Kim J, Kwak HJ, Cha JY, *et al.* Metformin suppresses lipopolysaccharide (LPS)-induced

inflammatory response in murine macrophages via Activating Transcription Factor-3 (ATF-3) induction. *J Biol Chem* 2014; **289**: 23246–55.
197 Kelly B, Tannahill GM, Murphy MP, O'Neill LAJ. Metformin inhibits the production of reactive oxygen species from NADH: Ubiquinone oxidoreductase to limit induction of interleukin-1β (IL-1β) and boosts interleukin-10 (IL-10) in lipopolysaccharide (LPS)-activated macrophages. *J Biol Chem* 2015; **290**: 20348–59.
198 Chaudhuri A, Ghanim H, Vora M, *et al.* Exenatide exerts a potent antiinflammatory effect. *J Clin Endocrinol Metab* 2012; **97**: 198–207.
199 Makdissi A, Ghanim H, Vora M, *et al.* Sitagliptin exerts an antinflammatory action. *J Clin Endocrinol Metab* 2012; **97**: 3333–41.
200 Ussher JR, Drucker DJ. Cardiovascular actions of incretin-based therapies. *Circ Res* 2014; **114**: 1788–803.
201 Dai Y, Dai D, Wang X, Ding Z, Mehta JL. DPP-4 Inhibitors Repress NLRP3 Inflammasome and Interleukin-1beta via GLP-1 Receptor in Macrophages Through Protein Kinase C Pathway. *Cardiovasc Drugs Ther* 2014; **28**: 425–32.
202 Goldfine AB, Fonseca V, Jablonski KA, Pyle L, Staten MA, Shoelson SE. The effects of salsalate on glycemic control in patients with type 2 diabetes: A randomized trial. *Ann Intern Med* 2010; **152**: 346–57.
203 Goldfine AB, Fonseca V, Jablonski KA, *et al.* Salicylate (Salsalate) in patients with type 2 diabetes: A randomized trial. *Ann Intern Med* 2013; **159**: 1–12.
204 De Rotte MCFJ, De Jong PHP, Den Boer E, *et al.* Effect of methotrexate use and erythrocyte methotrexate polyglutamate on glycosylated hemoglobin in rheumatoid arthritis. *Arthritis Rheumatol* 2014; **66**: 2026–36.
205 Everett BM, Pradhan AD, Solomon DH, *et al.* Rationale and design of the Cardiovascular Inflammation Reduction Trial: A test of the inflammatory hypothesis of atherothrombosis. *Am Heart J* 2013; **166**: 199-207.e15.
206 Kiortsis DN, Mavridis AK, Vasakos S, Nikas SN, Drosos AA. Effects of infliximab treatment on insulin resistance in patients with rheumatoid arthritis and ankylosing spondylitis. *Ann Rheum Dis* 2005; **64**: 765–6.
207 Marra M, Campanati A, Testa R, *et al.* Effect of Etanercept on insulin sensitivity in nine patients with psoriasis. *Int J Immunopathol Pharmacol* 2007; **20**: 731–6.
208 Bernstein LE, Berry J, Kim S, Canavan B, Grinspoon SK. Effects of etanercept in patients with the metabolic syndrome. *Arch Intern Med* 2006; **166**: 902–8.
209 Dominguez H, Storgaard H, Rask-Madsen C, *et al.* Metabolic and Vascular Effects of Tumor Necrosis Factor-α Blockade with Etanercept in Obese Patients with Type 2 Diabetes. *J Vasc Res* 2005; **42**: 517–25.
210 Paquot N, Castillo MJ, Lefèbvre PJ, Scheen AJ. No Increased Insulin Sensitivity after a Single Intravenous Administration of a Recombinant Human Tumor Necrosis Factor Receptor: Fc Fusion Protein in Obese Insulin-Resistant Patients [1]. *J Clin Endocrinol Metab* 2000; **85**: 1316–9.
211 Stanley TL, Zanni M V., Johnsen S, *et al.* TNF-α antagonism with etanercept decreases glucose and increases the proportion of high molecular weight adiponectin in obese subjects with features of the metabolic syndrome. *J Clin Endocrinol Metab* 2011; **96**: E146–50.
212 Genovese MC, Fleischmann R, Hagino O, *et al.* The Effect of Sarilumab in Combination with Dmards on Fasting Glucose and Glycosylated Hemoglobin in Patients with Rheumatoid Arthritis with and without Diabetes. In: 2017 American College of Rheumatology Annual Meeting. 2017: 69.
213 Joseph JP, Reyes E, Guzman J, *et al.* CXCR2 Inhibition - a novel approach to treating CoronAry heart DiseAse (CICADA): Study protocol for a randomised controlled trial. *Trials* 2017; **18**: 473.
214 Genovese MC, Fleischmann R, Kivitz AJ, *et al.* Sarilumab plus methotrexate in patients with active rheumatoid arthritis and inadequate response to methotrexate: Results of a phase III study. *Arthritis Rheumatol* 2015; **67**: 1424–37.
215 Fleischmann R, van Adelsberg J, Lin Y, *et al.* Sarilumab and Nonbiologic Disease-

Modifying Antirheumatic Drugs in Patients With Active Rheumatoid Arthritis and Inadequate Response or Intolerance to Tumor Necrosis Factor Inhibitors. *Arthritis Rheumatol* 2017; **69**: 277–90.
216 Dinarello CA, van der Meer JWM. Treating inflammation by blocking interleukin-1 in humans. *Semin Immunol* 2013; **25**: 469–84.
217 Weber A, Wasiliew P, Kracht M. Interleukin-1 (IL-1) pathway. Sci. Signal. 2010; **3**: cm1–cm1.
218 Larsen CM, Faulenbach M, Vaag A, *et al.* Interleukin-1–Receptor Antagonist in Type 2 Diabetes Mellitus. *N Engl J Med* 2007; **356**: 1517–26.
219 Larsen CM, Faulenbach M, Vaag A, Ehses JA, Donath MY, Mandrup-Poulsen T. Sustained effects of interleukin-1 receptor antagonist treatment in type 2 diabetes. *Diabetes Care* 2009; **32**: 1663–8.
220 Ridker PM, Everett BM, Thuren T, *et al.* Antiinflammatory Therapy with Canakinumab for Atherosclerotic Disease. *N Engl J Med* 2017; : NEJMoa1707914.
221 Everett BM, Donath MY, Pradhan AD, *et al.* Anti-Inflammatory Therapy with Canakinumab for the Prevention and Management of Diabetes. *J Am Coll Cardiol* 2018; **71**. DOI:10.1016/j.jacc.2018.03.002.
222 Burgess S, Ference BA, Staley JR, *et al.* Association of LPA Variants With Risk of Coronary Disease and the Implications for Lipoprotein(a)-Lowering Therapies: A Mendelian Randomization Analysis. *JAMA Cardiol* 2018. DOI:10.1001/JAMACARDIO.2018.1470.
223 Kahn SE, Hull RL, Utzschneider KM. Mechanisms linking obesity to insulin resistance and type 2 diabetes. *Nature* 2006; **444**: 840–6.
224 Cimini FA, Barchetta I, Porzia A, *et al.* Circulating IL-8 levels are increased in patients with type 2 diabetes and associated with worse inflammatory and cardiometabolic profile. *Acta Diabetol* 2017; **54**: 961–7.
225 Marques-Vidal P, Schmid R, Bochud M, *et al.* Adipocytokines, Hepatic and Inflammatory Biomarkers and Incidence of Type 2 Diabetes. The CoLaus Study. *PLoS One* 2012; **7**: e51768.
226 Ridker PM, Everett BM, Thuren T, *et al.* Antiinflammatory Therapy with Canakinumab for Atherosclerotic Disease. *N Engl J Med* 2017; : NEJMoa1707914.
227 Ahola-Olli A V., Würtz P, Havulinna AS, *et al.* Genome-wide Association Study Identifies 27 Loci Influencing Concentrations of Circulating Cytokines and Growth Factors. *Am J Hum Genet* 2017; **100**: 40–50.
228 Burton PR, Clayton DG, Cardon LR, *et al.* Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007; **447**: 661–78.
229 Sliz E, Kalaoja M, Ahola-Olli A, *et al.* Genome-wide association study identifies seven novel loci associating with circulating cytokines and cell adhesion molecules in Finns. *J Med Genet* 2019; **56**: 607–16.
230 Sun BB, Maranville JC, Peters JE, *et al.* Genomic atlas of the human plasma proteome. *Nature* 2018; **558**: 73–9.
231 Suhre K, Arnold M, Bhagwat AM, *et al.* Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat Commun* 2017; **8**: 14357.
232 Sidore C, Busonero F, Maschio A, *et al.* Genome sequencing elucidates Sardinian genetic architecture and augments association analyses for lipid and blood inflammatory markers. *Nat Genet* 2015; **47**: 1272–81.
233 Ayele FT, Doumatey A, Huang H, *et al.* Genome-wide associated loci influencing interleukin (IL)-10, IL-1Ra, and IL-6 levels in African Americans. *Immunogenetics* 2012; **64**: 351–9.
234 Naitza S, Porcu E, Steri M, *et al.* A genome-wide association scan on the levels of markers of inflammation in sardinians reveals associations that underpin its complex regulation. *PLoS Genet* 2012; **8**. DOI:10.1371/journal.pgen.1002480.
235 Melzer D, Perry JRB, Hernandez D, *et al.* A Genome-Wide Association Study Identifies Protein Quantitative Trait Loci (pQTLs). *PLoS Genet* 2008; **4**: e1000072.

236 Gibson G. Rare and common variants: Twenty arguments. *Nat Rev Genet* 2012; **13**: 135–45.
237 Hanada T, Yoshimura A. Regulation of cytokine signaling and inflammation. *Cytokine Growth Factor Rev* 2002; **13**: 413–21.
238 Duncan BB, Schmidt MI, Pankow JS, *et al.* Low-Grade Systemic Inflammation and the Development of Type 2 Diabetes: The Atherosclerosis Risk in Communities Study. *Diabetes* 2003; **52**: 1799–805.
239 Koloverou E, Panagiotakos DB, Georgousopoulou EN, *et al.* Single and combined effects of inflammatory markers on 10 year diabetes incidence: The mediating role of adiposity—Results from the ATTICA cohort study. *Diabetes Metab Res Rev* 2018; **34**: e2939.
240 Dallmeier D, Larson MG, Wang N, Fontes JD, Benjamin EJ, Fox CS. Addition of Inflammatory Biomarkers Did Not Improve Diabetes Prediction in the Community: The Framingham Heart Study. *J Am Heart Assoc* 2012; **1**: e000869–e000869.
241 Day N, Oakes S, Luben R, *et al.* EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer. *Br J Cancer* 1999; **80 Suppl 1**: 95–103.
242 Riboli E, Hunt K, Slimani N, *et al.* European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr* 2002; **5**: 1113.
243 Hayat SA, Luben R, Keevil VL, *et al.* Cohort profile: A prospective cohort study of objective physical and cognitive capability and visual health in an ageing population of men and women in Norfolk (EPIC-Norfolk 3). *Int J Epidemiol* 2014; **43**: 1063–72.
244 Lotta LA, Gulati P, Day FR, *et al.* Integrative genomic analysis implicates limited peripheral adipose storage capacity in the pathogenesis of human insulin resistance. *Nat Genet* 2016; **49**: 17–26.
245 Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin Chem* 1972; **18**: 499–502.
246 Titan SM, Bingham S, Welch A, *et al.* Frequency of eating and concentrations of serum cholesterol in the Norfolk population of the European prospective investigation into cancer (EPIC-Norfolk): cross sectional study. *BMJ* 2001; **323**: 1286–8.
247 enCORE Software Platform. 2017.
248 Langenberg C, Sharp SJ, Franks PW, *et al.* Gene-Lifestyle Interaction and Type 2 Diabetes: The EPIC InterAct Case-Cohort Study. *PLoS Med* 2014; **11**. DOI:10.1371/journal.pmed.1001647.
249 Lindsay T, Westgate K, Wijndaele K, *et al.* Descriptive epidemiology of physical activity energy expenditure in UK adults (The Fenland study). *Int J Behav Nutr Phys Act* 2019; **16**: 1–13.
250 Kang YE, Kim JM, Joung KH, *et al.* The roles of adipokines, proinflammatory cytokines, and adipose tissue macrophages in obesity-associated insulin resistance in modest obesity and early metabolic dysfunction. *PLoS One* 2016; **11**: 1–14.
251 Makki K, Froguel P, Wolowczuk I. Adipose Tissue in Obesity-Related Inflammation and Insulin Resistance: Cells, Cytokines, and Chemokines. *ISRN Inflamm* 2013; **2013**: 1–12.
252 Supriya R, Tam BT, Yu AP, *et al.* Adipokines demonstrate the interacting influence of central obesity with other cardiometabolic risk factors of metabolic syndrome in Hong Kong Chinese adults. *PLoS One* 2018; **13**: e0201585.
253 Caër C, Rouault C, Le Roy T, *et al.* Immune cell-derived cytokines contribute to obesity-related inflammation, fibrogenesis and metabolic deregulation in human adipose tissue. *Sci Rep* 2017; **7**: 3000.
254 Skurk T, Alberti-Huber C, Herder C, Hauner H. Relationship between adipocyte size and adipokine expression and secretion. *J Clin Endocrinol Metab* 2007; **92**: 1023–33.
255 Marques-Vidal P, Bastardot F, Von Känel R, *et al.* Association between circulating cytokine levels, diabetes and insulin resistance in a population-based sample (CoLaus

study). *Clin Endocrinol (Oxf)* 2013; **78**: 232–41.
256 Fernandez-Real JM, Vayreda M, Richart C, *et al.* Circulating interleukin 6 levels, blood pressure, and insulin sensitivity in apparently healthy men and women. *J Clin Endocrinol Metab* 2001; **86**: 1154–9.
257 Brown WV. From the editor: Cytokines and HDL cholesterol. *J Clin Lipidol* 2015; **9**: 1.
258 Zuliani G, Volpato S, Blè A, *et al.* High interleukin-6 plasma levels are associated with low HDL-C levels in community-dwelling older adults: The InChianti study. *Atherosclerosis* 2007; **192**: 384–90.
259 Ebtehaj S, Gruppen EG, Parvizi M, Tietge UJF, Dullaart RPF. The anti-inflammatory function of HDL is impaired in type 2 diabetes: Role of hyperglycemia, paraoxonase-1 and low grade inflammation. *Cardiovasc Diabetol* 2017; **16**: 132.
260 Lehto S, Rönnemaa T, Haffner SM, Pyörälä K, Kallio V, Laakso M. Dyslipidemia and hyperglycemia predict coronary heart disease events in middle-aged patients with NIDDM. *Diabetes* 1997; **46**: 1354–9.
261 The Emerging Risk Factors Collaboration* TERF. Major Lipids, Apolipoproteins, and Risk of Vascular Disease. *JAMA* 2009; **302**: 1993.
262 MacMahon S, Duffy S, Rodgers A, *et al.* Blood cholesterol and vascular mortality by age, sex, and blood pressure: A meta-analysis of individual data from 61 prospective studies with 55 000 vascular deaths. *Lancet* 2007; **370**: 1829–39.
263 van der Vorst EPC, Theodorou K, Wu Y, *et al.* High-Density Lipoproteins Exert Pro-inflammatory Effects on Macrophages via Passive Cholesterol Depletion and PKC-NF-κB/STAT1-IRF1 Signaling. *Cell Metab* 2017; **25**: 197–207.
264 Straczkowski M, Dzienis-Straczkowska S, St??pie?? A, Kowalska I, Szelachowska M, Kinalska I. Plasma interleukin-8 concentrations are increased in obese subjects and related to fat mass and tumor necrosis factor-?? system. *J Clin Endocrinol Metab* 2002; **87**: 4602–6.
265 Kim CS, Park HS, Kawada T, *et al.* Circulating levels of MCP-1 and IL-8 are elevated in human obese subjects and associated with obesity-related parameters. *Int J Obes* 2006; **30**: 1347–55.
266 Schmidt FM, Weschenfelder J, Sander C, *et al.* Inflammatory cytokines in general and central obesity and modulating effects of physical Activity. *PLoS One* 2015; **10**: 1–17.
267 Park HS, Park JY, Yu R. Relationship of obesity and visceral adiposity with serum concentrations of CRP, TNF-α and IL-6. *Diabetes Res Clin Pract* 2005; **69**: 29–35.
268 Abbasi A, Sahlqvist AS, Lotta L, *et al.* A systematic review of biomarkers and risk of incident type 2 diabetes: An overview of epidemiological, prediction and aetiological research literature. *PLoS One* 2016; **11**. DOI:10.1371/journal.pone.0163721.
269 Thanassoulis G, O'Donnell CJ. Mendelian randomization: Nature's randomized trial in the post-genome era. *JAMA - J Am Med Assoc* 2009; **301**: 2386–8.
270 Shoelson SE, Lee J, Goldfine AB. Inflammation and insulin resistance. J. Clin. Invest. 2006; **116**: 1793–801.
271 McCarthy S, Das S, Kretzschmar W, *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* 2016; **48**: 1279–83.
272 Auton A, Abecasis GR, Altshuler DM, *et al.* A global reference for human genetic variation. Nature. 2015; **526**: 68–74.
273 Walter K, Min JL, Huang J, *et al.* The UK10K project identifies rare variants in health and disease. *Nature* 2015; **526**: 82–9.
274 Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007; **39**: 906–13.
275 Howie BN, Donnelly P, Marchini J. A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. *PLoS Genet* 2009; **5**: e1000529.
276 Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinforma Appl NOTE* 2010; **26**: 2190–219110.
277 Winkler TW, Kutalik Z, Gorski M, Lottaz C, Kronenberg F, Heid IM. EasyStrata:

evaluation and visualization of stratified genome-wide association meta-analysis data. *Bioinformatics* 2015; **31**: 259–61.
278 Pruim RJ, Welch RP, Sanna S, *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinforma Appl NOTE* 2010; **26**: 2336–233710.
279 Scott RA, Scott LJ, Mägi R, *et al.* An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. *Diabetes* 2017; **66**: 2888–902.
280 Gaulton KJ, Ferreira T, Lee Y, *et al.* Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat Genet* 2015; **47**: 1415–25.
281 Yang J, Ferreira T, Morris AP, *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* 2012; **44**: 369–75.
282 Welter D, MacArthur J, Morales J, *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* 2014; **42**: D1001–6.
283 van de Bunt M, Cortes A, IGAS Consortium I, Brown MA, Morris AP, McCarthy MI. Evaluating the Performance of Fine-Mapping Strategies at Common Variant GWAS Loci. *PLoS Genet* 2015; **11**: e1005535.
284 Giambartolomei C, Vukcevic D, Schadt EE, *et al.* Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genet* 2014; **10**: e1004383.
285 McLaren W, Gil L, Hunt SE, *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* 2016; **17**: 122.
286 Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, 2016 https://ggplot2.tidyverse.org/.
287 Bulik-Sullivan BK, Loh P-R, Finucane HK, *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 2015; **47**: 291–5.
288 Zheng J, Erzurumluoglu AM, Elsworth BL, *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* 2017; **33**: 272–9.
289 van der Harst P, Verweij N. Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease. *Circ Res* 2018; **122**: 433–43.
290 Giambartolomei C, Vukcevic D, Schadt EE, *et al.* Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genet* 2014; **10**: e1004383.
291 Bulik-Sullivan B, Finucane HK, Anttila V, *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* 2015; **47**: 1236–41.
292 Dufner A, Kisser A, Niendorf S, *et al.* The ubiquitin-specific protease USP8 is critical for the development and homeostasis of T cells. *Nat Immunol* 2015; **16**: 950–60.
293 Kalliolias GD, Ivashkiv LB. TNF biology, pathogenic mechanisms and emerging therapeutic strategies. *Nat Rev Rheumatol* 2016; **12**: 49–62.
294 Hasham SN, Pillarisetti S. Vascular lipases, inflammation and atherosclerosis. *Clin Chim Acta* 2006; **372**: 179–83.
295 Lettre G, Palmer CD, Young T, *et al.* Genome-Wide Association Study of Coronary Heart Disease and Its Risk Factors in 8,090 African Americans: The NHLBI CARe Project. *PLoS Genet* 2011; **7**: e1001300.
296 Kristiansson K, Perola M, Tikkanen E, *et al.* Genome-wide screen for metabolic syndrome susceptibility loci reveals strong lipid gene contribution but no evidence for common genetic basis for clustering of metabolic syndrome traits. *Circ Cardiovasc Genet* 2012; **5**: 242–9.
297 YS A, Ripatti S, Lindqvist I, *et al.* Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat Genet* 2009; **41**: 47–55.
298 Chen X, Xun K, Chen L, Wang Y. TNF- *α* , a potent lipid metabolism regulator. *Cell*

Biochem Funct 2009; **27**: 407–16.
299 Tacer KF, Kuzman D, Seliškar M, Pompon D, Rozman D. TNF-α interferes with lipid homeostasis and activates acute and proatherogenic processes. *Physiol Genomics* 2007; **31**: 216–27.
300 Moreno Velásquez I, Kumar J, Björkbacka H, *et al.* Duffy antigen receptor genetic variant and the association with Interleukin 8 levels. *Cytokine* 2015; **72**: 178–84.
301 Schnabel RB, Baumert J, Barbalic M, *et al.* Duffy antigen receptor for chemokines (Darc) polymorphism regulates circulating concentrations of monocyte chemoattractant protein-1 and other inflammatory mediators. *Blood* 2010; **115**: 5289–99.
302 Horuk R. The Duffy Antigen Receptor for Chemokines DARC/ACKR1. *Front Immunol* 2015; **6**: 279.
303 Reich D, Nalls MA, Kao WHL, *et al.* Reduced Neutrophil Count in People of African Descent Is Due To a Regulatory Variant in the Duffy Antigen Receptor for Chemokines Gene. *PLoS Genet* 2009; **5**: e1000360.
304 Rot A. Contribution of Duffy antigen to chemokine function. *Cytokine Growth Factor Rev* 2005; **16**: 687–94.
305 Velásquez IM, Frumento P, Johansson K, *et al.* Association of interleukin 8 with myocardial infarction: Results from the Stockholm Heart Epidemiology Program. *Int J Cardiol* 2014; **172**: 173–8.
306 Apostolakis S, Vogiatzi K, Amanatidou V, Spandidos DA. Interleukin 8 and cardiovascular disease. Cardiovasc. Res. 2009; **84**: 353–60.
307 de Winter RJ, Manten A, de Jong YP, Adams R, van Deventer SJ, Lie KI. Interleukin 8 released after acute myocardial infarction is mainly bound to erythrocytes. *Heart* 1997; **78**: 598–602.
308 Dunham I, Kundaje A, Aldred SF, *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012; **489**: 57–74.
309 Locke AE, Kahali B, Berndt SI, *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* 2015; **518**: 197–206.
310 Ferreira RC, Freitag DF, Cutler AJ, *et al.* Functional IL6R 358Ala Allele Impairs Classical IL-6 Receptor Signaling and Influences Risk of Diverse Inflammatory Diseases. *PLoS Genet* 2013; **9**: e1003444.
311 Russo RC, Garcia CC, Teixeira MM, Amaral FA. The CXCL8/IL-8 chemokine family and its receptors in inflammatory diseases. Expert Rev. Clin. Immunol. 2014; **10**: 593–619.
312 Võsa U, Claringbould A, Westra H-J, *et al.* Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. *bioRxiv* 2018; : 447367.
313 Auer PL, Teumer A, Schick U, *et al.* Rare and low-frequency coding variants in CXCR2 and other genes are associated with hematological traits. *Nat Genet* 2014; **46**: 629–34.
314 Astle WJ, Elding H, Jiang T, *et al.* The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* 2016; **167**: 1415-1429.e19.
315 Auburger G. 12q24 locus association with type 1 diabetes: SH2B3 or ATXN2 ? . *World J Diabetes* 2014; **5**: 316.
316 Fitau J, Boulday G, Coulon F, Quillard T, Charreau B. The adaptor molecule Lnk negatively regulates tumor necrosis factor-α-dependent VCAM-1 expression in endothelial cells through inhibition of the ERK1 and -2 pathways. *J Biol Chem* 2006; **281**: 20148–59.
317 Mori T, Suzuki-Yamazaki N, Takaki S. Lnk/Sh2b3 Regulates Adipose Inflammation and Glucose Tolerance through Group 1 ILCs. *Cell Rep* 2018; **24**: 1830–41.
318 Devallire J, Charreau B. The adaptor Lnk (SH2B3): An emerging regulator in vascular cells and a link between immune and inflammatory signaling. In: Biochemical Pharmacology. Elsevier, 2011: 1391–402.
319 Fadista J, Oskolkov N, Hansson O, Groop L. LoFtool: a gene intolerance score based on loss-of-function variants in 60 706 individuals. *Bioinformatics* 2016; **33**: btv602.
320 Hunt KA, Zhernakova A, Turner G, *et al.* Newly identified genetic risk variants for celiac disease related to the immune response. *Nat Genet* 2008; **40**: 395–402.

321 Trynka G, Hunt KA, Bockett NA, *et al.* Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat Genet* 2011; **43**: 1193–201.
322 Ellinghaus D, Jostins L, Spain SL, *et al.* Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nat Genet* 2016; **48**: 510–8.
323 Liu JZ, van Sommeren S, Huang H, *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet* 2015; **47**: 979–86.
324 Nikpay M, Goel A, Won HH, *et al.* A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet* 2015; **47**: 1121–30.
325 Todd JA, Walker NM, Cooper JD, *et al.* Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat Genet* 2007; **39**: 857–64.
326 Onengut-Gumuscu S, Chen W-M, Burren O, *et al.* Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat Genet* 2015; **47**: 381–6.
327 Saleh MA, McMaster WG, Wu J, *et al.* Lymphocyte adaptor protein LNK deficiency exacerbates hypertension and end-organ inflammation. *J Clin Invest* 2015; **125**: 1189–202.
328 Benner C, Spencer CCA, Havulinna AS, Salomaa V, Ripatti S, Pirinen M. FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* 2016; **32**: 1493–501.
329 Kichaev G, Yang WY, Lindstrom S, *et al.* Integrating Functional Data to Prioritize Causal Variants in Statistical Fine-Mapping Studies. *PLoS Genet* 2014; **10**: e1004722.
330 Prins BP, Kuchenbaecker KB, Bao Y, *et al.* Genome-wide analysis of health-related biomarkers in the UK Household Longitudinal Study reveals novel associations. *Sci Rep* 2017; **7**: 1–9.
331 Sesso HD, Buring JE, Rifai N, Blake GJ, Gaziano JM, Ridker PM. C-Reactive Protein and the Risk of Developing Hypertension. *J Am Med Assoc* 2003; **290**: 2945–51.
332 Wang TJ, Gona P, Larson MG, *et al.* Multiple biomarkers and the risk of incident hypertension. *Hypertension* 2007; **49**: 432–8.
333 Harrison DG. Inflammation, Immunity, and Hypertension. *Hypertension* 2017; **49**: 158–65.
334 Ehses JA, Ellingsgaard H, Böni-Schnetzler M, Donath MY. Pancreatic islet inflammation in type 2 diabetes: From α and β cell compensation to dysfunction. *Arch Physiol Biochem* 2009; **115**: 240–7.
335 Donath MY. Multiple benefits of targeting inflammation in the treatment of type 2 diabetes. *Diabetologia* 2016; **59**: 679–82.
336 Ridker PM, Thuren T, Zalewski A, Libby P. Interleukin-1β inhibition and the prevention of recurrent cardiovascular events: Rationale and Design of the Canakinumab Anti-inflammatory Thrombosis Outcomes Study (CANTOS). *Am Heart J* 2011; **162**: 597–605.
337 Cai T, Zhang Y, Ho Y-L, *et al.* Association of Interleukin 6 Receptor Variant With Cardiovascular Disease Effects of Interleukin 6 Receptor Blocking Therapy. *JAMA Cardiol* 2018; published online Aug 8. DOI:10.1001/jamacardio.2018.2287.
338 Finngen. FinnGen Documentation of R2 release. 2020. https://finngen.gitbook.io/finngen-documentation/-LvQ4yR2YFUM5eFTjieO/ (accessed March 31, 2020).
339 Carey DJ, Fetterolf SN, Davis FD, *et al.* The Geisinger MyCode community health initiative: An electronic health record-linked biobank for precision medicine research. *Genet Med* 2016; **18**: 906–13.
340 Suzuki K, Akiyama M, Ishigaki K, *et al.* Identification of 28 new susceptibility loci for type 2 diabetes in the Japanese population. *Nat Genet* 2019; : 1.
341 Gaziano JM, Concato J, Brophy M, *et al.* Million Veteran Program: A mega-biobank to

study genetic influences on health and disease. *J Clin Epidemiol* 2016; **70**: 214–23.
342 Dewey FE, Murray MF, Overton JD, *et al.* Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. *Science (80-)* 2016; **354**. DOI:10.1126/science.aaf6814.
343 Locke AE, Kahali B, Berndt SI, *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* 2015; **518**: 197–206.
344 Saxena R, Hivert MF, Langenberg C, *et al.* Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nat Genet* 2010; **42**: 142–8.
345 Scott RA, Lagou V, Welch RP, *et al.* Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat Genet* 2012; **44**: 991–1005.
346 Manning AK, Hivert M-F, Scott RA, *et al.* A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat Genet* 2012; **44**: 659–69.
347 Wheeler E, Leong A, Liu CT, *et al.* Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. *PLoS Med* 2017; **14**: e1002383.
348 Garbers C, Monhasery N, Aparicio-Siegmund S, *et al.* The interleukin-6 receptor Asp358Ala single nucleotide polymorphism rs2228145 confers increased proteolytic conversion rates by ADAM proteases. *Biochim Biophys Acta - Mol Basis Dis* 2014; **1842**: 1485–94.
349 Altshuler DL, Durbin RM, Abecasis GR, *et al.* A map of human genome variation from population-scale sequencing. *Nature* 2010; **467**: 1061–73.
350 Thomas NJ, Jones SE, Weedon MN, Shields BM, Oram RA, Hattersley AT. Frequency and phenotype of type 1 diabetes in the first six decades of life: A cross-sectional, genetically stratified survival analysis from UK Biobank. *Lancet Diabetes Endocrinol* 2017; **6**: 122–9.
351 Scott RA, Fall T, Pasko D, *et al.* Common genetic variants highlight the role of insulin resistance and body fat distribution in type 2 diabetes, independent of obesity. *Diabetes* 2014; **63**: 4378–87.
352 Lotta LA, Mokrosiński J, Mendes de Oliveira E, *et al.* Human Gain-of-Function MC4R Variants Show Signaling Bias and Protect against Obesity. *Cell* 2019; **177**: 597-607.e9.
353 Ference BA, Kastelein JJP, Ray KK, *et al.* Association of Triglyceride-Lowering LPL Variants and LDL-C-Lowering LDLR Variants with Risk of Coronary Heart Disease. *JAMA - J Am Med Assoc* 2019; **321**: 364–73.
354 Baron RM, Kenny DA. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *J Pers Soc Psychol* 1986; **51**: 1173–82.
355 Loh P-R, Kichaev G, Gazal S, Schoech AP, Price AL. Mixed-model association for biobank-scale datasets. *Nat Genet* 2018; **50**: 906–8.
356 Chawla A, Nguyen KD, Goh YPS. Macrophage-mediated inflammation in metabolic disease. Nat. Rev. Immunol. 2011; **11**: 738–49.
357 DeFronzo RA, Tripathy D, Schwenke DC, *et al.* Pioglitazone for Diabetes Prevention in Impaired Glucose Tolerance. *N Engl J Med* 2011; **364**: 1104–15.
358 Qu D, Liu J, Lau CW, Huang Y. IL-6 in diabetes and cardiovascular complications. *Br J Pharmacol* 2014; **171**: 3595–603.
359 Otsuka Y, Kiyohara C, Kashiwado Y, *et al.* Effects of tumor necrosis factor inhibitors and tocilizumab on the glycosylated hemoglobin levels in patients with rheumatoid arthritis; an observational study. *PLoS One* 2018; **13**: 1–16.
360 Dungan KM, Braithwaite SS, Preiser JC. Stress hyperglycaemia. *Lancet* 2009; **373**: 1798–807.
361 Nakamura M, Oda S, Sadahiro T, *et al.* Correlation between high blood IL-6 level, hyperglycemia, and glucose control in septic patients. *Crit Care* 2012; **16**: R58.

362 Foley CN, Staley JR, Breen PG, *et al.* A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. *bioRxiv* 2019; **44**: 592238.
363 Giambartolomei C, Liu JZ, Zhang W, *et al.* A Bayesian framework for multiple trait colocalization from summary association statistics. *Bioinformatics* 2018; **34**: 2538–45.
364 Williams SA, Kivimaki M, Langenberg C, *et al.* Plasma protein patterns as comprehensive indicators of health. *Nat Med* 2019; **25**: 1851–7.
365 Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011; **88**: 76–82.
366 Ashburner M, Ball CA, Blake JA, *et al.* Gene Ontology : tool for the unification of biology. *Nat Genet* 2000; **25**: 25–9.
367 Carbon S, Douglass E, Dunn N, *et al.* The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res* 2019; **47**: D330–8.
368 Carbon S, Ireland A, Mungall CJ, *et al.* AmiGO: Online access to ontology and annotation data. *Bioinformatics* 2009; **25**: 288–9.
369 Wallace C. Eliciting priors and relaxing the single causal variant assumption in colocalisation analyses. *PLoS Genet* 2020; **16**: e1008720.
370 Kraja AT, Chasman DI, North KE, *et al.* Pleiotropic genes for metabolic syndrome and inflammation. *Mol Genet Metab* 2014; **112**: 317–38.
371 Fernandes Silva L, Vangipurapu J, Kuulasmaa T, Laakso M. An intronic variant in the GCKR gene is associated with multiple lipids. *Sci Rep* 2019; **9**: 1–9.
372 Webb TR, Erdmann J, Stirrups KE, *et al.* Systematic Evaluation of Pleiotropy Identifies 6 Further Loci Associated With Coronary Artery Disease. *J Am Coll Cardiol* 2017; **69**: 823–36.
373 Chen VL, Chen Y, Du X, Handelman SK, Speliotes EK. Genetic variants that associate with cirrhosis have pleiotropic effects on human traits. *Liver Int* 2020; **40**: 405–15.
374 Watanabe K, Taskesen E, Van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* 2017; **8**: 1826.
375 Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 2011; **27**: 1739–40.
376 de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput Biol* 2015; **11**: 1–19.
377 Kaur Y, Wang DX, Liu HY, Meyre D. Comprehensive identification of pleiotropic loci for body fat distribution using the NHGRI-EBI Catalog of published genome-wide association studies. *Obes Rev* 2019; **20**: 385–406.
378 Lumsden AL, Mulugeta A, Zhou A, Hyppönen E. Apolipoprotein E (APOE) genotype-associated disease risks: a phenome-wide, registry-based, case-control study utilising the UK Biobank. *EBioMedicine* 2020; **59**. DOI:10.1016/j.ebiom.2020.102954.
379 Li P, Liu S, Lu M, *et al.* Hematopoietic-Derived Galectin-3 Causes Cellular and Systemic Insulin Resistance. *Cell* 2016; **167**: 973-984.e12.
380 Kingwell K. Diabetes: Turning down galectin 3 to combat insulin resistance. *Nat Rev Drug Discov* 2016; **16**: 18.
381 Bresalier RS, Byrd JC, Tessler D, *et al.* A circulating ligand for galectin-3 is a haptoglobin-related glycoprotein elevated in individuals with colon cancer. *Gastroenterology* 2004; **127**: 741–8.
382 Gordon T, Castelli WP, Hjortland MC, Kannel WB, Dawber TR. High density lipoprotein as a protective factor against coronary heart disease. The Framingham study. *Am J Med* 1977; **62**: 707–14.
383 Kaptoge S, Di Angelantonio E, Lowe G, *et al.* C-reactive protein concentration and risk of coronary heart disease, stroke, and mortality: An individual participant meta-analysis. *Lancet* 2010; **375**: 132–40.
384 Sack Jr GH. Serum amyloid A – a review. *Mol Med* 2018; **24**: 1–27.
385 Marhaug G, Dowton SB. Serum amyloid A: An acute phase apolipoprotein and precursor of AA amyloid. *Baillieres Clin Rheumatol* 1994; **8**: 553–73.
386 Tall AR, Yvan-Charvet L. Cholesterol, inflammation and innate immunity. *Nat Rev Immunol* 2015; **15**: 104–16.

387 Ebtehaj S, Gruppen EG, Bakker SJL, Dullaart RPF, Tietge UJF. HDL (High-Density Lipoprotein) Cholesterol Efflux Capacity Is Associated With Incident Cardiovascular Disease in the General Population. *Arterioscler Thromb Vasc Biol* 2019; **39**: 1874–83.
388 Yaghootkar H, Lotta LA, Tyrrell J, *et al.* Genetic evidence for a link between favorable adiposity and lower risk of type 2 diabetes, hypertension, and heart disease. *Diabetes* 2016; **65**: 2448–60.
389 Yaghootkar H, Scott RA, White CC, *et al.* Genetic evidence for a normal-weight 'metabolically obese' phenotype linking insulin resistance, hypertension, coronary artery disease, and type 2 diabetes. *Diabetes* 2014; **63**: 4369–77.
390 Salazar MR, Carbajal HA, Espeche WG, *et al.* Comparison of the abilities of the plasma triglyceride/high-density lipoprotein cholesterol ratio and the metabolic syndrome to identify insulin resistance. *Diabetes Vasc Dis Res* 2013; **10**: 346–52.
391 Shim K, Begum R, Yang C, Wang H. Complement activation in obesity, insulin resistance, and type 2 diabetes mellitus. *World J Diabetes* 2020; **11**: 1–12.
392 Engström G, Hedblad bo, Janzon L, Lindgärde F. Complement C3 and C4 in plasma and incidence of myocardial infarction and stroke: A population-based cohort study. *Eur J Prev Cardiol* 2007; **14**: 392–7.
393 Palikhe A, Sinisalo J, Seppänen M, *et al.* Serum Complement C3/C4 Ratio, a Novel Marker for Recurrent Cardiovascular Events. *Am J Cardiol* 2007; **99**: 890–5.
394 Onat A, Hergenç G, Can G, Kaya Z, Yüksel H. Serum complement C3: a determinant of cardiometabolic risk, additive to the metabolic syndrome, in middle-aged population. *Metabolism* 2010; **59**: 628–34.
395 Van Der Laan SW, Harshfield EL, Hemerich D, Stacey D, Wood AM, Asselbergs FW. From lipid locus to drug target through human genomics. *Cardiovasc Res* 2018; **114**: 1258–70.
396 Surakka I, Horikoshi M, Mägi R, *et al.* The impact of low-frequency and rare variants on lipid levels. *Nat Genet* 2015; **47**: 589–97.
397 Spracklen CN, Chen P, Kim YJ, *et al.* Association analyses of East Asian individuals and trans-ancestry analyses with European individuals reveal new loci associated with cholesterol and triglyceride levels. *Hum Mol Genet* 2017; **26**: 1770–84.
398 Jadhav KS, Bauer RC. Trouble with Tribbles-1: Elucidating the Mechanism of a Genome-Wide Association Study Locus. *Arterioscler Thromb Vasc Biol* 2019; **39**: 998–1005.
399 Banting FG, Best CH, Collip JB, Campbell WR, Fletcher AA. Pancreatic Extracts in the Treatment of Diabetes Mellitus. *Can Med Assoc J* 1922; **12**: 141–6.
400 Lang DA, Matthews DR, Peto J, Turner RC. Cyclic Oscillations of Basal Plasma Glucose and Insulin Concentrations in Human Beings. *N Engl J Med* 1979; **301**: 1023–7.
401 SUTHERLAND EW, DE DUVE C. Origin and distribution of the hyperglycemic-glycogenolytic factor of the pancreas. *J Biol Chem* 1948; **175**: 663–74.
402 Gribble FM, Reimann F. Function and mechanisms of enteroendocrine cells and gut hormones in metabolism. Nat. Rev. Endocrinol. 2019; **15**: 226–37.
403 Perley MJ, Kipnis DM. Plasma insulin responses to oral and intravenous glucose: studies in normal and diabetic sujbjects. *J Clin Invest* 1967; **46**: 1954–62.
404 ELRICK H, STIMMLER L, HLAD CJ, ARAI Y. Plasma Insulin Response to Oral and Intravenous Glucose Administration1. *J Clin Endocrinol Metab* 1964; **24**: 1076–82.
405 Unger RH, Ohneda A, Valverde I, Eisentraut AM, Exton J. Characterization of the responses of circulating glucagon-like immunoreactivity to intraduodenal and intravenous administration of glucose. *J Clin Invest* 1968; **47**: 48–65.
406 Brown JC, Pederson RA, Jorpes E, Mutt V. Preparation of highly active enterogastrone. *Can J Physiol Pharmacol* 1969; **47**: 113–4.
407 Brown JC, Dryburgh JR. A gastric inhibitory polypeptide. II. The complete amino acid sequence. *Can J Biochem* 1971; **49**: 867–72.
408 Unger RH, Eisentraut AM. Entero-Insular Axis. *Arch Intern Med* 1969; **123**: 261–6.
409 Mojsov S, Weir GC, Habener JF. Insulinotropin: Glucagon-like peptide I (7-37) co-

encoded in the glucagon gene is a potent stimulator of insulin release in the perfused rat pancreas. *J Clin Invest* 1987; **79**: 616–9.
410 Kreymann B, Ghatei MA, Williams G, Bloom SR. GLUCAGON-LIKE PEPTIDE-1 7-36: A PHYSIOLOGICAL INCRETIN IN MAN. *Lancet* 1987; **330**: 1300–4.
411 Nauck M, Stöckmann F, Ebert R, Creutzfeldt W. Reduced incretin effect in Type 2 (non-insulin-dependent) diabetes. *Diabetologia* 1986; **29**: 46–52.
412 Nauck MA, Homberger E, Siegel EG, *et al.* Incretin effects of increasing glucose loads in man calculated from venous insulin and C-peptide responses. *J Clin Endocrinol Metab* 1986; **63**: 492–8.
413 Creutzfeldt W, Ebert R. New developments in the incretin concept. *Diabetologia* 1985; **28**: 565–73.
414 Drucker DJ. The biology of incretin hormones. Cell Metab. 2006; **3**: 153–65.
415 Christensen MB, Gasbjerg LS, Heimbürger SM, Stensen S, Vilsbøll T, Knop FK. GIP's involvement in the pathophysiology of type 2 diabetes. *Peptides* 2020; **125**: 170178.
416 Ørskov C, Rabenhøj L, Wettergren A, Kofod H, Holst JJ. Tissue and plasma concentrations of amidated and glycine-extended glucagon-like peptide I in humans. *Diabetes* 1994; **43**: 535–9.
417 Ugleholdt R, Poulsen MLH, Holst PJ, *et al.* Prohormone convertase 1/3 is essential for processing of the glucose-dependent insulinotropic polypeptide precursor. *J Biol Chem* 2006; **281**: 11050–7.
418 Kieffer TJ, McIntosh CH, Pederson RA. Degradation of glucose-dependent insulinotropic polypeptide and truncated glucagon-like peptide 1 in vitro and in vivo by dipeptidyl peptidase IV. *Endocrinology* 1995; **136**: 3585–96.
419 Vilsbøll T, Krarup T, Madsbad S, Holst J. Defective amplification of the late phase insulin response to glucose by gip in obese type ii diabetic patients. *Diabetologia* 2002; **45**: 1111–9.
420 Meier JJ, Hücking K, Holst JJ, Deacon CF, Schmiegel WH, Nauck MA. Reduced Insulinotropic Effect of Gastric Inhibitory Polypeptide in First-Degree Relatives of Patients with Type 2 Diabetes. *Diabetes* 2001; **50**: 2497–504.
421 Christensen M, Vedtofte L, Holst JJ, Vilsbøll T, Knop FK. Glucose-dependent insulinotropic polypeptide: A bifunctional glucose-dependent regulator of glucagon and insulin secretion in humans. *Diabetes* 2011; **60**: 3103–9.
422 Gasbjerg LS, Bergmann NC, Stensen S, *et al.* Evaluation of the incretin effect in humans using GIP and GLP-1 receptor antagonists. *Peptides* 2020; **125**: 170183.
423 Holst JJ. The incretin system in healthy humans: The role of GIP and GLP-1. *Metabolism* 2019; **96**: 46–55.
424 Nauck MA, Meier JJ. The incretin effect in healthy individuals and those with type 2 diabetes: Physiology, pathophysiology, and response to therapeutic interventions. Lancet Diabetes Endocrinol. 2016; **4**: 525–36.
425 Schirra J, Nicolaus M, Roggel R, *et al.* Endogenous glucagon-like peptide 1 controls endocrine pancreatic secretion and antro-pyloro-duodenal motility in humans. *Gut* 2006; **55**: 243–51.
426 Gasbjerg LS, Helsted MM, Hartmann B, *et al.* Separate and combined glucometabolic effects of endogenous glucose-dependent insulinotropic polypeptide and glucagon-like peptide 1 in healthy individuals. *Diabetes* 2019; **68**: 906–17.
427 Aulinger BA, Bedorf A, Kutscherauer G, *et al.* Defining the role of GLP-1 in the enteroinsulinar axis in type 2 diabetes using DPP-4 inhibition and glp-1 receptor blockade. *Diabetes* 2014; **63**: 1079–92.
428 Højberg P V., Vilsbøll T, Rabøl R, *et al.* Four weeks of near-normalisation of blood glucose improves the insulin response to glucagon-like peptide-1 and glucose-dependent insulinotropic polypeptide in patients with type 2 diabetes. *Diabetologia* 2009; **52**: 199–207.
429 Christensen MB, Calanna S, Holst JJ, Vilsbløll T, Knop FK. Glucose-dependent insulinotropic polypeptide: Blood glucose stabilizing effects in patients with type 2 diabetes. *J Clin Endocrinol Metab* 2014; **99**: 418–26.

430 Meier JJ, Nauck MA. Is secretion of glucagon-like peptide-1 reduced in type 2 diabetes mellitus? *Nat Clin Pract Endocrinol Metab* 2008; **4**: 606–7.
431 Muscelli E, Mari A, Casolaro A, *et al.* Separate Impact of Obesity and Glucose Tolerance on the Patients. *Diabetes* 2008; **57**: 1340–8.
432 Toft-Nielsen MB, Damholt MB, Madsbad S, *et al.* Determinants of the impaired secretion of glucagon-like peptide-1 in type 2 diabetic patients. *J Clin Endocrinol Metab* 2001; **86**: 3717–23.
433 Vilsbøll T, Krarup T, Deacon CF, Madsbad S, Holst JJ. Reduced Postprandial Concentrations of Intact Biologically Active Glucagon-Like Peptide 1 in Type 2 Diabetic Patients. *Diabetes* 2001; **50**: 609–13.
434 Crockett SE, Mazzaferri EL, Cataland S. Gastric inhibitory polypeptide (GIP) in maturity onset diabetes mellitus. *Diabetes* 1976; **25**: 931–5.
435 Ross SA, Brown JC, Dupre J. Hypersecretion of gastric inhibitory polypeptide following oral glucose in diabetes mellitus. *Diabetes* 1977; **26**: 525–9.
436 Vollmer K, Holst JJ, Baller B, *et al.* Predictors of Incretin Concentrations in Subjects With Normal, Impaired, and Diabetic Glucose Tolerance. *Diabetes* 2008; **57**: 678–87.
437 Calanna S, Christensen M, Holst JJ, *et al.* Secretion of glucose-dependent insulinotropic polypeptide in patients with type 2 diabetes. *Diabetes Care* 2013; **36**: 3346–52.
438 Nauck MA, Baller B, Meier JJ. Gastric inhibitory polypeptide and glucagon-like peptide-1 in the pathogenesis of type 2 diabetes. *Diabetes* 2004; **53**. DOI:10.2337/diabetes.53.suppl_3.S190.
439 Nauck MA, Heimesaat MM, Orskov C, Holst JJ, Ebert R, Creutzfeldt W. Preserved incretin activity of glucagon-like peptide 1 [7-36 amide] but not of synthetic human gastric inhibitory polypeptide in patients with type- 2 diabetes mellitus. *J Clin Invest* 1993; **91**: 301–7.
440 Thondam SK, Daousi C, Wilding JPH, *et al.* Glucose-dependent insulinotropic polypeptide promotes lipid deposition in subcutaneous adipocytes in obese type 2 diabetes patients: A maladaptive response. *Am J Physiol - Endocrinol Metab* 2017; **312**: E224–33.
441 Vilsbøll T, Knop FK, Krarup T, *et al.* The Pathophysiology of Diabetes Involves a Defective Amplification of the Late-Phase Insulin Response to Glucose by Glucose-Dependent Insulinotropic Polypeptide - Regardless of Etiology and Phenotype. *J Clin Endocrinol Metab* 2003; **88**: 4897–903.
442 Knop FK, Vilsbøll T, Højberg P V., *et al.* The insulinotropic effect of GIP is impaired in patients with chronic pancreatitis and secondary diabetes mellitus as compared to patients with chronic pancreatitis and normal glucose tolerance. *Regul Pept* 2007; **144**: 123–30.
443 Jensen DH, Aaboe K, Henriksen JE, *et al.* Steroid-induced insulin resistance and impaired glucose tolerance are both associated with a progressive decline of incretin effect in first-degree relatives of patients with type 2 diabetes. *Diabetologia* 2012; **55**: 1406–16.
444 Marathe CS, Rayner CK, Jones KL, Horowitz M. Glucagon-like peptides 1 and 2 in health and disease: A review. *Peptides* 2013; **44**: 75–86.
445 Hare KJ, Knop FK, Asmar M, *et al.* Preserved inhibitory potency of GLP-1 on glucagon secretion in type 2 diabetes mellitus. *J Clin Endocrinol Metab* 2009; **94**: 4679–87.
446 Nauck MA, Kleine N, Ørskov C, Holst JJ, Willms B, Creutzfeldt W. Normalization of fasting hyperglycaemia by exogenous glucagon-like peptide 1 (7-36 amide) in Type 2 (non-insulin-dependent) diabetic patients. *Diabetologia* 1993; **36**: 741–4.
447 Willms B, Werner J, Holst JJ, Orskov C, Creutzfeldt W, Nauck MA. Gastric emptying, glucose responses, and insulin secretion after a liquid test meal: effects of exogenous glucagon-like peptide-1 (GLP-1)-(7-36) amide in type 2 (noninsulin-dependent) diabetic patients. *J Clin Endocrinol Metab* 1996; **81**: 327–32.
448 Christensen M, Calanna S, Sparre-Ulrich AH, *et al.* Glucose-dependent insulinotropic polypeptide augments glucagon responses to hypoglycemia in type 1 diabetes.

Diabetes 2015; **64**: 72–8.
449 Lund A, Vilsboll T, Bagger JI, Holst JJ, Knop FK. The separate and combined impact of the intestinal hormones, GIP, GLP-1, and GLP-2, on glucagon secretion in type 2 diabetes. *Am J Physiol - Endocrinol Metab* 2011; **300**: 1038–46.
450 Flint A, Raben A, Astrup A, Holst JJ. Glucagon-like peptide 1 promotes satiety and suppresses energy intake in humans. *J Clin Invest* 1998; **101**: 515–20.
451 Williams DL, Baskin DG, Schwartz MW. Evidence that intestinal glucagon-like peptide-1 plays a physiological role in satiety. *Endocrinology* 2009; **150**: 1680–7.
452 Dailey MJ, Moran TH. Glucagon-like peptide 1 and appetite. *Trends Endocrinol Metab* 2013; **24**: 85–91.
453 Hellström PM, Näslund E, Edholm T, *et al.* GLP-1 suppresses gastrointestinal motility and inhibits the migrating motor complex in healthy subjects and patients with irritable bowel syndrome. *Neurogastroenterol Motil* 2008; **20**: 649–59.
454 Meier JJ, Gallwitz B, Salmen S, *et al.* Normalization of glucose concentrations and deceleration of gastric emptying after solid meals during intravenous glucagon-like peptide 1 in patients with type 2 diabetes. *J Clin Endocrinol Metab* 2003; **88**: 2719–25.
455 Asmar M, Asmar A, Simonsen L, *et al.* The gluco-and liporegulatory and vasodilatory effects of glucose-dependent insulinotropic polypeptide (GIP) are abolished by an antagonist of the human GIP receptor. *Diabetes* 2017; **66**: 2363–71.
456 Asmar M, Tangaa W, Madsbad S, *et al.* On the role of glucose-dependent insulintropic polypeptide in postprandial metabolism in humans. *Am J Physiol - Endocrinol Metab* 2010; **298**: 614–21.
457 Schirra J, Nicolaus M, Woerle HJ, Struckmeier C, Katschinski M, Göke B. GLP-1 regulates gastroduodenal motility involving cholinergic pathways. *Neurogastroenterol Motil* 2009; **21**. DOI:10.1111/j.1365-2982.2008.01246.x.
458 Wadden TA, Hollander P, Klein S, *et al.* Weight maintenance and additional weight loss with liraglutide after low-calorie-diet-induced weight loss: The SCALE Maintenance randomized study. *Int J Obes* 2013; **37**: 1443–51.
459 Kim SJ, Nian C, McIntosh CHS. Activation of lipoprotein lipase by glucose-dependent insulinotropic polypeptide in adipocytes: A role for a protein kinase B, LKB1, and AMP-activated protein kinase cascade. *J Biol Chem* 2007; **282**: 8557–67.
460 Miyawaki K, Yamada Y, Ban N, *et al.* Inhibition of gastric inhibitory polypeptide signaling prevents obesity. *Nat Med* 2002; **8**: 738–42.
461 Getty-Kaushik L, Song DH, Boylan MO, Corkey BE, Wolfe MM. Glucose-Dependent Insulinotropic Polypeptide Modulates Adipocyte Lipolysis and Reesterification*. *Obesity* 2006; **14**: 1124–31.
462 Baba ASH, Harper JMM, Buttery PJ. Effects of gastric inhibitory polypeptide, somatostatin and epidermal growth factor on lipogenesis in ovine adipose explants. *Comp Biochem Physiol - B Biochem Mol Biol* 2000; **127**: 173–82.
463 Yip RG-C, Boylan MO, Kieffer TJ, Wolfe MM. Functional GIP receptors are present on adipocytes. *Endocrinology* 1998; **139**: 4004–7.
464 Timper K, Grisouard J, Sauter NS, *et al.* Glucose-dependent insulinotropic polypeptide induces cytokine expression, lipolysis, and insulin resistance in human adipocytes. *Am J Physiol - Endocrinol Metab* 2013; **304**: 1–13.
465 Oben J, Morgan L, Fletcher J, Marks V. Effect of the entero-pancreatic hormones, gastric inhibitory polypeptide and glucagon-like polypeptide-1(7-36) amide, on fatty acid synthesis in explants of rat adipose tissue. *J Endocrinol* 1991; **130**: 267–72.
466 Killion EA, Wang J, Yie J, *et al.* Anti-obesity effects of GIPR antagonists alone and in combination with GLP-1R agonists in preclinical models. *Sci Transl Med* 2018; **10**: eaat3392.
467 Miyawaki K, Yamada Y, Yano H, *et al.* Glucose intolerance caused by a defect in the entero-insular axis: A study in gastric inhibitory polypeptide receptor knockout mice. *Proc Natl Acad Sci* 1999; **96**: 14843–7.
468 Yip RGC, Wolfe MM. GIP biology and fat metabolism. Life Sci. 1999; **66**: 91–103.
469 Finan B, Müller TD, Clemmensen C, Perez-Tilve D, DiMarchi RD, Tschöp MH.

Reappraisal of GIP Pharmacology for Metabolic Diseases. Trends Mol. Med. 2016; **22**: 359–76.
470 Mroz PA, Finan B, Gelfanov V, *et al.* Optimized GIP analogs promote body weight lowering in mice through GIPR agonism not antagonism. *Mol Metab* 2019; **20**: 51–62.
471 Adriaenssens AE, Biggs EK, Darwish T, *et al.* Glucose-Dependent Insulinotropic Polypeptide Receptor-Expressing Cells in the Hypothalamus Regulate Food Intake. *Cell Metab* 2019; **30**: 987-996.e6.
472 Christensen MB, Gasbjerg LS, Heimburger SM, Stensen S, Vilsbøll T, Knop FK. GIP's involvement in the pathophysiology of type 2 diabetes. *Peptides* 2019. DOI:10.1016/j.marmicro.2019.101773.
473 McIntosh CHS, Bremsak I, Lynn FC, *et al.* Glucose-dependent insulinotropic polypeptide stimulation of lipolysis in differentiated 3T3-L1 cells: Wortmannin-sensitive inhibition by insulin. *Endocrinology* 1999; **140**: 398–404.
474 Bergmann NC, Lund A, Gasbjerg LS, *et al.* Effects of combined GIP and GLP-1 infusion on energy intake, appetite and energy expenditure in overweight/obese individuals: a randomised, crossover study. *Diabetologia* 2019; **62**: 665–75.
475 Møller CL, Vistisen D, Færch K, *et al.* Glucose-dependent insulinotropic polypeptide is associated with lower low-density lipoprotein but unhealthy fat distribution, independent of insulin: The addition-pro study. *J Clin Endocrinol Metab* 2016; **101**: 485–93.
476 Gasbjerg LS, Gabe MBN, Hartmann B, *et al.* Glucose-dependent insulinotropic polypeptide (GIP) receptor antagonists as anti-diabetic agents. *Peptides* 2018; **100**: 173–81.
477 Flatt PR. Dorothy hodgkin lecture 2008 gastric inhibitory polypeptide (GIP) revisited: A new therapeutic target for obesity-diabetes? *Diabet Med* 2008; **25**: 759–64.
478 Irwin N, Flatt PR. Therapeutic potential for GIP receptor agonists and antagonists. Best Pract. Res. Clin. Endocrinol. Metab. 2009; **23**: 499–512.
479 Heimbürger SM, Bergmann NC, Augustin R, Gasbjerg LS, Christensen MB, Knop FK. Glucose-dependent insulinotropic polypeptide (GIP) and cardiovascular disease. *Peptides* 2020; **125**. DOI:10.1016/j.peptides.2019.170174.
480 Greenwell AA, Chahade JJ, Ussher JR. Cardiovascular biology of the GIP receptor. *Peptides* 2020; **125**: 170228.
481 Jujić A, Atabaki-Pasdar N, Nilsson PM, *et al.* Glucose-dependent insulinotropic peptide and risk of cardiovascular events and mortality: a prospective study. *Diabetologia* 2020; **63**: 1043–54.
482 Scott RA, Amouyel P, Müller-nurasyid M. A genomic approach to therapeutic target validation identifies a glucose-lowering GLP1R variant protective for coronary heart disease. *Sci Transl Med* 2016; **8**: 1–13.
483 Christensen MB, Lund A, Calanna S, *et al.* Glucose-Dependent Insulinotropic Polypeptide (GIP) inhibits bone resorption independently of insulin and glycemia. *J Clin Endocrinol Metab* 2018; **103**: 288–94.
484 Berglund LM, Lyssenko V, Ladenvall C, *et al.* Glucose-dependent insulinotropic polypeptide stimulates osteopontin expression in the vasculature via endothelin-1 and CREB. *Diabetes* 2016; **65**: 239–54.
485 Planas-Rigol E, Terrades-Garcia N, Corbera-Bellalta M, *et al.* Endothelin-1 promotes vascular smooth muscle cell migration across the artery wall: A mechanism contributing to vascular remodelling and intimal hyperplasia in giant-cell arteritis. *Ann Rheum Dis* 2017; **76**: 1623–33.
486 Wang KX, Denhardt DT. Osteopontin: Role in immune regulation and stress responses. *Cytokine Growth Factor Rev* 2008; **19**: 333–45.
487 Asmar M, Simonsen L, Madsbad S, Stallknecht B, Holst JJ, Bülow J. Glucose-dependent insulinotropic polypeptide may enhance fatty acid re-esterification in subcutaneous abdominal adipose tissue in lean humans. *Diabetes* 2010; **59**: 2160–3.
488 Brown JC, Dryburgh JR, Ross SA, Dupre J. Identification and actions of gastric inhibitory polypeptide. *Recent Prog Horm Res* 1975; **Vol.31**: 487–532.
489 Bailey CJ. GIP analogues and the treatment of obesity-diabetes. *Peptides* 2020; **125**.

DOI:10.1016/j.peptides.2019.170202.
490 Mohammad S, Patel RT, Bruno J, Panhwar MS, Wen J, McGraw TE. A Naturally Occurring GIP Receptor Variant Undergoes Enhanced Agonist-Induced Desensitization, Which Impairs GIP Control of Adipose Insulin Sensitivity. *Mol Cell Biol* 2014; **34**: 3618–29.
491 Gault VA, Parker JC, Harriott P, Flatt PR, O'Harte FPM. Evidence that the major degradation product of glucose-dependent insulinotropic polypeptide, GIP(3-42), is a GIP receptor antagonist in vivo. *J Endocrinol* 2002; **175**: 525–33.
492 Sparre-Ulrich AH, Hansen LS, Svendsen B, *et al.* Species-specific action of (Pro3)GIP - A full agonist at human GIP receptors, but a partial agonist and competitive antagonist at rat and mouse GIP receptors. *Br J Pharmacol* 2016; **173**: 27–38.
493 Sparre-Ulrich AH, Gabe MN, Gasbjerg LS, *et al.* GIP(3–30)NH2 is a potent competitive antagonist of the GIP receptor and effectively inhibits GIP-mediated insulin, glucagon, and somatostatin release. *Biochem Pharmacol* 2017; **131**: 78–88.
494 Gabe MBN, Sparre-Ulrich AH, Pedersen MF, *et al.* Human GIP(3-30)NH2 inhibits G protein-dependent as well as G protein-independent signaling and is selective for the GIP receptor with high-affinity binding to primate but not rodent GIP receptors. *Biochem Pharmacol* 2018; **150**: 97–107.
495 Astrup A, Rössner S, Van Gaal L, *et al.* Effects of liraglutide in the treatment of obesity: a randomised, double-blind, placebo-controlled study. *Lancet* 2009; **374**: 1606–16.
496 Drucker DJ, Buse JB, Taylor K, *et al.* Exenatide once weekly versus twice daily for the treatment of type 2 diabetes: a randomised, open-label, non-inferiority study. *Lancet* 2008; **372**: 1240–50.
497 Pfeffer MA, Claggett B, Diaz R, *et al.* Lixisenatide in patients with type 2 diabetes and acute coronary syndrome. *N Engl J Med* 2015; **373**: 2247–57.
498 Holman RR, Bethel MA, Mentz RJ, *et al.* Effects of once-weekly exenatide on cardiovascular outcomes in type 2 diabetes. *N Engl J Med* 2017; **377**: 1228–39.
499 Piteau S, Olver A, Kim SJ, *et al.* Reversal of islet GIP receptor down-regulation and resistance to GIP by reducing hyperglycemia in the Zucker rat. *Biochem Biophys Res Commun* 2007; **362**: 1007–12.
500 Knerr PJ, Mowery SA, Finan B, Perez-Tilve D, Tschöp MH, DiMarchi RD. Selection and progression of unimolecular agonists at the GIP, GLP-1, and glucagon receptors as drug candidates. *Peptides* 2020; **125**: 170225.
501 Frias JP, Bastyr EJ, Vignati L, *et al.* The Sustained Effects of a Dual GIP/GLP-1 Receptor Agonist, NNC0090-2746, in Patients with Type 2 Diabetes. *Cell Metab* 2017; **26**: 343-352.e2.
502 Schmitt C, Portron A, Jadidi S, Sarkar N, DiMarchi R. Pharmacodynamics, pharmacokinetics and safety of multiple ascending doses of the novel dual glucose-dependent insulinotropic polypeptide/glucagon-like peptide-1 agonist RG7697 in people with type 2 diabetes mellitus. *Diabetes, Obes Metab* 2017; **19**: 1436–45.
503 Frias JP, Nauck MA, Van J, *et al.* Efficacy and safety of LY3298176, a novel dual GIP and GLP-1 receptor agonist, in patients with type 2 diabetes: a randomised, placebo-controlled and active comparator-controlled phase 2 trial. *Lancet* 2018; **392**: 2180–93.
504 Frias JP, Bastyr EJ, Vignati L, *et al.* The Sustained Effects of a Dual GIP/GLP-1 Receptor Agonist, NNC0090-2746, in Patients with Type 2 Diabetes. *Cell Metab* 2017; **26**: 343-352.e2.
505 Gasbjerg LS, Gabe MBN, Hartmann B, *et al.* Glucose-dependent insulinotropic polypeptide (GIP) receptor antagonists as anti-diabetic agents. *Peptides* 2018; **100**: 173–81.
506 Floris M, Olla S, Schlessinger D, Cucca F. Genetic-Driven Druggable Target Identification and Validation. *Trends Genet* 2018; **34**: 558–70.
507 Müssig K, Staiger H, Machicao F, *et al.* Association of type 2 diabetes candidate polymorphisms in KCNQ1 with incretin and insulin secretion. *Diabetes* 2009; **58**: 1715–20.
508 Almgren P, Lindqvist A, Krus U, *et al.* Genetic determinants of circulating GIP and GLP-

1 concentrations. *JCI Insight* 2017; **2**. DOI:10.1172/jci.insight.93306.
509 Stefan N, Machicao F, Staiger H, *et al.* Polymorphisms in the gene encoding adiponectin receptor 1 are associated with insulin resistance and high liver fat. *Diabetologia* 2005; **48**: 2282–91.
510 Lindgren O, Carr RD, Deacon CF, *et al.* Incretin Hormone and Insulin Responses to Oral *Versus* Intravenous Lipid Administration in Humans. *J Clin Endocrinol Metab* 2011; **96**: 2519–24.
511 Deacon CF, Johnsen AH, Holst JJ. Degradation of glucagon-like peptide-1 by human plasma in vitro yields an N-terminally truncated peptide that is a major endogenous metabolite in vivo. *J Clin Endocrinol Metab* 1995; **80**: 952–7.
512 Deacon CF, Nauck MA, Meier J, Hücking K, Holst JJ. Degradation of Endogenous and Exogenous Gastric Inhibitory Polypeptide in Healthy and in Type 2 Diabetic Subjects as Revealed Using a New Assay for the Intact Peptide [1]. *J Clin Endocrinol Metab* 2000; **85**: 3575–81.
513 Emilsson V, Ilkov M, Lamb JR, *et al.* Co-regulatory networks of human serum proteins link genetics to disease. *Science (80-)* 2018; **361**. DOI:10.1126/science.aaq1327.
514 Hoffmann TJ, Theusch E, Haldar T, *et al.* A large electronic-health-record-based genome-wide study of serum lipids. *Nat Genet* 2018. DOI:10.1038/s41588-018-0064-5.
515 Dichgans M, Malik R, König IR, *et al.* Shared genetic susceptibility to ischemic stroke and coronary artery disease : A genome-wide analysis of common variants. *Stroke* 2014; **45**: 24–36.
516 Turcot V, Lu Y, Highland HM, *et al.* Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity. *Nat Genet* 2018; **50**: 26–35.
517 Pulit SL, Stoneman C, Morris AP, *et al.* Meta-Analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum Mol Genet* 2019; **28**: 166–74.
518 Nitz I, Fisher E, Welkert C, *et al.* Association analyses of GIP and GIPR polymorphisms with traits of the metabolic syndrome. *Mol Nutr Food Res* 2007; **51**: 1046–52.
519 Finan B, Ma T, Ottaway N, *et al.* Unimolecular dual incretins maximize metabolic benefits in rodents, monkeys, and humans. *Sci Transl Med* 2013; **5**. DOI:10.1126/scitranslmed.3007218.
520 Coskun T, Sloop KW, Loghin C, *et al.* LY3298176, a novel dual GIP and GLP-1 receptor agonist for the treatment of type 2 diabetes mellitus: From discovery to clinical proof of concept. *Mol Metab* 2018; **18**: 3–14.
521 Wilson JM, Nikooienejad A, Robins DA, *et al.* The dual glucose-dependent insulinotropic peptide and glucagon-like peptide-1 receptor agonist, tirzepatide, improves lipoprotein biomarkers associated with insulin resistance and cardiovascular risk in patients with type 2 diabetes. *Diabetes, Obes Metab* 2020. DOI:10.1111/dom.14174.
522 Mori Y, Matsui T, Hirano T, Yamagishi SI. GIP as a potential therapeutic target for atherosclerotic cardiovascular disease– a systematic review. *Int J Mol Sci* 2020; **21**: 8–10.
523 Nagashima M, Watanabe T, Terasaki M, *et al.* Native incretins prevent the development of atherosclerotic lesions in apolipoprotein e knockout mice. *Diabetologia* 2011; **54**: 2649–59.
524 Gerstein HC, Colhoun HM, Dagenais GR, *et al.* Dulaglutide and cardiovascular outcomes in type 2 diabetes (REWIND): a double-blind, randomised placebo-controlled trial. *Lancet* 2019; **394**: 121–30.
525 Hernandez AF, Green JB, Janmohamed S, *et al.* Albiglutide and cardiovascular outcomes in patients with type 2 diabetes and cardiovascular disease (Harmony Outcomes): a double-blind, randomised placebo-controlled trial. *Lancet* 2018; **392**: 1519–29.
526 Marso SP, Daniels GH, Frandsen KB, *et al.* Liraglutide and cardiovascular outcomes in

type 2 diabetes. *N Engl J Med* 2016; **375**: 311–22.
527 Marso SP, Bain SC, Consoli A, *et al.* Semaglutide and cardiovascular outcomes in patients with type 2 diabetes. *N Engl J Med* 2016; **375**: 1834–44.
528 Saxena R, Hivert MF, Langenberg C, *et al.* Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nat Genet* 2010; **42**: 142–8.
529 Lyssenko V, Eliasson L, Kotova O, *et al.* Pleiotropic effects of GIP on islet function involve osteopontin. *Diabetes* 2011; **60**: 2424–33.
530 Riboli E. Nutrition and cancer: Background and rationale of the European prospective investigation into cancer and nutrition (EPIC). *Ann Oncol* 1992; **3**: 783–91.
531 Lotta LA, Pietzner M, Stewart ID, *et al.* Cross-platform genetic discovery of small molecule products of metabolism and application to clinical outcomes. *bioRxiv* 2020; : 2020.02.03.932541.
532 Bateman A. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res* 2019; **47**: D506–15.
533 Loh PR, Tucker G, Bulik-Sullivan BK, *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* 2015; **47**: 284–90.
534 Nelson CP, Goel A, Butterworth AS, *et al.* Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat Genet* 2017; **49**: 1385–91.
535 Allara E, Morani G, Carter P, *et al.* Genetic Determinants of Lipids and Cardiovascular Disease Outcomes: A Wide-Angled Mendelian Randomization Investigation. *Circ Genomic Precis Med* 2019; **12**: 543–51.
536 Malik R, Chauhan G, Traylor M, *et al.* Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet* 2018; **50**: 524–37.
537 Prokopenko I, Poon W, Mägi R, *et al.* A Central Role for GRB10 in Regulation of Islet Function in Man. *PLoS Genet* 2014; **10**: 1–13.
538 Langenberg C, Sharp S, Forouhi NG, *et al.* Design and cohort description of the InterAct Project: An examination of the interaction of genetic and lifestyle factors on the incidence of type 2 diabetes in the EPIC Study. *Diabetologia* 2011; **54**: 2272–82.
539 Shungin D, Winkler TW, Croteau-Chonka DC, *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature* 2015; **518**: 187–96.
540 Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian randomization. *Stat Methods Med Res* 2017; **26**: 2333–55.
541 Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* 2015; **4**: 1–16.
542 Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011; **88**: 76–82.
543 Buse JB, Ginsberg HN, Bakris GL, *et al.* Primary prevention of cardiovascular diseases in people with diabetes mellitus: a scientific statement from the American Heart Association and the American Diabetes Association. *Circulation* 2007; **115**: 114–26.
544 Huang Y, Mahley RW. Apolipoprotein E: Structure and function in lipid metabolism, neurobiology, and Alzheimer's diseases. *Neurobiol Dis* 2014; **72**: 3–12.
545 Hotamisligil GS. Inflammation and metabolic disorders. Nat. Rev. 2006; **444**: 860–7.
546 Belmont JW, Boudreau A, Leal SM, *et al.* A haplotype map of the human genome. *Nature* 2005; **437**: 1299–320.
547 Fadista J, Manning AK, Florez JC, Groop L. The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. *Eur J Hum Genet* 2016; **24**: 1202–5.
548 Thomas D, Xie R, Gebregziabher M. Two-stage sampling designs for gene association studies. *Genet Epidemiol* 2004; **27**: 401–14.
549 Satagopan JM, Venkatraman ES, Begg CB. Two-stage designs for gene-disease association studies with sample size constraints. *Biometrics* 2004; **60**: 589–97.
550 Froom P, Melamed S, Kristal-Boneh E, Benbassat J, Ribak J. Healthy volunteer effect

in industrial workers. *J Clin Epidemiol* 1999; **52**: 731–5.
551 Fry A, Littlejohns TJ, Sudlow C, *et al.* Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants with Those of the General Population. *Am J Epidemiol* 2017; **186**: 1026–34.
552 Hindorff LA, Bonham VL, Brody LC, *et al.* Prioritizing diversity in human genomics research. *Nat Rev Genet* 2018; **19**: 175–85.
553 Bustamante CD, De La Vega FM, Burchard EG. Genomics for the world. *Nature* 2011; **475**: 163–5.
554 Popejoy AB, Fullerton SM. Genomics is failing on diversity. *Nature* 2016; **538**: 161–4.
555 Smolen JS, Beaulieu A, Rubbert-Roth A, *et al.* Effect of interleukin-6 receptor inhibition with tocilizumab in patients with rheumatoid arthritis (OPTION study): a double-blind, placebo-controlled, randomised trial. *Lancet* 2008; **371**: 987–97.
556 Genovese MC, Fleischmann R, Kivitz AJ, *et al.* Sarilumab plus methotrexate in patients with active rheumatoid arthritis and inadequate response to methotrexate: Results of a phase III study. *Arthritis Rheumatol* 2015; **67**: 1424–37.
557 Emery P, Keystone E, Tony HP, *et al.* IL-6 receptor inhibition with tocilizumab improves treatment outcomes in patients with rheumatoid arthritis refractory to anti-tumour necrosis factor biologicals: Results from a 24-week multicentre randomised placebo-controlled trial. *Ann Rheum Dis* 2008; **67**: 1516–23.
558 Xie HG, Kim RB, Wood AJ, Stein CM. Molecular basis of ethnic differences in drug disposition and response. *Annu Rev Pharmacol Toxicol* 2001; **41**: 815–50.
559 Soldin OP, Mattison DR. Sex differences in pharmacokinetics and pharmacodynamics. *Clin Pharmacokinet* 2009; **48**: 143–57.
560 Kauppi L, Jeffreys AJ, Keeney S. Where the crossovers are: Recombination distributions in mammals. *Nat Rev Genet* 2004; **5**: 413–24.
561 Fernández-Rhodes L, Young KL, Lilly AG, *et al.* Importance of Genetic Studies of Cardiometabolic Disease in Diverse Populations. *Circ Res* 2020; : 1816–40.
562 Columb MO, Atkinson MS. Statistical analysis: sample size and power estimations. *BJA Educ* 2016; **16**: 159–61.
563 Mayeux R. Biomarkers: Potential Uses and Limitations. *NeuroRx* 2004; **1**: 182–8.
564 Spencer CCA, Su Z, Donnelly P, Marchini J. Designing genome-wide association studies: Sample size, power, imputation, and the choice of genotyping chip. *PLoS Genet* 2009; **5**. DOI:10.1371/journal.pgen.1000477.
565 Yengo L, Sidorenko J, Kemper KE, *et al.* Meta-analysis of genome-wide association studies for height and body mass index in ~ 700 , 000 individuals of European ancestry. *Hum Mol Genet* 2018; : 1–25.
566 Robinson MW, Harmon C, O'Farrelly C. Liver immunology and its role in inflammation and homeostasis. *Cell Mol Immunol* 2016; **13**: 267–76.
567 Anstee QM, Day CP. The genetics of NAFLD. *Nat Rev Gastroenterol Hepatol* 2013; **10**: 645–55.
568 Buch S, Stickel F, Trépo E, *et al.* A genome-wide association study confirms PNPLA3 and identifies TM6SF2 and MBOAT7 as risk loci for alcohol-related cirrhosis. *Nat Genet* 2015; **47**: 1443–8.
569 Chalasani N, Guo X, Loomba R, *et al.* Genome-Wide Association Study Identifies Variants Associated with Histologic Features of Nonalcoholic Fatty Liver Disease. *Gastroenterology* 2011; **139**: 1567–76.
570 Chambers JC, Zhang W, Sehmi J, *et al.* Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nat Genet* 2011; **43**: 1131–8.
571 Kitamoto T, Kitamoto A, Yoneda M, *et al.* Genome-wide scan revealed that polymorphisms in the PNPLA3, SAMM50, and PARVB genes are associated with development and progression of nonalcoholic fatty liver disease in Japan. *Hum Genet* 2013; **132**: 783–92.
572 Mann JP, Anstee QM. PNPLA3 and obesity: A synergistic relationship in NAFLD. *Nat Rev Gastroenterol Hepatol* 2017; **14**: 506–7.
573 Speliotes EK, Yerges-Armstrong LM, Wu J, *et al.* Genome-wide association analysis

identifies variants associated with nonalcoholic fatty liver disease that have distinct effects on metabolic traits. *PLoS Genet* 2011; **7**. DOI:10.1371/journal.pgen.1001324.
574 Stender S, Kozlitina J, Nordestgaard BG, Tybjærg-Hansen A, Hobbs HH, Cohen JC. Adiposity amplifies the genetic risk of fatty liver disease conferred by multiple loci. *Nat Genet* 2017; **49**: 842–7.
575 Kingsmore KM, Grammer AC, Lipsky PE. Drug repurposing to improve treatment of rheumatic autoimmune inflammatory diseases. *Nat Rev Rheumatol* 2020; **16**: 32–52.
576 Bresnihan B, Alvaro-Gracia JM, Cobby M, *et al.* Treatment of rheumatoid arthritis with recombinant human interleukin-1 receptor antagonist. *Arthritis Rheum* 1998; **41**: 2196–204.
577 Weinblatt ME, Kremer JM, Bankhurst AD, *et al.* A Trial of Etanercept, a Recombinant Tumor Necrosis Factor Receptor:Fc Fusion Protein, in Patients with Rheumatoid Arthritis Receiving Methotrexate. *N Engl J Med* 1999; **340**: 253–9.
578 Kim SJ, Nian C, McIntosh CHS. GIP increases human adipocyte LPL expression through CREB and TORC2-mediated trans-activation of the LPL gene. *J Lipid Res* 2010; **51**: 3145–57.
579 Yip RGC, Wolfe MM. GIP biology and fat metabolism. *Life Sci* 1999; **66**: 91–103.
580 Mroz PA, Finan B, Gelfanov V, *et al.* Optimized GIP analogs promote body weight lowering in mice through GIPR agonism not antagonism. *Mol Metab* 2019; **20**: 51–62.
581 Egger M, Smith GD, Phillips AN. Meta-analysis: Principles and procedures. Br. Med. J. 1997; **315**: 1533–7.
582 van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Stat Methods Med Res* 2007; **16**: 219–42.
583 van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. *J Stat Softw* 2011; **45**: 1–67.
584 Opgen-rhein R, Strimmer K. Inferring gene dependency networks from genomic longitudinal data: a functional data approach. *Revstat* 2006; **4**: 53–65.
585 Schafer J, Strimmer K. A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics. *Stat Appl Genet Mol Biol* 2005; **4**.
586 Opgen-rhein R, Strimmer K. USING REGULARIZED DYNAMIC CORRELATION TO INFER GENE DEPENDENCY NETWORKS FROM TIME-SERIES MICROARRAY DATA Rainer Opgen-Rhein and Korbinian Strimmer Department of Statistics , University of Munich ,. *4th Int Work Comput Syst Biol WCSB 2006* 2005; : 73–6.
587 Revelle WR. psych: Procedures for Personality and Psychological Research. 2017. https://www.scholars.northwestern.edu/en/publications/psych-procedures-for-personality-and-psychological-research (accessed Oct 16, 2020).
588 Rubin DB. Multiple Imputation for Nonresponse in Surveys. Hoboken, NJ, USA: John Wiley & Sons, Inc., 1987 DOI:10.1002/9780470316696.
589 Schwarzer G, Mair P, Hatzinger R. meta : An R Package for Meta-Analysis meta : An R Package for Meta-Analysis. 2016.
590 Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants: Fig. 1. *Bioinformatics* 2015; **31**: 3555–7.

Supplementary Materials

Supplementary Methods

Methods S4.1. Additional methods for observational epidemiology analyses.

Systematic review and meta-analysis of prospective studies investigating the association of IL-6 levels with incident type 2 diabetes

A systematic review and meta-analysis of studies investigating the association of IL-6 levels with incident type 2 diabetes in population-based cohorts was conducted. A previous meta-analysis had been performed by Wang et al[186] for studies with a publication date up to 10 February 2012. In addition to studies identified by Wang et al[186], novel results from the EPIC-Norfolk study and from studies identified with a new and updated systematic search were included in the meta-analysis reported in this manuscript. A review protocol has not previously been published and will be described in detail in this section. EPIC-Norfolk constitutes the largest study to date and the one with the longest duration of follow-up (**Table S4.2**). A search of the PubMed electronic database was performed to identify studies investigating the association of IL-6 levels with incident type 2 diabetes. Prospective studies published between 10 February 2012 and 30 October 2018 were considered. A full list of search terms can be found in **Table S4.3**. Studies focusing on prevalent type 2 diabetes, type 1 diabetes or gestational diabetes were excluded as were review articles and animal studies. Additionally, studies were excluded if they included cohorts that overlapped those already included in the meta-analysis by Wang and colleagues. The initial database search yielded 379 recent publications, the titles of which were initially screened using the above-mentioned criteria. Two of the authors (NB and RLS) independently screened the titles of the identified publications and conferred to ensure consistency. Following this, 14 relevant publications were identified, and their abstracts were further screened by the same authors to ensure consistency. Overall, three recently published studies[225,239,240] were identified and their results were included in the meta-analysis, which included 15 studies overall. Quality ratings for the 15 selected studies were adjudicated based on criteria evaluating study size and representativeness, reliability of exposure measurement, reliability of type 2 diabetes ascertainment, and adjustment for possible confounders (**Table S4.4**).

Estimates of association (risk ratios, odds ratios, hazard ratios – all assumed to approximate the hazard ratio) with type 2 diabetes were calculated per 1 log pg/mL higher IL-6 levels for studies included in the Wang *et al*[186] analysis. In the case of the three newly identified studies, only the results of Koloverou *et al*[239] were presented per 1 log pg/mL higher IL-6 levels. The results of Dallmeier *et al*[240] and Marques-Vidal *et al*[225] were originally presented per 1 SD higher log IL-6 levels. These estimates were converted to units per 1 log pg/mL higher IL-6 levels using a conversion factor of 0.8 (corresponding to the SD of log IL-6 levels in the EPIC-Norfolk study). Study characteristics and relative risks for the included studies are summarised in **Table S4.2**. Fixed-effects meta-analysis was conducted including all 15 available studies consisting of a total of 5,421 incident type 2 diabetes cases and 31,562 non-cases. The combined hazard ratio was reported per 1 log pg/mL higher IL-6 levels. The I^2 heterogeneity statistic was used to estimate significant heterogeneity between study effect estimates.

To assess potential publication bias, the Egger test[581] was used. Furthermore, funnel plots illustrating the relationship between the hazard ratio and standard error of each study were drawn (**Figure S4.11**). No significant evidence of publication bias was found (P = 0.44). The funnel plot showed that two studies were on the contour line and could potentially bias the overall estimate (**Figure S4.11**). Subgroup analysis excluding these two studies demonstrated no change in the overall estimate (1.24; 95% CI, 1.17, 1.32; $P=1x10^{-12}$). We also sought to estimate potential bias induced via differing IL-6 measurement methods between studies. Three studies that measured IL-6 levels using techniques other than electrochemiluminescent

methods such as ELISA were excluded. The remaining prospective studies were meta-analysed, resulting in the same effect estimate as the main analysis (HR 1.24; 95% CI, 1.16, 1.33; P = $1x10^{-10}$). Differential IL-6 measurements between studies were therefore ruled out as a potential source of bias.

DEXA protocol in Fenland.
In the Fenland study overall and regional body fat mass was quantified by dual-energy X-ray absorptiometry (DEXA) using a Lunar Prodigy advanced fan beam scanner (GE Healthcare, Bedford, UK), encore v14.10.022 and CoreScan® software (GE Healthcare, Bedford UK). All body images were manually processed, and demarcations corrected following a standardized protocol. The trunk included neck, chest, abdominal and pelvic areas. The legs were defined as the region below the lower borders of the trunk. The abdomen was defined as the portion of the trunk between ribs and pelvis, defined by the outline of the iliac crest. The gluteofemoral region included hips and upper thighs. Visceral abdominal fat mass was estimated using subcutaneous fat width and the anteroposterior thickness of the abdominal wall. These measurements were used to extrapolate abdominal subcutaneous fat mass. Visceral abdominal fat mass was calculated by subtracting the subcutaneous abdominal fat mass from the total abdominal fat mass.

Methods S4.2. Additional methods for genetic association analyses.
Re-scaling the association estimates of IL6R Asp358Ala with continuous metabolic traits
To aid the interpretation of association estimates with different continuous metabolic traits (e.g. HbA1c or fasting insulin), the associations were reported in both clinical and standardized units of each trait per copy of the *IL6R* 358Ala allele. Clinical units aid interpretation of the association results, while standard deviation units enable comparison of association estimates across outcomes. Estimates from summary-level results were converted to standardized units or to clinical units using the standard deviation (SD) in the Fenland or in the UK Biobank as summarized in **Table S4.6**.
In these analyses, individuals carrying the *IL6R* 358Ala allele had numerically lower WHR adjusted for BMI, a risk factor for type 2 diabetes[30] However, the difference in WHR adjusted for BMI observed for the *IL6R* 358Ala allele is unlikely to explain its association with lower risk of type 2 diabetes. From previous analyses, a 1-SD increase in WHR adjusted for BMI is associated with an odds ratio for type 2 diabetes of 1.79[30] On the basis of this estimate, given that each copy of *IL6R* 358Ala is associated with 0.004 SD lower BMI-adjusted WHR, the expected odds ratio for type 2 diabetes would be 0.998 (i.e. a very small estimated effect size) if the association of the *IL6R* 358Ala allele was entirely due to BMI-adjusted WHR. Since the 95% confidence interval estimated for the association between the *IL6R* 358Ala allele and type 2 diabetes excludes 0.998 (i.e. odds ratio per allele from 0.970 to 0.989), it is unlikely that the association of this variant with diabetes is only due to an association with more favourable fat distribution.

Estimation of heritability and variance in the risk for type 2 diabetes and coronary disease explained by the IL6R 358Ala allele
In stage 2, estimates of heritability for type 2 diabetes and coronary heart disease were obtained using linkage disequilibrium score regression[287] in UK Biobank using the HapMap3 European ancestry CEU population as reference panel (**Table S4.9**). Variants included in the regression model were restricted to those present in the reference panel with minor allele frequency above 5% to retain only common variants. Type 2 diabetes and coronary heart disease prevalence in the general population were assumed to be equivalent to the prevalence in UK Biobank, 5.51% and 5.50% respectively. The heritability in these conditions explained by the *IL6R* 358Ala allele was estimated by dividing the heritability of the disease by the variance explained by the genetic variant.

Exclusion of misdiagnosed cases of type 1 diabetes using a polygenic score approach

To assess whether the association with type 2 diabetes of *IL6R* 358Ala might be affected by misclassification of cases of type 1 diabetes as type 2 diabetes, an approach based on a 29-variant polygenic risk score[350] for type 1 diabetes was used. Using individual-level data from UK Biobank, this polygenic score was calculated for each participants by adding the number of copies of each contributing variant weighted by its per allele association estimate with type 1 diabetes as previously described[350].

Individuals with values of the polygenic score below the median in people with diabetes (type 1 and type 2 combined; value in UK Biobank, 13.04) have been shown to be highly unlikely to have type 1 diabetes (sensitivity to detect type 1 diabetes cases for exclusion, 96%; false negative rate, 4%)[350] This means that if one were to restrict an analysis to people below that value, only 4% of type 1 diabetes cases originally included would remain in the analysis. This restriction does not bias the association of the *IL6R* 358Ala allele as the *IL6R* locus is not included in the type 1 diabetes genetic score and therefore is in linkage equilibrium with (i.e. inherited randomly regarding) the loci included in the score.

Given that type 1 diabetes accounts for ~10% of diabetes in the general population, one would expect that amongst the approximately 22,000 cases of type 2 diabetes included in this analysis, cases of type 1 diabetes erroneously classified as type 2 diabetes would be 2,200 if there were a 100% misclassification. Even in the extreme scenario of 100% misclassification, restricting to only people with polygenic score for type 1 diabetes below the median would lower the number of type 1 diabetes present in the analysis to just 44 (0.4% of overall diabetes cases; corresponding to 1,100 misclassified cases left in the halved sample size multiplied by 4% false negatives). More likely scenarios of a lower misclassification rate yield numbers of misclassified type 1 diabetes cases left in the sensitivity analysis in the single digits (**Figures S4.7B and S4.7C**). On this basis, the association with type 2 diabetes was estimated after restricting for participants below the median of the type 1 diabetes polygenic score.

Projection of the potential benefit of IL6R blocking therapy on the risk of type 2 diabetes.
Both experimental[310] and genetic[88,160] association studies have shown that rs2228145-C (358Ala) closely mimics some of the effects of IL-6R inhibitory therapy. However, an association of this variant with lower risk of type 2 diabetes, as shown in this study, does not necessarily mean that IL-6R inhibitory therapy will produce clinically meaningful reduction in the risk of type 2 diabetes. Even if the effects of a genetic variant and that of a drug on the target were qualitatively the same, two differences between genetic variant and treatment with the drug in randomized clinical trials would remain. First, genetic variants are usually associated with small differences in the activity of the target gene, while drugs usually have large effects on target activity (different effect size)[222,353]. Second, differences in target gene activity associated with a genetic variant are lifelong, while the effect of drugs is usually assessed in trials of short duration (duration of exposure). With several assumptions, these differences can sometimes be modelled to formulate projections of the likely efficacy of drug treatment based on association magnitude of genetic variants that mimic that drug treatment.

In this study, projections of potential benefit of IL-6R blocking therapy on the risk of type 2 diabetes in a primary prevention setting were formulated using (a) the genetic association of *IL6R* 358Ala with type 2 diabetes from this study, (b) the genetic association of *IL6R* 358Ala with C-reactive protein (CRP; used as biomarker of target engagement for IL-6R) in this and other studies[160] and (c) the effects of IL-6R blocking therapy on CRP (the most-widely used biomarker of target engagement for IL-6R blocking therapy) in randomized clinical trials of tocilizumab, an IL-6R inhibitor[88].

First, to model the different effect size, ratios between the absolute differences in CRP levels for the Asp358Ala variant and for IL-6R inhibitory therapy were obtained. Estimates of the absolute difference in CRP between IL-6R inhibitory therapy (tocilizumab 4 or 8 mg/kg) and placebo was obtained from a published meta-analysis[88] of randomized controlled trials.

Estimates of the absolute difference in CRP per copy of the 358Ala allele were obtained on the basis of previously-published genetic association studies[160]. Because CRP was $\log_e$-transformed in these studies, the reported estimate reflected the percentage rather than the absolute difference (i.e., 7.5% lower CRP per allele). To obtain an absolute difference comparable to the one reported in the trials (which were conducted in people with autoimmune disease and hence high average CRP levels), the 7.5% difference was applied to the mean CRP level of individuals contributing to the trial estimate (i.e., mean CRP in trial participants, 26 mg/L; absolute difference per copy of 358Ala, -1.9 mg/L). This assumes that the percentage difference in CRP due to the 358Ala allele stays the same at different starting levels of CRP. To assess this, conditional quantile regression (CQR) and subsequent meta-regression were performed in 14,695 individuals from EPIC-Norfolk. Using the CQR model, the association of 358Ala with ln-CRP was estimated at every 5th percentile of the ln-CRP distribution. Standard errors were calculated using bootstrapping with 200 replicates. CQR models were adjusted for age, sex and the first four genetic principal components. Meta-regression across the CQR estimates per quantile was performed to estimate the difference in ln-CRP per copy of 358Ala across the quantiles of ln-CRP. The meta-regression model was adjusted using the same covariates as the CQR. This meta-regression model did not show evidence that the variant association estimates vary at different ln-CRP levels (P=0.81; **Figure S4.2**). Therefore, ratios between the absolute differences in CRP levels for the 358Ala allele and IL-6R inhibitory therapy were estimated (i.e. 4.7 for tocilizumab 4 mg/kg / 358Ala and 10.9 for tocilizumab 8 mg/kg / 358Ala). These ratios were used to re-scale the estimate of association with type 2 diabetes from genetic association studies of Asp358Ala while accounting for the larger absolute effect of IL-6R inhibitory therapy on CRP. This re-scaling makes the critical assumption that the effects of IL-6R inhibition on diabetes risk estimated by a small difference due to the genetic variant will scale linearly even for large (~5-10 fold) differences between variant and treatment effects.

This estimate attempts to correct for the different effect size of genetic variant vs drug but still does not correct for the duration of exposure. To account for that, it was assumed that the rescaled estimate represented the projected effect of the drug on diabetes risk for a lifelong-exposure (or a long-term exposure of several decades as genetic association studies used for these estimates were conducted in people with an average age of ~59 years). Projections for shorter exposure were obtained by re-scaling the life-long estimate for increasing numbers representing the ratio of exposure between lifelong and a given shorter exposure. Estimates are reported for ratios ranging from 1 to 10, to capture a range of hypothetical durations of treatment. This projection assumes a constant additive risk of the IL-6 pathway on type 2 diabetes risk, which is in line with assumptions of genetic association studies of common variants and previously used Mendelian randomization approaches in this and other settings. These projections are relative to primary occurrence of type 2 diabetes in a general population setting and do not reflect the likely efficacy of treatment in people with high diabetes risk or with inflammatory conditions (e.g. people with rheumatoid arthritis).

Methods S7.1. Fenland glycaemic traits GWAS

Fasting glucose and fasting insulin in Fenland[244] (**Table 7.1**) were the values of circulating glucose (in mmol/L), insulin (natural log-transformed and expressed in log-pmol/L) measured in whole blood after overnight fasting. Two-hour glucose was the value of glucose (in mmol/L) measured in plasma two-hours after a 75-gram oral glucose challenge (**Table 7.1**). Glucose levels were quantified using the Dimension RxL Integrated Chemistry System (Siemens, Germany). Insulin levels were quantified using the 1235 AutoDELFIA automatic immunoassay system using a two-step time resolved fluorometric assay (Kit No. B080-101, Perkin Elmer, USA). Individuals were excluded if they had prevalent type 1 or type 2 diabetes (defined by physician diagnosis); reported use of diabetes medication(s); or had fasting glucose levels >=7 mmol/L, 2-hr glucose levels >=11.1mmol/L, or HbA1c >= 6.5%. A total of 8,729, 7,428 and 8,619 participants were included in the analysis for fasting glucose, fasting insulin and 2-hour glucose respectively.

Fenland[244] genotyping and imputation were performed as described in the main text. Only samples genotyped using the Affymetrix UK Biobank Axiom Array were included. Samples with low call rates < 95%, extreme heterozygosity, gender mismatch with X chromosome variants, duplicates, first- or second-degree relatives or outlying ethnic ancestry were removed prior to association testing. Following sample QC, we applied variant QC thresholds for call rate (< 95%), Hardy-Weinberg Equilibrium (HWE) $P<1x10^{-6}$, imputation quality < 0.4 and minor allele frequency (MAF) < 1%. Association testing for 2-hour glucose was performed under an additive model adjusting for age, sex, BMI and the first 10 principal components using SNPTEST v2.4.1. Association testing for fasting glucose and insulin was performed in the same manner while including age^2 as an additional covariate. Following this, additional variants were excluded if they were tri-allelic; had a minor allele count (MAC)<3; demonstrated a standard error ≥10; or were missing an effect estimate, standard error, or imputation quality.

Methods S7.2. Pairwise colocalisation of GIP levels with cardiometabolic traits and disease endpoints

The SOMAscan® 4K system (SomaLogic, Boulder, Colorado, USA) used two SOMAmers to quantify GIP levels, namely X16292_288 and X5755_29, that targeted different portions of the GIP precursor protein. As GIP levels measured by SOMAmers represented relative abundances, we aimed to assess whether the underlying genetics at *GIPR* were comparable to those of absolute quantitation Methods S7.uch as ELISA. In addition, we aimed to determine which of the SOMAmers represented the best fasting GIP measure available in our cohort.

To address this, we employed a pairwise Bayesian genetic colocalisation framework to estimate whether the associations with cardiometabolic traits at the *GIPR* locus were shared between the two GIP SOMAmers and the Almgren *et al.* 2017[508] GIP measures. As an association with 2-hr GIP levels was also found at the *GIPR* locus, this trait was also included. To contextualise the analysis, we included T2D, CHD, BMI, 2-hour glucose adjusted for BMI (2-hr glucose adjBMI) and LDL as cardiometabolic traits of interest, the data sources of each are described in (**Table S7.1**). Summary statistics from the Fenland 2-hr glucose adjBMI GWAS were preferred to those from previous efforts[344], as outlined in the main text. The details of this GWAS are described in **Methods S7.1**. Using GWAS summary statistics for each trait, the 1Mb regions either side of E354 were extracted. Insertions and deletions as well as any variants with a standard error of 0 were removed from the analysis and only variants present in all datasets were considered. All traits were then aligned to the GIP-raising alleles for consistency. Bayesian colocalisation analysis was then conducted using the COLOC[290] R package between each pair of traits using beta estimates and corresponding trait variances as well as either the case proportion for case-control traits or trait standard deviations for quantitative traits. This was done to estimate posterior probabilities (PP_{coloc}) denoting evidence of colocalisation: H0 – no signal; H1 – signal unique to trait 1; H2 – signal unique to trait 2; H3 – two independent causal variants in the same locus driving the association signal for the respective traits and H4 – presence of a causal variant shared between two traits. The prior

probability that a variant is associated with trait 1, p1, was set to $1x10^{-4}$, hence the prior probability that any variant is associated with trait 1 is 1 in 10,000. The same settings were used for p2. We assigned a prior probability of $1x10^{-5}$ for p12, which equates to a prior probability of 1 in 100,000 that a single variant is associated with both traits. T2D and CHD were treated as case-control traits and all other traits as quantitative. Pairwise PP_{coloc} estimates were considered significant if they met the following criteria: (H4 + H3 ≥ 0.9 & H4/H3 ≥ 3. All data analysis was performed using R version 3.6.3.

Methods S7.3. Partial correlations between X-12283 and known metabolites

To attempt to find an indication of the metabolite class and putative functional pathway of X-12283, we estimated partial correlations between X-12283 levels and the levels of other metabolites measured using the Metabolon platform among EPIC-Norfolk participants. Briefly, a partial correlation estimates the pairwise correlation between two metabolites A and B, while adjusting this estimate for all other metabolites correlated with each metabolite of interest. We performed this analysis using measures from 11,966 individuals, from the two measurement sets of each approximately 6000 quasi-randomly selected individuals. In order to retain sufficient statistical power, we considered only the 883 metabolites with less than 50% missingness within both of the two measurement sets.

First, missing metabolite measures were imputed within each measurement set, using multivariate imputation by chained equations (MICE) [582] with the R package "mice"[583] version 3.6.0. Prior to imputation, metabolite levels were natural log-transformed, winsorised to five SDs, and residuals taken by regressing out the effects of age and sex. Imputation was repeated a total of 20 times, generating 20 sets of fully imputed results. Following imputation, measures were standardised (mean = 0, SD =1). For each imputation, partial correlations between metabolite pairs were calculated using the R package "GeneNet"[584–586] version 1.2.14. Partial correlation estimates were transformed using Fisher's z transformation and the R package "psych"[587] version 1.9.12.31, and then pooled across the 20 imputations for each measurement set, using Rubin's rules[588]. Estimates for the two measurement sets were then meta-analysed, using a fixed effect, inverse variance weighted method and R package "meta"[589] version 4.12-0, and finally back transformed to correlation estimates.

Partial correlation estimates of more than 0.1 were then used to draw a gaussian graphical model (GGM) to visualise the correlation network, using X-12283 as the central node. Partial correlations were considered significant at a Bonferroni significance threshold of $P \leq 1.28x10^{-7}$, accounting for the 389,403 metabolite pairs tested. All data analysis was performed using R version 3.6.3, GGM plots were visualised using Cytoscape version 3.2.1.

Supplementary Tables

Table S1.1: The search strategy used for the systematic literature review. The search was limited to: Clinical trials, journal articles, meta-analyses in English on PubMed in human populations without any disease after June 2007 = First large-scale GWAS (WTCCC)

Search strategy used:

1. Cytokine* AND genome-wide association
2. Cytokine* AND GWAS
3. Cytokine* AND Varia*
4. Cytokine* AND SNP
5. Cytokine* AND Association
6. Cytokine* AND Associat*
7. Inflammat* AND genome-wide association
8. Inflammat* AND GWAS
9. Inflammat* AND Varia*
10. Inflammat* AND SNP
11. Inflammat* AND Association
12. Inflammat* AND GWAS
13. Inflammat* AND Associat*
14. Interleukin* AND genome-wide association
15. Interleukin* AND GWAS
16. Interleukin* AND Varia*
17. Interleukin* AND SNP
18. Interleukin* AND Association
19. Interleukin* AND GWAS
20. Interleukin* AND Associat*
21. IL* AND genome-wide association
22. IL* AND GWAS
23. IL* AND Varia*
24. IL* AND SNP
25. IL* AND Association
26. IL* AND GWAS
27. IL* AND Associat*
28. Interferon* AND genome-wide association
29. Interferon* AND GWAS
30. Interferon* AND Varia*
31. Interferon* AND SNP
32. Interferon* AND Association
33. Interferon* AND GWAS
34. Interferon* AND Associat*
35. IFN* AND genome-wide association
36. IFN* AND GWAS
37. IFN* AND Varia*
38. IFN* AND SNP
39. IFN* AND Association
40. IFN* AND GWAS
41. IFN* AND Associat*
42. Tumour Necrosis Factor* AND genome-wide association
43. Tumour Necrosis Factor* AND GWAS
44. Tumour Necrosis Factor* AND Varia*
45. Tumour Necrosis Factor* AND SNP
46. Tumour Necrosis Factor* AND Association
47. Tumour Necrosis Factor* AND GWAS
48. Tumour Necrosis Factor* AND Associat*
49. TNF* AND genome-wide association
50. TNF* AND GWAS
51. TNF* AND Varia*
52. TNF* AND SNP
53. TNF* AND Association
54. TNF* AND GWAS
55. TNF* AND Associat*

Table S3.1: Number of variants removed from each cytokine meta-analysis at each quality control step.

QC filter	IFNγ	IL-6	IL-8	TNFα
Number of SNPs before QC	20,889,287	21,005,838	20,792,591	20,991,893
Drop if less than 50% sample size genotyped	2,861,949	2,886,839	2,728,592	2,868,406
Drop if number studies < 2 per SNP	0	0	0	0
MAC <20	0	0	0	0
Indels	1,791,026	1,796,779	1,793,926	1,797,320
Tri-allelic SNPs	22,290	22,472	22,370	22,520
Frequency SE > 0.1	9	9	9	9
Drop if minfreq[a] <0.05 & maxfreq[b] >0.5	0	0	0	0
Drop if maxfreq >0.95 & minfreq <0.5	0	0	0	0
Number of SNPs After QC	16,214,013	16,229,739	16,247,694	16,303,638
Number of SNPs at p<5E-08[c]	153	131	1004	710

Abbreviations: QC, Quality control; IFNγ, Interferon-γ; IL-6, Interleukin-6; IL-8, Interleukin-8; TNFα, Tumour necrosis factor-α; MAC, Minor allele count; Indels, Insertions and deletions; SE, Standard error; minfreq, Minimum frequency; maxfreq, Maximum frequency

a. Minimum frequency observed for a particular SNP across the 4 studies sub-studies included in each meta-analysis
b. Maximum frequency observed for a particular SNP across the 4 studies sub-studies included in each meta-analysis
c. The total number of genome-wide significant SNPs for each meta-analysis

Table S4.1: Summary of the studies participating in the different analyses of the manuscript.

Study stage [a]	Outcome	Cases overall, N	Non-cases (for case-control studies) or participants (for continuous trait studies) overall, N	Participating study	PubMed ID for cohort description	Website (URL)
1	Type 2 diabetes	5,421	31,562	15 prospective studies	10466767; 23264288; 23130155; 23251619; 28834086	http://www.srl.cam.ac.uk/epic/
	2-hour glucose	-	10,218	Fenland	27841877	http://www.mrc-epid.cam.ac.uk/research/studies/fenland/
	Fasting glucose	-	10,344			
	HbA1c	-	10,331			
	Fasting insulin	-	8,428			
	DEXA traits	-	10,344			
	Iron	-	8,980			
	Transferrin	-	8,988			
	Ferritin	-	8,958			
2	Type 2 diabetes	22,182	424,361	UK Biobank	25826379	http://www.ukbiobank.ac.uk/
		109,191	223,608	Million Veteran Program	26441289	https://www.research.va.gov/mvp/
		55,005	400,308	DIAMANTE	30297969	N/A
		36,614	155,150	Suzuki *et al.*	30718926	http://jenger.riken.jp/en/result
		11,006	82,655	Finngen	NA	https://www.finngen.fi/en
		26,616	64,558	Geisinger	26866580	https://www.geisinger.org/mycode
	Type 1 diabetes	12,079	12,262	T1DGC	17130525	https://repository.niddk.nih.gov/studies/t1dgc/
		11,182	321,617	Million Veteran Program	26441289	https://www.research.va.gov/mvp/
		948	424,361	UK Biobank	25826379	http://www.ukbiobank.ac.uk/
	Coronary heart disease	24,890	427,309			

Study stage [a]	Outcome	Cases overall, N	Non-cases (for case-control studies) or participants (for continuous trait studies) overall, N	Participating study	PubMed ID for cohort description	Website (URL)
	Fasting glucose	-	133,010	MAGIC	22885924; 22581228	https://www.magicinvestigators.org/
	2-hour glucose	-	42,854			
	Fasting insulin	-	108,557			
	HbA1c	-	479,942	MAGIC; UK Biobank	28898252; 25826379	https://www.magicinvestigators.org/; http://www.ukbiobank.ac.uk
	Non-fasted glucose	-	355,817	UK Biobank	25826379	http://www.ukbiobank.ac.uk
	Body mass index	-	772,066	GIANT; UK Biobank	25673413; 25826379	https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium; http://www.ukbiobank.ac.uk
	Hip circumference	-	604,143			
	Waist circumference	-	615,305			
	Waist-to-hip ratio adjusted for BMI	-	625,123			
	Waist-to-hip ratio unadjusted	-	602,940			
3 [b]	Type 2 diabetes (observational association)	411	7,010	EPIC-Norfolk	10466767	http://www.srl.cam.ac.uk/epic/
	Coronary heart disease (observational association)	1,123	5,917			
	Type 2 diabetes (genetic association)	17,006	330,572	UK Biobank	25826379	http://www.ukbiobank.ac.uk/
	Coronary heart disease (genetic association)	18,770	333,247			

Abbreviations: N, Number of participants; HbA1c; Glycated haemoglobin; BMI; Body mass index; EPIC, European prospective investigation of cancer; DIAMANTE, Diabetes meta-analysis of trans-ethnic association studies; MAGIC, Meta-Analyses of glucose and insulin-related traits consortium; T1DGC, Type 1 diabetes genetics consortium.

a. Traits are continuous unless otherwise stated as a disease or a polygenic score
b. EPIC-Norfolk contributed to both the observational and genetic mediation analyses whereas UK Biobank contributed to only the genetic mediation analysis

Table S4.2: Study characteristics and adjusted risk ratios from studies of incident type 2 diabetes associated with levels of IL-6.

Study, year (country)	Cases, N	Non-cases, N	Proportion of women (%)	Study design	Duration (years of follow-up)	Type 2 diabetes ascertainment	Exposure measurement method	Risk Ratio (95% CI)	Covariates	Source
Pradhan et al. 2001 (US)	188	362	100	Nested case-control	4	Self-report based on ADA criteria and verified by primary care records	ELISA (R&D Systems, Minneapolis, MN)	1.70 (0.61-4.74)	Age, FH of diabetes, smoking, physical activity, alcohol, PMT, fasting status, BMI, and FI	a
Duncan et al. 2003 (US)	581	572	60	Case cohort	9	1. Reported physician diagnosis; 2. use of anti-diabetes medications; 3. FG ≥7.0 mmol/L; 4. non-fasting glucose of ≥11.1 mmol/L	ELISA (R&D Systems, Minneapolis, MN)	1.59 (1.02-2.48)	Age, centre, sex, ethnicity, FH of diabetes, hypertension, BMI, WHR, FG, and FI	a
Krakoff et al. 2003 (US)	71	71	66	Nested case-control	4.6	WHO criteria	ELISA (Quantikine High Sensitivity; R&D Systems, Oxon, U.K.)	0.75 (0.45-1.25)	Age, WC, FPG, PG2h, HbA1c, and FI	a
Spranger et al. 2003 (Germany)	188	377	21	Nested case-control	2.3	Self-report with cases validated using a questionnaire completed by the primary care physician	ELISA (R&D Systems, Minneapolis, MN)	2.20 (1.31-3.69)	Age, sex, sporting activities, smoking, alcohol, educational attainment, BMI, WHR, and HbA1c	a
Hu et al. 2004 (US)	737	785	100	Nested case-control	10	Unconfirmed self-report of one of: 1. Classic symptoms plus elevated glucose levels; 2. two elevated plasma glucose if no symptoms; 3. treatment with agents or insulin, ADA criteria for cases after 1998	ELISA (Quantikine HS Immunoassay kit)	1.52 (1.18-1.96)	Age, race, time at blood drawn, alcohol, physical activity, smoking, FH of diabetes, PMT, MP, diet score, fasting status, and BMI	a
Liu et al. 2007 (US)	1584	4317	100	Nested case-control	5.9	Self-report	ELISA (R&D Systems)	1.19 (1.02-1.39)	Age, race, clinic, time of blood draw, BMI, alcohol, physical activity, smoking, PMT, FH of diabetes, and FG	a
Thorand et al. 2007 F (Germany)	222	809	100	Case cohort	10.8	Self-report with cases validated using a questionnaire mailed to the treating physician or by medical chart review	High-sensitivity latex enhanced nephelometric assay on a BN II analyser (Dade-Behring, Marburg, Germany)	1.65 (1.16-2.35)	Age, survey, smoking, alcohol, physical activity, SBP, total-to-HDL cholesterol ratio, BMI, and FH of diabetes	a
Thorand et al. 2007 M (Germany)	305	889	0	Case cohort	10.8	Self-report with cases validated using a questionnaire mailed to the treating physician or by medical chart review	High sensitivity immunoradiometric assay (IRMA) (men aged 45–64 years) or a high-sensitivity latex enhanced nephelometric assay on a BN II analyser (men aged 35–44 years) (Dade-	1.34 (1.06-1.69)	Age, survey, smoking, alcohol, physical activity, SBP, total-to-HDL cholesterol ratio, BMI, and FH of diabetes	a

Study, year (country)	Cases, N	Non-cases, N	Proportion of women (%)	Study design	Duration (years of follow-up)	Type 2 diabetes ascertainment	Exposure measurement method	Risk Ratio (95% CI)	Covariates	Source
							Behring, Marburg, Germany)			
Wannamethee et al. 2007 (UK)	108	3599	0	Cohort	5	Self-report confirmed via primary care records	ELISA (R&D Systems)	1.26 (1.05-1.51)	Age, social class, physical activity, smoking, alcohol, use of statins, CHD or stroke, SBP, treatment for hypertension, BMI, and HOMA-IR	a
Ley et al. 2008 (Canada)	86	406	58	Cohort	10	1. FG ≥7.0 mmol/L or PG2h ≥11.1 mmol/L; 2. current use of insulin or oral hypoglycaemic agents; 3. self-report (not verified from records)	ELISA (BioSource International, Camarillo, CA)	0.98 (0.74-1.30)	Age, sex, TG, HDL-C, hypertension, WC, and IGT	a
Bertoni et al. 2010 (US)	410	5161	53	Cohort	4.7	Using hypoglycaemic medication or FG ≥7 mmol/L	ELISA (Quantikine HS Human IL-6 Immunoassay; R&D Systems)	1.30 (1.09-1.55)	Age, sex, race, education, site, alcohol, smoking, exercise, SBP, antihypertensive medication, HOMA-IR, and BMI	a
Dallmeier et al., 2012 (USA)	162	2476	56	Cohort	6.6	Fasting glucose level ≥126 mg/dL or use of insulin or oral hypoglycaemic medications at follow-up	ELISA (R&D Systems, Minneapolis, MN)	1.09 (0.84-1.40)	Age, sex, cohort, BMI, FG, SBP, HDL-C, TG, and smoking	b
Marques-Vidal et al., 2012 (Switzerland)	208	3634	56.7	Cohort	5.5	Fasting plasma glucose ≥7.0 mmol/L and/or presence of oral hypoglycaemic or insulin treatment	Multiplexed particle-based flow cytometric cytokine assay	1.06 (0.85-1.32)	Age, FH of type 2 diabetes, height, WC, resting heart rate, presence of hypertension, HDL-C, TG, FG and serum uric acid	b
Koloverou et al., 2018 (Greece)	160	1094	51	Cohort	10	Unconfirmed self-report or ADA criteria i.e. use of antidiabetic medication (hypoglycaemic drugs and/or insulin use) or fasting plasma glucose ≥126 mg/dL	ELISA	1.08 (0.19-6.09)	Age, sex, FH of diabetes, smoking, adherence to the Mediterranean diet, physical activity, hypertension, hypercholesterolaemia status and BMI	b
EPIC-Norfolk, (UK)	411	7010	61	Cohort	16.7	1. WHO criteria (HbA1c >6.5% (48mmol/mol)) 2. Electronic health records (ICD 10: E11) 3. Self-report	Electrochemiluminescence (Meso Scale Discovery, US)	1.25 (1.10-1.41)	Age, sex, BMI, WHR, ethnicity, education level, FH of type 2 diabetes, smoking status, average units of alcohol per week and average physical activity per week	c

Abbreviations: ADA, American Diabetes Association; BMI, body mass index; CHD, coronary heart disease; FG, fasting glucose; FH, family history; FI, fasting insulin; FPG, fasting plasma glucose; HDL-C, HDL cholesterol; IGT, impaired glucose tolerance; MP, menopausal; PG2h, 2-h plasma glucose; PMT, postmenopausal with hormone replacement therapy; SBP, systolic blood pressure; TG, triglycerides; WHO, World Health Organization; WHR, waist-to-hip ratio.

Sources:

a. Wang et al., 2013[186]
b. Systematic search
c. This study

Table S4.3: Systematic literature review search terms. The following terms were searched in the PubMed database to identify prospective studies that have investigated the association between IL-6 levels and incident type 2 diabetes.

	Search Terms
#5	#1 AND #2 AND #3 AND #4
#4	"2012/02/10"[Date - Publication]: "2018/10/30"[Date - Publication]
#3	"Cohort Studies"[MeSH] OR Cohort Study OR Studies, Cohort OR Study, Cohort OR Studies, Historical Cohort OR Cohort Studies, Historical Cohort Study, Historical OR Historical Cohort Study OR Study, Historical OR Analysis, Cohort OR Analyses, Cohort OR Cohort Analyses OR Cohort Analysis OR Incidence Studies OR Incidence Study OR Studies, Incidence OR Study, Incidence OR "Incidence"[MeSH] OR "Cohort Studies"[MeSH] OR Case-cohort study OR Nested case-control study OR Survey
#2	Diabetes Mellitus, Type 2[MeSH] OR NIDDM OR Maturity-Onset Diabetes OR Diabetes Mellitus, Noninsulin-Dependent OR Diabetes Mellitus, Adult-Onset OR Adult-Onset Diabetes Mellitus OR Diabetes Mellitus, Adult Onset OR Diabetes Mellitus, Ketosis-Resistant OR Diabetes Mellitus, Ketosis Resistant OR Ketosis-Resistant Diabetes Mellitus OR Diabetes Mellitus, Maturity-Onset OR Diabetes Mellitus, Maturity Onset OR Diabetes Mellitus, Non-Insulin Dependent OR Diabetes Mellitus, Non-Insulin-Dependent OR Non-Insulin-Dependent Diabetes Mellitus OR Diabetes Mellitus, Noninsulin Dependent OR Diabetes Mellitus, Slow-Onset OR Diabetes Mellitus, Slow Onset OR Slow-Onset Diabetes Mellitus OR Diabetes Mellitus, Stable OR Stable Diabetes Mellitus OR Diabetes Mellitus, Type II OR Maturity-Onset Diabetes Mellitus OR Maturity Onset Diabetes Mellitus OR MODY OR Type 2 Diabetes Mellitus OR Noninsulin-Dependent Diabetes Mellitus OR T2D OR T2DM OR Type 2 Diabetes[tiab] OR Type 2 diabetes mellitus OR diabetes[ti]
#1	Interleukin-6[MeSH] OR Interleukin 6 OR IL6 OR B-Cell Stimulatory Factor 2 OR B-Cell Stimulatory Factor-2 OR Differentiation Factor-2, B-Cell OR Differentiation Factor 2, B Cell OR B-Cell Differentiation Factor-2 OR B Cell Differentiation Factor 2 OR BSF-2 OR Hybridoma Growth Factor OR Growth Factor, Hybridoma OR IFN-beta 2 OR Plasmacytoma Growth Factor OR Growth Factor, Plasmacytoma OR Hepatocyte-Stimulating Factor OR Hepatocyte Stimulating Factor OR MGI-2 OR Myeloid Differentiation-Inducing Protein OR Differentiation-Inducing Protein, Myeloid OR Myeloid Differentiation Inducing Protein OR B-Cell Differentiation Factor OR B Cell Differentiation Factor OR Differentiation Factor, B-Cell OR Differentiation Factor, B Cell OR IL-6 OR Interferon beta-2 OR Interferon beta 2 OR beta-2, Interferon OR B Cell Stimulatory Factor-2 OR B Cell Stimulatory Factor 2

Table S4.4: Quality ratings of the studies included in the meta-analysis of association between IL-6 levels and incident type 2 diabetes.

Study, year	Size and representativeness	Reliably measured exposure	Ascertainment of diabetes	Adjustment for possible confounders	Overall quality score
Pradhan *et al.* 2001	1	1	1	2	5
Duncan *et al.* 2003	1	1	1	1	4
Krakoff *et al.* 2003	0	1	1	1	3
Spranger *et al.* 2003	1	1	1	1	4
Hu *et al.* 2004	1	1	0	2	4
Liu *et al.* 2007	1	1	0	1	3
Thorand *et al.* 2007 M [a]	1	0	1	2	4
Thorand *et al.* 2007 F [a]	1	0	1	2	4
Wannamethee *et al.* 2007	1	1	1	1	4
Ley *et al.* 2008	0	1	1	1	3
Bertoni *et al.* 2010	1	1	1	1	4
Dallmeier *et al.* 2012	1	1	1	1	4
Marques-Vidal *et al.* 2012	1	1	1	1	4
Koloverou *et al.* 2018	1	1	1	2	5
EPIC-Norfolk, 2018	1	1	2	2	6

Abbreviations: M, Males; F, Females; EPIC, European prospective investigation of cancer

a. IL-6 levels were measured differently between males aged 45-64 years compared to males aged 35-44 years and all females in the study therefore estimates were analysed separately.

Each study received up to 6 points based on the criteria below:

1. Size and representativeness: the study sample was representative of participants (i.e. sample size was more than 500, or more than 80% of eligible participants were invited, or more than 80% agreed to participate). [1 point]
2. Reliably measured exposure (ELISA = 1, Electrochemiluminescent assay = 1, particle-based flow cytometry = 1, light-scatter based methods other than flow cytometry = 0). [1 point]
3. The study reliably assessed outcome (Only self-reported = 0, objective glycaemic measures = +1 point to score, use of health records = +1 point to score). [2 points]
4. Adjustment for possible confounders (Adjusted only for age and sex = 0, Adjusted for age, sex and self-reported risk factors = 1, additionally adjusted for anthropometry = 2, adjusted for glycaemic measures = -1 point from score) [2 points]

The arbitrary categorization of overall quality score was low (0 to 2 points), good (3 or 4 points), or high (5 or 6 points).

Table S4.5: Fenland participant DEXA and glycemic measures.

Variable	Value
Country	United Kingdom
Number of participants	10,344
Baseline age, mean years (SD)	48 (8)
Female sex, N (%)	5,528 (53)
Arms fat mass, median (IQR) in grams	2,516 (2,046-3,099)
Trunk fat mass, median (IQR) in grams	13,142 (9,570-17,268)
Abdominal fat mass total, median (IQR) in grams	2,145 (1,435-2,991)
Subcutaneous abdominal fat mass, median (IQR) in grams	1,239 (921-1,667)
Visceral abdominal fat mass, median (IQR) in grams	766 (317-1,409)
Gluteofemoral fat mass, median (IQR) in grams	4,124 (3,298-5,174)
Leg fat mass, median (IQR) in grams	8,285 (6,652-10,446)
Fasting glucose, mean (SD) in mmol/L	4.83 (0.66)
2-hour glucose, mean (SD) in mmol/L	5.29 (1.70)
HbA1c, mean (SD) in %	5.54 (0.46)
Fasting insulin, median (IQR) in pmol/L	38.3 (26.5-57.2)

Abbreviations: SD, Standard deviation; N, Number of participants; IQR, Interquartile range; mmol/L, millimoles per Litre; pmol/L, picomoles per Litre; HbA1c, glycated haemoglobin

Table S4.6: Standard deviation values used to convert estimates for continuous metabolic traits between clinical and standardized units and their source.

Metabolic trait	Clinical units	Standard deviation	Source study
2-hour glucose	mmol/L	1.7	Fenland
Fasting plasma glucose	mmol/L	0.66	Fenland
HbA1c	%	0.49	Fenland
Fasting insulin	Log (pmol/L)	0.60	Fenland
Non-fasted glucose	mmol/L	0.87	UK Biobank
BMI	kg/m2	4.8	UK Biobank
Hip circumference	cm	9.2	UK Biobank
Waist circumference	cm	13.5	UK Biobank
WHRadjBMI	ratio	0.056	UK Biobank
WHR	ratio	0.09	UK Biobank

Abbreviations: BMI, Body mass index; WHR, Waist-to-hip ratio; WHRadjBMI, WHR adjusted for BMI; mmol, millimoles; L, Litre; pmol, picomoles; Kg, Kilograms; m, meters; cm, centimetres
Variables with a skewed distribution were normalized using the natural log transformation.

Table S4.7: Polygenic risk scores used in Stage 2 and their sources.

Phenotype	SNP ID	EA [a]	Beta	PMID
BMI	rs543874	A	0.205	
BMI	rs3101336	T	0.395	
BMI	rs12566985	G	0.450	
BMI	rs17024393	T	0.026	
BMI	rs657452	A	0.399	
BMI	rs11165643	C	0.419	
BMI	rs12401738	G	0.366	
BMI	rs2820292	A	0.441	
BMI	rs11583200	C	0.399	
BMI	rs977747	T	0.426	
BMI	rs7903146	C	0.290	
BMI	rs17094222	T	0.209	
BMI	rs11191560	T	0.081	
BMI	rs7899106	A	0.051	
BMI	rs11030104	A	0.203	
BMI	rs3817334	C	0.402	
BMI	rs12286929	A	0.476	
BMI	rs4256980	C	0.354	
BMI	rs2176598	T	0.248	
BMI	rs7138803	G	0.365	
BMI	rs11057405	G	0.100	
BMI	rs12429545	G	0.130	
BMI	rs9581854	C	0.175	
BMI	rs1441264	G	0.402	
BMI	rs9540493	A	0.461	
BMI	rs7141420	C	0.481	
BMI	rs10132280	C	0.303	
BMI	rs12885454	C	0.352	
BMI	rs11847697	C	0.051	
BMI	rs16951275	T	0.235	
BMI	rs7164727	C	0.343	
BMI	rs3736485	A	0.468	
BMI	rs1558902	T	0.394	
BMI	rs3888190	C	0.390	
BMI	rs12446632	G	0.138	25673413[309]
BMI	rs758747	C	0.287	
BMI	rs9925964	A	0.354	
BMI	rs2650492	G	0.288	
BMI	rs2080454	C	0.390	
BMI	rs4787491	A	0.465	
BMI	rs12940622	G	0.439	
BMI	rs1000940	A	0.305	
BMI	rs9914578	C	0.213	
BMI	rs6567160	T	0.235	
BMI	rs7239883	G	0.377	
BMI	rs7243357	T	0.178	
BMI	rs1808579	C	0.482	
BMI	rs2287019	C	0.181	
BMI	rs3810291	G	0.342	
BMI	rs2075650	A	0.144	
BMI	rs29941	A	0.326	
BMI	rs17724992	A	0.270	
BMI	rs13021737	A	0.169	
BMI	rs10182181	A	0.493	
BMI	rs1016287	T	0.295	
BMI	rs7599312	G	0.266	
BMI	rs11126666	G	0.258	
BMI	rs492400	C	0.429	
BMI	rs2176040	A	0.348	
BMI	rs1528435	C	0.378	
BMI	rs11688816	G	0.457	
BMI	rs2121279	C	0.121	
BMI	rs17203016	A	0.196	
BMI	rs1460676	T	0.159	
BMI	rs6091540	C	0.295	
BMI	rs2836754	T	0.377	
BMI	rs1516725	T	0.138	
BMI	rs13078960	T	0.195	
BMI	rs2365389	C	0.422	

Phenotype	SNP ID	EA [a]	Beta	PMID
BMI	rs16851483	G	0.067	
BMI	rs6804842	A	0.432	
BMI	rs3849570	C	0.350	
BMI	rs10938397	A	0.429	
BMI	rs13107325	C	0.072	
BMI	rs17001654	C	0.148	
BMI	rs11727676	T	0.093	
BMI	rs2112347	T	0.368	
BMI	rs7715256	G	0.430	
BMI	rs2207139	A	0.169	
BMI	rs205262	A	0.275	
BMI	rs13191362	A	0.122	
BMI	rs751414	G	0.283	
BMI	rs9400239	T	0.309	
BMI	rs9374842	C	0.231	
BMI	rs13201877	A	0.130	
BMI	rs9641123	G	0.399	
BMI	rs1167827	A	0.431	
BMI	rs2245368	C	0.176	
BMI	rs6465468	G	0.299	
BMI	rs17405819	T	0.295	
BMI	rs16907751	C	0.099	
BMI	rs2033732	T	0.252	
BMI	rs10968576	A	0.315	
BMI	rs1928295	T	0.430	
BMI	rs4740619	T	0.453	
BMI	rs10733682	A	0.473	
BMI	rs6477694	C	0.359	
WHRadjBMI	rs998584	A	0.05	30575882[30]
WHRadjBMI	rs1936805	T	0.041	
WHRadjBMI	rs7133378	G	0.039	
WHRadjBMI	rs2371767	G	0.04	
WHRadjBMI	rs2791550	G	0.037	
WHRadjBMI	rs10923724	T	0.035	
WHRadjBMI	rs10195252	T	0.031	
WHRadjBMI	rs1294410	C	0.031	
WHRadjBMI	rs718314	G	0.034	
WHRadjBMI	rs3786897	G	0.03	
WHRadjBMI	rs10919388	C	0.033	
WHRadjBMI	rs714515	G	0.028	
WHRadjBMI	rs17451107	T	0.029	
WHRadjBMI	rs2236519	A	0.03	
WHRadjBMI	rs6861681	A	0.029	
WHRadjBMI	rs1443512	A	0.031	
WHRadjBMI	rs9837325	C	0.034	
WHRadjBMI	rs459193	A	0.029	
WHRadjBMI	rs2167750	T	0.027	
WHRadjBMI	rs12608504	A	0.026	
WHRadjBMI	rs2145272	G	0.026	
WHRadjBMI	rs2294239	A	0.025	
WHRadjBMI	rs797486	A	0.037	
WHRadjBMI	rs1055144	T	0.03	
WHRadjBMI	rs605203	A	0.025	
WHRadjBMI	rs12214804	C	0.045	
WHRadjBMI	rs4738141	G	0.028	
WHRadjBMI	rs2276824	C	0.024	
WHRadjBMI	rs17819328	G	0.023	
WHRadjBMI	rs634869	T	0.023	
WHRadjBMI	rs2925979	T	0.024	
WHRadjBMI	rs905938	T	0.025	
WHRadjBMI	rs8066985	A	0.022	
WHRadjBMI	rs10502148	C	0.023	
WHRadjBMI	rs1569135	A	0.021	
WHRadjBMI	rs7492628	G	0.024	
WHRadjBMI	rs6556301	T	0.022	
WHRadjBMI	rs9792666	A	0.057	
WHRadjBMI	rs2845885	C	0.045	
WHRadjBMI	rs3810068	T	0.023	
WHRadjBMI	rs7801581	T	0.023	
WHRadjBMI	rs12936587	G	0.02	

Phenotype	SNP ID	EA [a]	Beta	PMID
WHRadjBMI	rs3747577	C	0.023	
WHRadjBMI	rs757608	A	0.02	
WHRadjBMI	rs143384	A	0.02	
WHRadjBMI	rs951252	G	0.019	
WHRadjBMI	rs2428549	G	0.021	
WHRadjBMI	rs2073267	G	0.025	
WHRadjBMI	rs601339	A	0.025	
WHRadjBMI	rs910382	G	0.019	
WHRadjBMI	rs711869	G	0.019	
WHRadjBMI	rs1051921	G	0.023	
WHRadjBMI	rs11263432	T	0.02	
WHRadjBMI	rs7598832	C	0.019	
WHRadjBMI	rs6719672	G	0.025	
WHRadjBMI	rs780159	G	0.018	
WHRadjBMI	rs6688233	T	0.022	
WHRadjBMI	rs622217	T	0.019	
WHRadjBMI	rs16891532	A	0.034	
WHRadjBMI	rs9644033	A	0.022	
WHRadjBMI	rs10264590	A	0.018	
WHRadjBMI	rs6446204	C	0.02	
WHRadjBMI	rs10462028	A	0.019	
WHRadjBMI	rs8030605	A	0.026	
WHRadjBMI	rs10992408	G	0.024	
WHRadjBMI	rs4779526	A	0.021	
WHRadjBMI	rs4727695	A	0.029	
WHRadjBMI	rs6853254	T	0.019	
WHRadjBMI	rs9388766	C	0.018	
WHRadjBMI	rs6581662	T	0.02	
WHRadjBMI	rs3741378	C	0.024	
WHRadjBMI	rs13256367	A	0.018	
WHRadjBMI	rs2993481	T	0.021	
WHRadjBMI	rs11724804	G	0.017	
WHRadjBMI	rs4420638	A	0.022	
WHRadjBMI	rs1045241	C	0.018	
WHRadjBMI	rs4450871	A	0.018	
WHRadjBMI	rs7589318	G	0.017	
WHRadjBMI	rs2836179	G	0.017	
WHRadjBMI	rs11592754	C	0.023	
WHRadjBMI	rs303084	A	0.019	
WHRadjBMI	rs9844972	C	0.033	
WHRadjBMI	rs2047937	C	0.016	
WHRadjBMI	rs3764002	C	0.018	
WHRadjBMI	rs727428	T	0.016	
WHRadjBMI	rs10980802	G	0.016	
WHRadjBMI	rs11747001	A	0.018	
WHRadjBMI	rs2222543	G	0.016	
WHRadjBMI	rs39312	C	0.016	
WHRadjBMI	rs7235010	A	0.018	
WHRadjBMI	rs9362097	G	0.016	
WHRadjBMI	rs9583489	C	0.017	
WHRadjBMI	rs1805741	C	0.017	
WHRadjBMI	rs2254069	A	0.024	
WHRadjBMI	rs2235529	C	0.021	
WHRadjBMI	rs17041868	C	0.03	
WHRadjBMI	rs13154197	G	0.027	
WHRadjBMI	rs2298632	C	0.015	
WHRadjBMI	rs1876829	C	0.019	
WHRadjBMI	rs1440372	C	0.016	
WHRadjBMI	rs4902632	A	0.021	
WHRadjBMI	rs6433219	A	0.017	
WHRadjBMI	rs7122422	C	0.015	
WHRadjBMI	rs2444770	T	0.021	
WHRadjBMI	rs10512606	C	0.027	
WHRadjBMI	rs11079041	A	0.016	
WHRadjBMI	rs4849294	T	0.015	
WHRadjBMI	rs1316979	T	0.029	
WHRadjBMI	rs4851221	G	0.019	
WHRadjBMI	rs1144	C	0.015	
WHRadjBMI	rs3851294	G	0.025	
WHRadjBMI	rs1105881	G	0.015	

Phenotype	SNP ID	EA [a]	Beta	PMID
WHRadjBMI	rs11051005	A	0.016	
WHRadjBMI	rs6920788	T	0.016	
WHRadjBMI	rs7680787	T	0.014	
WHRadjBMI	rs12459350	A	0.014	
WHRadjBMI	rs4704389	A	0.014	
WHRadjBMI	rs6932767	T	0.017	
WHRadjBMI	rs1474921	A	0.015	
WHRadjBMI	rs380654	G	0.014	
WHRadjBMI	rs8055190	C	0.034	
WHRadjBMI	rs9647379	G	0.014	
WHRadjBMI	rs2398893	A	0.015	
WHRadjBMI	rs2821391	A	0.015	
WHRadjBMI	rs998749	A	0.013	
WHRadjBMI	rs12828318	A	0.018	
WHRadjBMI	rs10844642	A	0.014	
WHRadjBMI	rs1328757	T	0.014	
WHRadjBMI	rs1053593	G	0.014	
WHRadjBMI	rs1498126	C	0.017	
WHRadjBMI	rs1190982	T	0.015	
WHRadjBMI	rs12774134	C	0.02	
WHRadjBMI	rs11726981	C	0.015	
WHRadjBMI	rs7800072	G	0.014	
WHRadjBMI	rs10880321	G	0.014	
WHRadjBMI	rs7823561	A	0.014	
WHRadjBMI	rs4751628	G	0.014	
WHRadjBMI	rs7242873	G	0.024	
WHRadjBMI	rs2061705	G	0.013	
WHRadjBMI	rs17326656	T	0.015	
WHRadjBMI	rs13406302	C	0.014	
WHRadjBMI	rs2452877	A	0.013	
WHRadjBMI	rs7114403	A	0.013	
WHRadjBMI	rs2283847	T	0.013	
WHRadjBMI	rs11187537	C	0.015	
WHRadjBMI	rs6752964	C	0.02	
WHRadjBMI	rs3789615	C	0.012	
WHRadjBMI	rs11893688	T	0.014	
WHRadjBMI	rs332105	G	0.013	
WHRadjBMI	rs8030277	T	0.013	
WHRadjBMI	rs10963067	C	0.022	
WHRadjBMI	rs174829	G	0.014	
WHRadjBMI	rs2057869	A	0.014	
WHRadjBMI	rs421168	G	0.013	
WHRadjBMI	rs10745659	G	0.012	
WHRadjBMI	rs7612999	A	0.014	
WHRadjBMI	rs6449133	T	0.013	
WHRadjBMI	rs2595004	C	0.016	
WHRadjBMI	rs7919055	C	0.032	
WHRadjBMI	rs2066107	T	0.013	
WHRadjBMI	rs2058914	G	0.013	
WHRadjBMI	rs2333496	T	0.012	
WHRadjBMI	rs10887759	A	0.016	
WHRadjBMI	rs12684047	T	0.016	
WHRadjBMI	rs4239275	T	0.012	
WHRadjBMI	rs6486060	G	0.012	
WHRadjBMI	rs2320125	T	0.012	
WHRadjBMI	rs807067	T	0.011	
WHRadjBMI	rs17167945	G	0.015	
WHRadjBMI	rs36232	G	0.015	
WHRadjBMI	rs9750952	C	0.014	
WHRadjBMI	rs710122	G	0.012	
WHRadjBMI	rs9896963	C	0.014	
WHRadjBMI	rs9305545	G	0.016	
WHRadjBMI	rs747249	A	0.012	
WHRadjBMI	rs6874524	T	0.013	
WHRadjBMI	rs676556	G	0.014	
WHRadjBMI	rs6908042	A	0.011	
WHRadjBMI	rs2701523	A	0.012	
WHRadjBMI	rs1360485	T	0.011	
WHRadjBMI	rs544668	T	0.011	
WHRadjBMI	rs7091853	C	0.011	

Phenotype	SNP ID	EA [a]	Beta	PMID
WHRadjBMI	rs1979527	A	0.013	
WHRadjBMI	rs1278493	G	0.011	
WHRadjBMI	rs12631066	C	0.013	
WHRadjBMI	rs2823096	A	0.014	
WHRadjBMI	rs10891483	T	0.017	
WHRadjBMI	rs362275	C	0.011	
WHRadjBMI	rs6496127	G	0.01	
WHRadjBMI	rs7235891	C	0.011	
WHRadjBMI	rs1156979	A	0.011	
WHRadjBMI	rs505102	C	0.013	
WHRadjBMI	rs10507524	C	0.018	
WHRadjBMI	rs13028903	T	0.011	
WHRadjBMI	rs12440695	C	0.01	
WHRadjBMI	rs2590440	G	0.013	
WHRadjBMI	rs2240328	T	0.014	
WHRadjBMI	rs3758938	T	0.011	
WHRadjBMI	rs4454603	C	0.01	
WHRadjBMI	rs1328369	T	0.01	
WHRadjBMI	rs12186798	G	0.013	
WHRadjBMI	rs15285	C	0.011	
Hip-specific score for higher WHR	rs10195252	T	0.031	
Hip-specific score for higher WHR	rs9837325	C	0.034	
Hip-specific score for higher WHR	rs2167750	T	0.027	
Hip-specific score for higher WHR	rs2145272	G	0.026	
Hip-specific score for higher WHR	rs605203	A	0.025	
Hip-specific score for higher WHR	rs2845885	C	0.045	
Hip-specific score for higher WHR	rs12936587	G	0.02	
Hip-specific score for higher WHR	rs3747577	C	0.023	
Hip-specific score for higher WHR	rs601339	A	0.025	
Hip-specific score for higher WHR	rs10992408	G	0.024	
Hip-specific score for higher WHR	rs4450871	A	0.018	
Hip-specific score for higher WHR	rs11592754	C	0.023	
Hip-specific score for higher WHR	rs727428	T	0.016	
Hip-specific score for higher WHR	rs2235529	C	0.021	
Hip-specific score for higher WHR	rs13154197	G	0.027	
Hip-specific score for higher WHR	rs7122422	C	0.015	
Hip-specific score for higher WHR	rs1316979	T	0.029	
Hip-specific score for higher WHR	rs6920788	T	0.016	
Hip-specific score for higher WHR	rs12459350	A	0.014	
Hip-specific score for higher WHR	rs4704389	A	0.014	
Hip-specific score for higher WHR	rs2061705	G	0.013	
Hip-specific score for higher WHR	rs2320125	T	0.012	
Waist-specific score for higher WHR	rs17451107	T	0.029	
Waist-specific score for higher WHR	rs1055144	T	0.03	
Waist-specific score for higher WHR	rs1569135	A	0.021	
Waist-specific score for higher WHR	rs757608	A	0.02	
Waist-specific score for higher WHR	rs7598832	C	0.019	
Waist-specific score for higher WHR	rs6719672	G	0.025	
Waist-specific score for higher WHR	rs11724804	G	0.017	
Waist-specific score for higher WHR	rs11747001	A	0.018	
Waist-specific score for higher WHR	rs39312	C	0.016	
Waist-specific score for higher WHR	rs9362097	G	0.016	
Waist-specific score for higher WHR	rs11079041	A	0.016	
Waist-specific score for higher WHR	rs4849294	T	0.015	
Waist-specific score for higher WHR	rs6932767	T	0.017	
Waist-specific score for higher WHR	rs9647379	G	0.014	
Waist-specific score for higher WHR	rs998749	A	0.013	
Waist-specific score for higher WHR	rs10844642	A	0.014	
Waist-specific score for higher WHR	rs17326656	T	0.015	
Waist-specific score for higher WHR	rs6752964	C	0.02	
Waist-specific score for higher WHR	rs11893688	T	0.014	
Waist-specific score for higher WHR	rs332105	G	0.013	
Waist-specific score for higher WHR	rs10963067	C	0.022	
Waist-specific score for higher WHR	rs2058914	G	0.013	
Waist-specific score for higher WHR	rs710122	G	0.012	
Waist-specific score for higher WHR	rs9305545	G	0.016	
Waist-specific score for higher WHR	rs676556	G	0.014	
Waist-specific score for higher WHR	rs6908042	A	0.011	
Waist-specific score for higher WHR	rs544668	T	0.011	
Waist-specific score for higher WHR	rs12631066	C	0.013	

Phenotype	SNP ID	EA [a]	Beta	PMID
Waist-specific score for higher WHR	rs2823096	A	0.014	
Waist-specific score for higher WHR	rs10891483	T	0.017	
Waist-specific score for higher WHR	rs362275	C	0.011	
Waist-specific score for higher WHR	rs13028903	T	0.011	
Waist-specific score for higher WHR	rs2590440	G	0.013	
Waist-specific score for higher WHR	rs2240328	T	0.014	
Waist-specific score for higher WHR	rs4454603	C	0.01	
Waist-specific score for higher WHR	rs1328369	T	0.01	
HbA1c via glycemic mechanisms	rs560887	C	0.0284	
HbA1c via glycemic mechanisms	rs11708067	A	0.0132	
HbA1c via glycemic mechanisms	rs8192675	T	0.0112	
HbA1c via glycemic mechanisms	rs13134327	A	0.0131	
HbA1c via glycemic mechanisms	rs7756992	G	0.0123	
HbA1c via glycemic mechanisms	rs4607517	A	0.0306	28898252[347]
HbA1c via glycemic mechanisms	rs11558471	A	0.015	
HbA1c via glycemic mechanisms	rs2383208	A	0.0142	
HbA1c via glycemic mechanisms	rs579459	C	0.0107	
HbA1c via glycemic mechanisms	rs17747324	C	0.0149	
HbA1c via glycemic mechanisms	rs11603334	G	0.012	
HbA1c via glycemic mechanisms	rs10830963	G	0.0196	
Fasting plasma glucose	rs10438234	C	0.013	
Fasting plasma glucose	rs10747083	A	0.013	
Fasting plasma glucose	rs10811661	T	0.024	
Fasting plasma glucose	rs10814916	C	0.016	
Fasting plasma glucose	rs10830963	G	0.078	
Fasting plasma glucose	rs11039149	A	0.025	
Fasting plasma glucose	rs11558471	A	0.029	
Fasting plasma glucose	rs11603334	G	0.019	
Fasting plasma glucose	rs11607883	G	0.021	
Fasting plasma glucose	rs11619319	G	0.02	
Fasting plasma glucose	rs11708067	A	0.023	
Fasting plasma glucose	rs1280	T	0.026	
Fasting plasma glucose	rs1334577	A	0.013	
Fasting plasma glucose	rs16913693	T	0.043	
Fasting plasma glucose	rs17264369	A	0.017	
Fasting plasma glucose	rs174576	C	0.02	
Fasting plasma glucose	rs2191348	T	0.029	
Fasting plasma glucose	rs2302593	C	0.014	
Fasting plasma glucose	rs2657879	G	0.012	
Fasting plasma glucose	rs340874	C	0.013	
Fasting plasma glucose	rs3829109	G	0.017	
Fasting plasma glucose	rs4258313	G	0.035	
Fasting plasma glucose	rs4502156	T	0.022	
Fasting plasma glucose	rs4869272	T	0.018	
Fasting plasma glucose	rs573225	A	0.063	22581228[346]; 22885924[345]; 24947364[351]
Fasting plasma glucose	rs576674	G	0.017	
Fasting plasma glucose	rs6113722	G	0.035	
Fasting plasma glucose	rs6943153	T	0.015	
Fasting plasma glucose	rs730497	A	0.057	
Fasting plasma glucose	rs7651090	G	0.013	
Fasting plasma glucose	rs7708285	G	0.011	
Fasting plasma glucose	rs7756992	G	0.014	
Fasting plasma glucose	rs780093	C	0.028	
Fasting plasma glucose	rs7903146	T	0.022	
Fasting plasma glucose	rs9814873	A	0.0098	
Fasting plasma glucose	rs983309	T	0.026	
Fasting plasma glucose	rs10305492	G	0.071	
Fasting plasma glucose	rs651007	A	0.0099	
Impaired insulin secretion	rs10811661	T	0.083	
Impaired insulin secretion	rs10830963	G	0.076	
Impaired insulin secretion	rs10946398	C	0.061	
Impaired insulin secretion	rs11603334	G	0.051	
Impaired insulin secretion	rs11605924	A	0.069	
Impaired insulin secretion	rs11672660	T	0.13	
Impaired insulin secretion	rs12686676	G	0.015	
Impaired insulin secretion	rs12779790	G	0.068	
Impaired insulin secretion	rs13266634	C	0.041	
Impaired insulin secretion	rs174550	T	0.031	
Impaired insulin secretion	rs4502156	T	0.05	
Impaired insulin secretion	rs4607517	A	0.032	

Phenotype	SNP ID	EA [a]	Beta	PMID
Impaired insulin secretion	rs5015480	C	0.061	
Impaired insulin secretion	rs5219	T	0.038	
Impaired insulin secretion	rs560887	T	0.047	
Impaired insulin secretion	rs7903146	T	0.06	
Impaired insulin secretion	rs7957197	T	0.062	
Impaired insulin secretion	rs933360	T	0.079	
Insulin resistance Scott	rs10195252	T	0.017	
Insulin resistance Scott	rs17036328	T	0.021	
Insulin resistance Scott	rs2745353	T	0.011	
Insulin resistance Scott	rs2943645	T	0.019	
Insulin resistance Scott	rs3822072	A	0.012	
Insulin resistance Scott	rs459193	G	0.015	
Insulin resistance Scott	rs4846565	G	0.013	
Insulin resistance Scott	rs4865796	A	0.015	
Insulin resistance Scott	rs6536208	A	0.017	
Insulin resistance Scott	rs731839	G	0.015	
Insulin resistance	rs9425291	A	1	
Insulin resistance	rs7973683	C	1	
Insulin resistance	rs2745353	T	1	
Insulin resistance	rs2943645	T	1	
Insulin resistance	rs1045241	C	1	
Insulin resistance	rs8032586	C	1	
Insulin resistance	rs6937438	A	1	
Insulin resistance	rs4865796	A	1	
Insulin resistance	rs11231693	A	1	
Insulin resistance	rs10195252	T	1	
Insulin resistance	rs4738141	G	1	
Insulin resistance	rs3822072	A	1	
Insulin resistance	rs492400	T	1	
Insulin resistance	rs6066149	G	1	
Insulin resistance	rs966544	G	1	
Insulin resistance	rs17386142	C	1	
Insulin resistance	rs2126259	T	1	
Insulin resistance	rs11577194	T	1	
Insulin resistance	rs11130329	A	1	
Insulin resistance	rs6536208	A	1	
Insulin resistance	rs308971	G	1	
Insulin resistance	rs12525532	T	1	
Insulin resistance	rs731839	G	1	
Insulin resistance	rs459193	G	1	
Insulin resistance	rs718314	G	1	
Insulin resistance	rs4846565	G	1	
Insulin resistance	rs2249105	A	1	
Insulin resistance	rs2434612	G	1	
Insulin resistance	rs6887914	C	1	
Insulin resistance	rs10995441	G	1	
Insulin resistance	rs4804311	A	1	
Insulin resistance	rs4976033	G	1	
Insulin resistance	rs8101064	T	1	
Insulin resistance	rs3861397	G	1	
Insulin resistance	rs972283	G	1	
Insulin resistance	rs7323406	A	1	
Insulin resistance	rs1011685	C	1	
Insulin resistance	rs132985	C	1	
Insulin resistance	rs7227237	C	1	
Insulin resistance	rs17169104	G	1	
Insulin resistance	rs2699429	C	1	
Insulin resistance	rs7176058	A	1	
Insulin resistance	rs645040	T	1	
Insulin resistance	rs9492443	C	1	
Insulin resistance	rs683135	A	1	
Insulin resistance	rs754814	T	1	
Insulin resistance	rs17402950	G	1	
Insulin resistance	rs7005992	C	1	
Insulin resistance	rs4804833	A	1	
Insulin resistance	rs498313	A	1	
Insulin resistance	rs295449	A	1	
Insulin resistance	rs3864041	T	1	
Insulin resistance	rs9881942	A	1	
MC4R β-arrestin biased GoF score	rs2229616	T	1	31002796[352]

Phenotype	SNP ID	EA [a]	Beta	PMID
MC4R β -arrestin biased GoF score	rs52820871	G	1	
MC4R β -arrestin biased GoF score	affx89015531	A	1	
MC4R β -arrestin biased GoF score	affx89019383	C	1	

Abbreviations: WHRadjBMI; Waist-to-hip ratio adjusted for BMI; BMI, Body mass index; HbA1c, glycated haemoglobin; *MC4R,* Melanocortin 4 Receptor; GoF, Gain-of-function, EA; Effect allele; PMID, PubMed identification number.

a. EA represents the trait-raising allele

Table S4.8: Correlation of IL-6 levels with DEXA adiposity measures in Fenland.

Adiposity trait (units)[a]	Correlation with IL-6 levels [b c]	P-value
Waist-to-hip ratio	0.17	<0.0001
BMI in kg/m^2	0.25	<0.0001
Body fat percentage in %	0.23	<0.0001
Abdominal fat mass in grams	0.27	<0.0001
Gluteofemoral fat mass in grams	0.20	<0.0001
Leg fat mass in grams	0.18	<0.0001
Abdominal to gluteofemoral fat mass ratio	0.17	<0.0001
Peripheral fat mass in grams	0.20	<0.0001
Visceral fat mass in grams	0.24	<0.0001
Subcutaneous fat mass in grams	0.20	<0.0001

Abbreviations: BMI, Body Mass Index; IL-6, interleukin-6.
Adiposity measures accounted for 9% of the variance in IL-6 levels

a. DEXA adiposity traits were inverse-rank normal transformed
b. IL-6 levels were natural log transformed (pg/mL)
c. Pearson's correlation coefficients are reported

Table S4.9: Variance and heritability in type 2 diabetes and coronary disease occurrence explained by *IL6R* Asp358Ala.

Outcome	Variance explained (%)	Heritability explained (%)
Type 2 diabetes	0.007	0.04
Coronary heart disease	0.007	0.07

Table S4.10: Association of 358Ala with type 2 diabetes conditioning on rs2481065.

Analysis	Type 2 diabetes cases	Controls	OR (95% CI)	P-value
Main	22,182	424,361	0.96 (0.95-0.98)	$3x10^{-4}$
Conditioned on rs2481065	22,182	424,361	0.97 (0.95-0.99)	0.007

Abbreviations: OR, Odds ratio; CI, Confidence interval

Table S5.1: Inflammatory proteins with evidence of an association with T2D or CHD.

Target	Full Name	Gene
6Ckine	C-C motif chemokine 21	*CCL21*
a1-Antitrypsin	Alpha-1-antitrypsin	*SERPINA1*
a1-Microglobulin	Alpha-1-microglobulin	*AMBP*
ABP1	Amiloride-sensitive amine oxidase [copper-containing]	*AOC1*
ACE	Angiotensin-converting enzyme	*ACE*
Activin AC	Inhibin beta A chain:Inhibin beta C chain heterodimer	*INHBA|INHBC*
ADAM 9	Disintegrin and metalloproteinase domain-containing protein 9	*ADAM9*
Adiponectin	Adiponectin	*ADIPOQ*
AIF1	Allograft inflammatory factor 1	*AIF1*
ALCAM	CD166 antigen	*ALCAM*
aldolase C	Fructose-bisphosphate aldolase C	*ALDOC*
Alpha-1B-glycoprotein	Alpha-1B-glycoprotein	*A1BG*
Angiostatin	Angiostatin	*PLG*
ANXA9	Annexin A9	*ANXA9*
Apo A-I	Apolipoprotein A-I	*APOA1*
Apo B	Apolipoprotein B	*APOB*
Apo E	Apolipoprotein E	*APOE*
Apo L1	Apolipoprotein L1	*APOL1*
ARF1	ADP-ribosylation factor 1	*ARF1*
ARGI1	Arginase-1	*ARG1*
AT1B2	Sodium/potassium-transporting ATPase subunit beta-2	*ATP1B2*
ATS13	A disintegrin and metalloproteinase with thrombospondin motifs 13	*ADAMTS13*
AZGP1	Zinc-alpha-2-glycoprotein	*AZGP1*
b2-Microglobulin	Beta-2-microglobulin	*0*
B4GT1	Beta-1,4-galactosyltransferase 1	*B4GALT1*
BGLR	Beta-glucuronidase	*GUSB*
BT3A3	Butyrophilin subfamily 3 member A3	*BTN3A3*
BTNL8	Butyrophilin-like protein 8	*BTNL8*
BTNL9	Butyrophilin-like protein 9	*BTNL9*
C1-Esterase Inhibitor	Plasma protease C1 inhibitor	*SERPING1*
C1QBP	Complement component 1 Q subcomponent-binding protein, mitochondrial	*C1QBP*
C1QR1	Complement component C1q receptor	*CD93*
C1RL1	Complement C1r subcomponent-like protein	*C1RL*
C4a	C4a anaphylatoxin	*C4A|C4B*
C4b-binding protein alpha chain	C4b-binding protein alpha chain	*C4BPA*
C5b, 6 Complex	Complement C5b-C6 complex	*C5|C6*
C9	Complement component C9	*C9*
CAD17	Cadherin-17	*CDH17*
Calpain I	Calpain I	*CAPN1|CAPNS1*
CARD9	Caspase recruitment domain-containing protein 9	*CARD9*
Caspase-3	Caspase-3	*CASP3*
CATE	Cathepsin E	*CTSE*

Target	Full Name	Gene
CATF	Cathepsin F	*CTSF*
Cathepsin B	Cathepsin B	*CTSB*
Cathepsin D	Cathepsin D	*CTSD*
Cathepsin H	Cathepsin H	*CTSH*
Cathepsin V	Cathepsin L2	*CTSV*
CATZ	Cathepsin Z	*CTSZ*
CCL28	C-C motif chemokine 28	*CCL28*
CD109	CD109 antigen	*CD109*
CD23	Low affinity immunoglobulin epsilon Fc receptor	*FCER2*
CD248	Endosialin	*CD248*
CD30 Ligand	Tumor necrosis factor ligand superfamily member 8	*TNFSF8*
CD36 ANTIGEN	Platelet glycoprotein 4	*CD36*
CD46	Membrane cofactor protein	*CD46*
CD48	CD48 antigen	*CD48*
CD59	CD59 glycoprotein	*CD59*
CD5L	CD5 antigen-like	*CD5L*
CD7	T-cell antigen CD7	*CD7*
CD8A	T-cell surface glycoprotein CD8 alpha chain	*CD8A*
CEAM8	Carcinoembryonic antigen-related cell adhesion molecule 8	*CEACAM8*
CECR1	Adenosine deaminase CECR1	*CECR1*
cIAP-1	Baculoviral IAP repeat-containing protein 2	*BIRC2*
CIG49	Interferon-induced protein with tetratricopeptide repeats 3	*IFIT3*
CLC4C	C-type lectin domain family 4 member C	*CLEC4C*
CLC4G	C-type lectin domain family 4 member G	*CLEC4G*
CLC4K	C-type lectin domain family 4 member K	*CD207*
CLM6	CMRF35-like molecule 6	*CD300C*
CM35H	CMRF35-like molecule 8	*CD300A*
CO1A1	Collagen alpha-1(I) chain	*COL1A1*
Coactosin-like protein	Coactosin-like protein	*COTL1*
Coagulation Factor VII	Coagulation factor VII	*F7*
COLEC12	Collectin-12	*COLEC12*
Collagen Type III	Collagen Type III	*COL3A1*
Complement receptor type 1	Complement receptor type 1	*CR1*
Complement receptor type 2	Complement receptor type 2	*CR2*
CPN2	Carboxypeptidase N subunit 2	*CPN2*
CREG1	Protein CREG1	*CREG1*
CRIP1	Cysteine-rich protein 1	*CRIP1*
CRIP2	Cysteine-rich protein 2	*CRIP2*
Cripto	Teratocarcinoma-derived growth factor 1	*TDGF1*
CRP	C-reactive protein	*CRP*
CRTAM	Cytotoxic and regulatory T-cell molecule	*CRTAM*
CXCL16, soluble	C-X-C motif chemokine 16	*CXCL16*
Cystatin C	Cystatin-C	*CST3*

Target	Full Name	Gene
DAF	Complement decay-accelerating factor	*CD55*
DAN	Neuroblastoma suppressor of tumorigenicity 1	*NBL1*
DAPK2	Death-associated protein kinase 2	*DAPK2*
DC-SIGN	CD209 antigen	*CD209*
DLL1	Delta-like protein 1	*DLL1*
DLL4	Delta-like protein 4	*DLL4*
DPP2	Dipeptidyl peptidase 2	*DPP7*
DQA2	HLA class II histocompatibility antigen, DQ alpha 2 chain	*HLA-DQA2*
DR6	Tumor necrosis factor receptor superfamily member 21	*TNFRSF21*
DRB3	HLA class II histocompatibility antigen, DR beta 3 chain	*HLA-DRB3*
Dtk	Tyrosine-protein kinase receptor TYRO3	*TYRO3*
DTX1	E3 ubiquitin-protein ligase DTX1	*DTX1*
ECM1	Extracellular matrix protein 1	*ECM1*
EFNB1	Ephrin-B1	*EFNB1*
EFNB2	Ephrin-B2	*EFNB2*
EMAP-2	Endothelial monocyte-activating polypeptide 2	*AIMP1*
EMBP	Bone marrow proteoglycan	*PRG2*
Endothelin 2	Endothelin-2	*EDN2*
ENPP2	Ectonucleotide pyrophosphatase/phosphodiesterase family member 2	*ENPP2*
Ephrin-A2	Ephrin-A2	*EFNA2*
Epithelial cell kinase	Ephrin type-A receptor 2	*EPHA2*
Epo	Erythropoietin	*EPO*
ERK-1	Mitogen-activated protein kinase 3	*MAPK3*
Factor B	Complement factor B	*CFB*
FADD	FAS-associated death domain protein	*FADD*
FAIM3	Fas apoptotic inhibitory molecule 3	*FAIM3*
Fas ligand, soluble	Tumor necrosis factor ligand superfamily member 6, soluble form	*FASLG*
Fas, soluble	Tumor necrosis factor receptor superfamily member 6	*FAS*
FCAMR	High affinity immunoglobulin alpha and immunoglobulin mu Fc receptor	*FCAMR*
FCG3B	Low affinity immunoglobulin gamma Fc region receptor III-B	*FCGR3B*
FCRL1	Fc receptor-like protein 1	*FCRL1*
FGL1	Fibrinogen-like protein 1	*FGL1*
Fibrinogen	Fibrinogen	*FGA|FGB|FGG*
Ficolin-3	Ficolin-3	*FCN3*
FN1.3	Fibronectin Fragment 3	*FN1*
Fractalkine/CX3CL-1	Fractalkine	*CX3CL1*
FSTL3	Follistatin-related protein 3	*FSTL3*
FUT10	Alpha-(1,3)-fucosyltransferase 10	*FUT10*
Galectin-3	Galectin-3	*LGALS3*
GAS-6	Growth arrest-specific protein 6	*GAS6*
GBP2	Guanylate-binding protein 2	*GBP2*
GCP-2	C-X-C motif chemokine 6	*CXCL6*
G-CSF	Granulocyte colony-stimulating factor	*CSF3*

Target	Full Name	Gene
GDF-11/8	Growth/differentiation factor 11/8	*GDF11|MSTN*
Gelsolin	Gelsolin	*GSN*
GEM	GTP-binding protein GEM	*GEM*
GGH	Gamma-glutamyl hydrolase	*GGH*
GGT2	Inactive gamma-glutamyltranspeptidase 2	*GGT2*
GIB	Phospholipase A2	*PLA2G1B*
Glypican 3	Glypican-3	*GPC3*
GM-CSF	Granulocyte-macrophage colony-stimulating factor	*CSF2*
GNS	N-acetylglucosamine-6-sulfatase	*GNS*
GON1	Progonadoliberin-1	*GNRH1*
GP116	Adhesion G protein-coupled receptor F5	*ADGRF5*
gp130, soluble	Interleukin-6 receptor subunit beta	*IL6ST*
GP1BA	Platelet glycoprotein Ib alpha chain	*GP1BA*
gpIIbIIIa	Integrin alpha-IIb: beta-3 complex	*ITGA2B|ITGB3*
GPNMB	Transmembrane glycoprotein NMB	*GPNMB*
Granulysin	Granulysin	*GNLY*
granzyme A	Granzyme A	*GZMA*
Granzyme B	Granzyme B	*GZMB*
GRB14	Growth factor receptor-bound protein 14	*GRB14*
GRB2-related adapter protein 2	GRB2-related adapter protein 2	*GRAP2*
GRN	Granulins	*GRN*
Haptoglobin, Mixed Type	Haptoglobin	*HP*
HBAZ	Hemoglobin subunit zeta	*HBZ*
HBD-1	Beta-defensin 1	*DEFB1*
HD-5	Defensin-5	*DEFA5*
Hemoglobin	Hemoglobin	*HBA1|HBB*
HERC5	E3 ISG15--protein ligase HERC5	*HERC5*
Hexosaminidase B	Beta-hexosaminidase subunit beta	*HEXB*
HO-1	Heme oxygenase 1	*HMOX1*
HS71B	Heat shock 70 kDa protein 1B	*HSPA1B*
HSP 70	Heat shock protein 70	*HSPA2*
HSP70 protein 8	Heat shock cognate 71 kDa protein	*HSPA8*
HVEM	Tumor necrosis factor receptor superfamily member 14	*TNFRSF14*
ICAM4	Intercellular adhesion molecule 4	*ICAM4*
IDE	Insulin-degrading enzyme	*IDE*
IFIT2	Interferon-induced protein with tetratricopeptide repeats 2	*IFIT2*
IgA	Immunoglobulin A	*IGHA1|IGHA2*
IgD	Immunoglobulin D	*IGHD*
IgE	Immunoglobulin E	*IGHE*
IGFBP-2	Insulin-like growth factor-binding protein 2	*IGFBP2*
IGF-I sR	Insulin-like growth factor 1 receptor	*IGF1R*
IGLL1	Immunoglobulin lambda-like polypeptide 1	*IGLL1*
ihh	Indian hedgehog protein	*IHH*

Target	Full Name	Gene
IL-1 R4	Interleukin-1 receptor-like 1	*IL1RL1*
IL-1 sRI	Interleukin-1 receptor type 1	*IL1R1*
IL-1 sRII	Interleukin-1 receptor type 2	*IL1R2*
IL-18 BPa	Interleukin-18-binding protein	*IL18BP*
IL-1F6	Interleukin-36 alpha	*IL36A*
IL-1Rrp2	Interleukin-1 receptor-like 2	*IL1RL2*
IL-21	Interleukin-21	*IL21*
IL-23	Interleukin-23	*IL12B|IL23A*
IL-23 R	Interleukin-23 receptor	*IL23R*
IL-6 sRa	Interleukin-6 receptor subunit alpha	*IL6R*
ILT-2	Leukocyte immunoglobulin-like receptor subfamily B member 1	*LILRB1*
ILT-4	Leukocyte immunoglobulin-like receptor subfamily B member 2	*LILRB2*
INDO	Indoleamine 2,3-dioxygenase 1	*IDO1*
IP-10	C-X-C motif chemokine 10	*CXCL10*
ISGF3	Interferon regulatory factor 9	*IRF9*
I-TAC	C-X-C motif chemokine 11	*CXCL11*
JAG1	Protein jagged-1	*JAG1*
Kallikrein 7	Kallikrein-7	*KLK7*
KI2L2	Killer cell immunoglobulin-like receptor 2DL2	*KIR2DL2*
KI2L3	Killer cell immunoglobulin-like receptor 2DL3	*KIR2DL3*
KI2S2	Killer cell immunoglobulin-like receptor 2DS2	*KIR2DS2*
KIF3A	Kinesin-like protein KIF3A	*KIF3A*
KLC1	Kinesin light chain 1	*KLC1*
KMT2C	Histone-lysine N-methyltransferase 2C	*KMT2C*
LAG-1	C-C motif chemokine 4-like	*CCL4L1*
LAG-3	Lymphocyte activation gene 3 protein	*LAG3*
LEAP2	Liver-expressed antimicrobial peptide 2	*LEAP2*
Leptin	Leptin	*LEP*
LFA-1 beta-2	Integrin beta-2	*ITGB2*
LGMN	Legumain	*LGMN*
LGP2	Probable ATP-dependent RNA helicase DHX58	*DHX58*
Lipocalin 2	Neutrophil gelatinase-associated lipocalin	*LCN2*
LIRA3	Leukocyte immunoglobulin-like receptor subfamily A member 3	*LILRA3*
LIRA4	Leukocyte immunoglobulin-like receptor subfamily A member 4	*LILRA4*
LIRA5	Leukocyte immunoglobulin-like receptor subfamily A member 5	*LILRA5*
LIRB4	Leukocyte immunoglobulin-like receptor subfamily B member 4	*LILRB4*
LL-37	Antibacterial protein LL-37	*CAMP*
L-plastin	Plastin-2	*LCP1*
LPLC1	BPI fold-containing family B member 1	*BPIFB1*
LRC32	Leucine-rich repeat-containing protein 32	*LRRC32*
LRP8	Low-density lipoprotein receptor-related protein 8	*LRP8*
LRRK2	Leucine-rich repeat serine/threonine-protein kinase 2	*LRRK2*
LTOR3	Ragulator complex protein LAMTOR3	*LAMTOR3*

Target	Full Name	Gene
L-VEGF165	Isoform L-VEGF165	*VEGFA*
Lymphotoxin a1/b2	Lymphotoxin alpha1:beta2	*LTA|LTB*
MANBA	Beta-mannosidase	*MANBA*
MBL	Mannose-binding protein C	*MBL2*
M-CSF R	Macrophage colony-stimulating factor 1 receptor	*CSF1R*
MFAP5	Microfibrillar-associated protein 5	*MFAP5*
MICA	MHC class I polypeptide-related sequence A	*MICA*
MICB	MHC class I polypeptide-related sequence B	*MICB*
Midkine	Midkine	*MDK*
MIP-3b	C-C motif chemokine 19	*CCL19*
MLF1	Myeloid leukemia factor 1	*MLF1*
MMP-8	Neutrophil collagenase	*MMP8*
MO2R1	Cell surface glycoprotein CD200 receptor 1	*CD200R1*
Moesin	Moesin	*MSN*
MSP	Hepatocyte growth factor-like protein	*MST1*
MSP R	Macrophage-stimulating protein receptor	*MST1R*
Mx1	Interferon-induced GTP-binding protein Mx1	*MX1*
Myeloperoxidase	Myeloperoxidase	*MPO*
NCAM-120	Neural cell adhesion molecule 1, 120 kDa isoform	*NCAM1*
NCAM-L1	Neural cell adhesion molecule L1	*L1CAM*
NEUR1	Sialidase-1	*NEU1*
NFASC	Neurofascin	*NFASC*
NKp46	Natural cytotoxicity triggering receptor 1	*NCR1*
Notch 1	Neurogenic locus notch homolog protein 1	*NOTCH1*
NPFF	Pro-FMRFamide-related neuropeptide FF	*NPFF*
NR1H2	Oxysterols receptor LXR-beta	*NR1H2*
NR1H4	Bile acid receptor	*NR1H4*
Osteocalcin	Osteocalcin	*BGLAP*
OX2G	OX-2 membrane glycoprotein	*CD200*
p15-INK4b	Cyclin-dependent kinase 4 inhibitor B	*CDKN2B*
PACAP	Marginal zone B- and B1-cell-specific protein	*PACAP*
PAFAH	Platelet-activating factor acetylhydrolase	*PLA2G7*
PDE5A	cGMP-specific 3',5'-cyclic phosphodiesterase	*PDE5A*
PDGFD	Platelet-derived growth factor D	*PDGFD*
PEBB	Core-binding factor subunit beta	*CBFB*
PGM1	Phosphoglucomutase-1	*PGM1*
PGRP-L	N-acetylmuramoyl-L-alanine amidase	*PGLYRP2*
PIGR	Polymeric immunoglobulin receptor	*PIGR*
PKHA1	Pleckstrin homology domain-containing family A member 1	*PLEKHA1*
prosaposin	Prosaposin	*PSAP*
Proteasome beta chain	Proteasome subunit beta type-4	*PSMB4*
Proteasome subunit alpha type 5	Proteasome subunit alpha type-5	*PSMA5*
PSB2	Proteasome subunit beta type-2	*PSMB2*

Target	Full Name	Gene
P-Selectin	P-selectin	*SELP*
PSME1	Proteasome activator complex subunit 1	*PSME1*
PSP	Lithostathine-1-alpha	*REG1A*
PTPRS	Receptor-type tyrosine-protein phosphatase S	*PTPRS*
PXDN	Peroxidasin homolog	*PXDN*
PYGL	Glycogen phosphorylase, liver form	*PYGL*
RAB14	Ras-related protein Rab-14	*RAB14*
RAB4A	Ras-related protein Rab-4A	*RAB4A*
Rb	Retinoblastoma-associated protein	*RB1*
RBP	Retinol-binding protein 4	*RBP4*
REG1B	Lithostathine-1-beta	*REG1B*
REG3G	Regenerating islet-derived protein 3-gamma	*REG3G*
RGAP1	Rac GTPase-activating protein 1	*RACGAP1*
RNF8	E3 ubiquitin-protein ligase RNF8	*RNF8*
ROR2	Tyrosine-protein kinase transmembrane receptor ROR2	*ROR2*
S100A13	Protein S100-A13	*S100A13*
SAA	Serum amyloid A-1 protein	*SAA1*
SAP3	Ganglioside GM2 activator	*GM2A*
SARP-2	Secreted frizzled-related protein 1	*SFRP1*
SCF	Kit ligand	*KITLG*
SDC3	Syndecan-3	*SDC3*
SEC13	Protein SEC13 homolog	*SEC13*
SECTM1	Secreted and transmembrane protein 1	*SECTM1*
SELS	Selenoprotein S	*VIMP*
SEM4A	Semaphorin-4A	*SEMA4A*
SEM4D	Semaphorin-4D	*SEMA4D*
Sema E	Semaphorin-3C	*SEMA3C*
Semaphorin-7A	Semaphorin-7A	*SEMA7A*
sE-Selectin	E-selectin	*SELE*
SG11A	Sperm-associated antigen 11A	*SPAG11A*
SH21B	SH2 domain-containing protein 1B	*SH2D1B*
sICAM-1	Intercellular adhesion molecule 1	*ICAM1*
sICAM-2	Intercellular adhesion molecule 2	*ICAM2*
sICAM-5	Intercellular adhesion molecule 5	*ICAM5*
SLAF6	SLAM family member 6	*SLAMF6*
sLeptin R	Leptin receptor	*LEPR*
sL-Selectin	L-Selectin	*SELL*
Sonic Hedgehog	Sonic hedgehog protein	*SHH*
SP-D	Pulmonary surfactant-associated protein D	*SFTPD*
Sperm-associated antigen 11	Sperm-associated antigen 11B	*SPAG11B*
SPLC2	BPI fold-containing family A member 2	*BPIFA2*
sRAGE	Advanced glycosylation end product-specific receptor, soluble	*AGER*
STAT1	Signal transducer and activator of transcription 1-alpha/beta	*STAT1*

Target	Full Name	Gene
STAT3	Signal transducer and activator of transcription 3	*STAT3*
sTie-2	Angiopoietin-1 receptor, soluble	*TEK*
STOM	Erythrocyte band 7 integral membrane protein	*STOM*
suPAR	Urokinase plasminogen activator surface receptor	*PLAUR*
SWP70	Switch-associated protein 70	*SWAP70*
Syntenin 1	Syntenin-1	*SDCBP*
SYT11	Synaptotagmin-11	*SYT11*
TAFI	Carboxypeptidase B2	*CPB2*
TCCR	Interleukin-27 receptor subunit alpha	*IL27RA*
TECK	C-C motif chemokine 25	*CCL25*
TGF-b R III	Transforming growth factor beta receptor type 3	*TGFBR3*
Thrombin	Thrombin	*F2*
Thrombopoietin Receptor	Thrombopoietin Receptor	*MPL*
TIGIT	T-cell immunoreceptor with Ig and ITIM domains	*TIGIT*
TIMP-2	Metalloproteinase inhibitor 2	*TIMP2*
TLR1	Toll-like receptor 1	*TLR1*
TLR4:MD-2 complex	Toll-like receptor 4:Lymphocyte antigen 96 complex	*TLR4|LY96*
TLR5	Toll-like receptor 5	*TLR5*
TNF sR-II	Tumor necrosis factor receptor superfamily member 1B	*TNFRSF1B*
TNFSF15	Tumor necrosis factor ligand superfamily member 15	*TNFSF15*
TPSN	Tapasin	*TAPBP*
TR	Transferrin receptor protein 1	*TFRC*
TRML2	Trem-like transcript 2 protein	*TREML2*
TRY3	Trypsin-3	*PRSS3*
Trypsin 2	Trypsin-2	*PRSS2*
TSG-6	Tumor necrosis factor-inducible gene 6 protein	*TNFAIP6*
TSP4	Thrombospondin-4	*THBS4*
UB2D2	Ubiquitin-conjugating enzyme E2 D2	*UBE2D2*
UB2D3	Ubiquitin-conjugating enzyme E2 D3	*UBE2D3*
Uromodulin	Uromodulin	*UMOD*
VCAM-1	Vascular cell adhesion protein 1	*VCAM1*
VEGF sR2	Vascular endothelial growth factor receptor 2	*KDR*
VEGF-D	Vascular endothelial growth factor D	*FIGF*
WNT3A	Protein Wnt-3a	*WNT3A*
WNT5A	Protein Wnt-5a	*WNT5A*

Table S5.2: Sentinel variants for the 85 loci used in multi-trait colocalisation.

Variant	Chr	Pos	Gene	Consequence	Locus
rs55730499	6	1.61E+08	*LPA*	intron variant	1
rs10882101	10	94462427	-	intergenic variant	2
rs602633	1	1.1E+08	*CELSR2*	downstream gene variant	3
rs2972144	2	2.27E+08	-	intergenic variant	4
rs6511720	19	11202306	*LDLR*	intron variant	5
rs7412	19	45412079	*APOE*	missense variant	6
rs11637783	15	79139000	*MORF4L1*	intron variant	7
rs3184504	12	1.12E+08	*SH2B3*	missense variant	8
rs115505614	5	1.02E+08	*GIN1*	3 prime UTR variant	9
rs974819	11	1.04E+08	*RP11-563P16.1*	intron variant non-coding transcript variant	10
rs72802342	16	75234872	*CTRB2*	downstream gene variant	11
rs5213	11	17408404	*KCNJ11*	3 prime UTR variant	12
rs3768321	1	40035928	*SNORA55*	upstream gene variant	13
rs11591147	1	55505647	*PCSK9*	missense variant	14
rs2954029	8	1.26E+08	*RP11-136O12.2*	intron variant non-coding transcript variant	15
rs1260326	2	27730940	*GCKR*	missense variant splice region variant	16
rs2246942	10	91004886	*LIPA*	intron variant	17
rs4736819	8	41509915	*ANK1*	downstream gene variant	18
rs1561198	2	85809989	*VAMP8*	downstream gene variant	19
rs3918226	7	1.51E+08	*NOS3*	intron variant	20
rs601945	6	32573415	*HLA-DRB1*	intron variant	21
rs28505901	9	1.39E+08	*GPSM1*	intron variant	22
rs243024	2	60583665	*AC007381.2*	downstream gene variant	23
rs507666	9	1.36E+08	*ABO*	intron variant non-coding transcript variant	24
rs56348580	12	1.21E+08	*HNF1A*	synonymous variant	25
rs1250229	2	2.16E+08	*FN1*	upstream gene variant	26
rs9379084	6	7231843	*RREB1*	missense variant	27
rs3887925	3	1.87E+08	*ST6GAL1*	intron variant	28
rs668948	2	21291529	-	intergenic variant	29
rs4845625	1	1.54E+08	*IL6R*	intron variant	30
rs2028150	2	65655012	*SPRED2*	intron variant	31
rs2681492	12	90013089	*ATP2B1*	intron variant	32
rs3740390	10	1.05E+08	*AS3MT*	intron variant	33
rs12980942	19	41832231	*TGFB1*	downstream gene variant	34
rs8107974	19	19388500	*TM6SF2*	upstream gene variant	35
rs964184	11	1.17E+08	*ZNF259*	3 prime UTR variant	36
rs2925979	16	81534790	*CMIP*	intron variant	37
rs1783541	11	65294799	*SCYL1*	intron variant	38
rs348330	1	2.3E+08	*ABCB10*	intron variant	39
rs17689007	8	9974824	*MSRA*	intron variant	40
rs15285	8	19824667	*LPL*	3 prime UTR variant	41
rs2280141	10	1.24E+08	*PLEKHA1*	downstream gene variant	42

Variant	Chr	Pos	Gene	Consequence	Locus
rs11070332	15	41809205	*LTK*	upstream gene variant	43
rs58730668	4	1.86E+08	*ACSL1*	intron variant	44
rs4678145	3	1.24E+08	*UMPS*	intron variant	45
rs2800733	6	1.27E+08	-	intergenic variant	46
rs2072633	6	31919578	*CFB*	intron variant	47
rs10840293	11	9751196	*SWAP70*	intron variant	48
rs6600191	16	295795	*ITFG3*	intron variant	49
rs3111316	19	13038415	*FARSA*	intron variant	50
rs34855406	17	40731411	*FAM134C*	downstream gene variant	51
rs6905288	6	43758873	*VEGFA*	downstream gene variant	52
rs2814993	6	34618893	*C6orf106*	intron variant	53
rs6984210	8	22033615	*BMP1*	intron variant	54
rs57327348	8	10808687	*XKR6*	intron variant	55
rs17680741	10	82251514	*TSPAN14*	intron variant	56
rs7617773	3	48193515	-	intergenic variant	57
rs1050362	16	72130815	*DHX38*	synonymous variant	58
rs11642430	16	30045789	*FAM57B*	upstream gene variant	59
rs738408	22	44324730	*PNPLA3*	synonymous variant	60
rs1580278	4	1.04E+08	-	intergenic variant	61
rs12001437	9	34074476	-	intergenic variant	62
rs12500824	4	77416627	*SHROOM3*	intron variant	63
rs4688760	3	49980596	*RBM6*	intron variant	64
rs7178762	15	63871292	*USP3*	intron variant	65
rs362307	4	3241845	*HTT*	3 prime UTR variant	66
rs7947761	11	1.01E+08	*ARHGAP42*	intron variant	67
rs73082363	3	49130378	*QRICH1*	intron variant	68
rs329122	5	1.34E+08	*JADE2*	intron variant	69
rs7124681	11	47529947	*CELF1*	intron variant	70
rs11926707	3	46925539	*PTH1R*	intron variant	71
rs10512861	3	1.32E+08	*DNAJC13*	downstream gene variant	72
rs699	1	2.31E+08	*AGT*	missense variant	73
rs2581787	3	53127677	*RFT1*	intron variant	74
rs3741414	12	57844049	*INHBC*	3 prime UTR variant	75
rs62007683	14	1.04E+08	*MARK3*	intron variant	76
rs145904381	1	1.51E+08	*BNIPL*	intron variant	77
rs78840640	7	23434606	*IGF2BP3*	intron variant	78
rs2269247	1	64107284	*PGM1*	intron variant	79
rs35497231	3	1.52E+08	-	intergenic variant	80
rs2727301	17	61965043	*RP11-630H24.3*	downstream gene variant	81
rs12802972	11	1704596	*FAM99B*	Non-coding transcript exon variant	82
rs80223330	4	1.2E+08	*PDE5A*	intron variant	83
rs76549422	20	31760953	*BPIFA2*	intron variant	84
rs2268382	7	1.3E+08	*CPA1*	intron variant	85

Table S5.3: Gene sets in which inflammatory proteins colocalising with T2D are significantly enriched.

Category	Gene Set	Total genes	Number of overlapping genes	P-value	FDR adjusted P-value	Overlapping genes
GWAS catalog	Blood protein levels	457	89	8.17E-09	1.5E-05	*TNFRSF1B, EPHA2, NBL1, LEPR, FCRL1, CRP, SLAMF6, CD48, FCGR3B, CREG1, SELL, NFASC, C4BPA, CD55, MBL2, PSAP, RBP4, CTSD, CD59, CTSF, MMP8, NCAM1, APOA1, CRTAM, C1RL, CPB2, LCP1, F7, LGALS3, SERPINA1, HP, GRN, MPO, CD300A, CD300C, TIMP2, CLEC4G, CD209, IL27RA, PGLYRP2, PLAUR, APOE, LILRB2, LILRA3, LILRA5, LILRB1, LILRB4, NCR1, PXDN, APOB, REG3G, REG1A, IL1R2, IL1R1, TNFAIP6, CST3, BPIFA2, BPIFB1, CTSZ, CECR1, MST1, ALCAM, ADIPOQ, CPN2, FGG, C9, IL6ST, THBS4, LEAP2, CSF1R, BTN3A3, AIF1, NEU1, CD109, ARG1, PLG, DLL1, GPNMB, AOC1, DEFB1, CTSB, FGL1, GGH, LY96, B4GALT1, CCL21, SEMA4D, TLR4, GSN*
Cancer gene neighbourhoods	CAR HPX	28	15	9.11E-08	3.9E-05	*C4BPA, MBL2, RBP4, F2, APOA1, CPB2, HP, APOB, FGB, FGA, FGG, C9, ARG1, PLG, FGL1*
Computational gene sets	CAR HPX	28	15	9.11E-08	7.8E-05	*C4BPA, MBL2, RBP4, F2, APOA1, CPB2, HP, APOB, FGB, FGA, FGG, C9, ARG1, PLG, FGL1*
GO cc	GO COLLAGEN CONTAINING EXTRACELLULAR MATRIX	103	31	2.86E-07	2.0E-04	*SDC3, MBL2, PSAP, CTSD, MDK, F2, PRG2, CTSF, MMP8, NCAM1, APOA1, MFAP5, F7, LGALS3, SERPINA1, SEMA7A, TIMP2, APOE, A1BG, PXDN, CTSZ, MST1, WNT5A, ADIPOQ, FGB, FGA, FGG, THBS4, PLG, CTSB, L1CAM*
GO cc	GO EXTRACELLULAR MATRIX	126	35	3.92E-07	2.0E-04	*SDC3, TGFBR3, MBL2, PSAP, CTSD, MDK, F2, PRG2, CD248, CTSF, LRRC32, MMP8, NCAM1, APOA1, MFAP5, F7, LGALS3, SERPINA1, SEMA7A, TIMP2, APOE, A1BG, PXDN, CTSZ, MST1, WNT5A, ADIPOQ, CPN2, FGB, FGA, FGG, THBS4, PLG, CTSB, L1CAM*
KEGG	KEGG COMPLEMENT AND COAGULATION CASCADES	44	17	4.42E-06	8.2E-04	*C4BPA, CD55, CR2, CR1, CD46, MBL2, CD59, F2, CPB2, F7, SERPINA1, PLAUR, FGB, FGA, FGG, C9, PLG*
Chemical and Genetic pertubation	CAIRO LIVER DEVELOPMENT DN	56	21	5.32E-07	1.8E-03	*CRP, PSAP, F2, PRG2, CPB2, LCP1, SERPINA1, HP, MPO, APOE, APOB, MST1, FGB, FGA, FGG, C9, GM2A, ARG1, PLG, GSN, LCN2*
Curated gene sets	CAIRO LIVER DEVELOPMENT DN	56	21	5.32E-07	2.9E-03	*CRP, PSAP, F2, PRG2, CPB2, LCP1, SERPINA1, HP, MPO, APOE, APOB, MST1, FGB, FGA, FGG, C9, GM2A, ARG1, PLG, GSN, LCN2*
GO cc	GO CELL SURFACE	266	54	1.26E-05	4.2E-03	*EPHA2, SDC3, LEPR, TGFBR3, FCRL1, SELL, FASLG, CD55, CR1, CD46, MBL2, CD59, F2, CD248, LRRC32, NCAM1, APOA1, PLA2G1B, LGALS3, IGHA2, IGHA1, B2M, SEMA7A, C1QBP, TIMP2, CD209, IL27RA, CEACAM8, PLAUR, LILRB2, LILRA5, LILRB1, IL1R1, CD93, CTSZ, WNT5A, ALCAM, ADIPOQ, TFRC, AIMP1, FGB, FGA, FGG, IL6ST, CSF1R, BTN3A3, CD109, PLG,*

Category	Gene Set	Total genes	Number of overlapping genes	P-value	FDR adjusted P-value	Overlapping genes
						ADAM9, B4GALT1, ROR2, TLR4, NOTCH1, L1CAM
Canonical Pathways	KEGG COMPLEMENT AND COAGULATION CASCADES	44	17	4.42E-06	9.7E-03	*C4BPA, CD55, CR2, CR1, CD46, MBL2, CD59, F2, CPB2, F7, SERPINA1, PLAUR, FGB, FGA, FGG, C9, PLG*
Canonical Pathways	NABA ECM GLYCOPROTEINS	25	12	8.78E-06	9.7E-03	*BGLAP, MFAP5, GAS6, PXDN, TNFAIP6, IGFBP2, ADIPOQ, FGB, FGA, FGG, THBS4, FGL1*
Cancer gene neighborhoods	CAR IGFBP1	17	9	4.75E-05	0.01	*C4BPA, MBL2, F2, APOA1, CPB2, APOB, C9, ARG1, PLG*
Canonical Pathways	NABA CORE MATRISOME	30	13	1.47E-05	0.01	*BGLAP, PRG2, MFAP5, GAS6, PXDN, TNFAIP6, IGFBP2, ADIPOQ, FGB, FGA, FGG, THBS4, FGL1*
Chemical and Genetic pertubation	HSIAO LIVER SPECIFIC GENES	64	21	6.72E-06	0.01	*CRP, C4BPA, MBL2, RBP4, F2, APOA1, CPB2, F7, SERPINA1, HP, APOE, APOB, MST1, CPN2, FGB, FGA, FGG, C9, ARG1, PLG, FGL1*
Curated gene sets	KEGG COMPLEMENT AND COAGULATION CASCADES	44	17	4.42E-06	0.01	*C4BPA, CD55, CR2, CR1, CD46, MBL2, CD59, F2, CPB2, F7, SERPINA1, PLAUR, FGB, FGA, FGG, C9, PLG*
Curated gene sets	HSIAO LIVER SPECIFIC GENES	64	21	6.72E-06	0.01	*CRP, C4BPA, MBL2, RBP4, F2, APOA1, CPB2, F7, SERPINA1, HP, APOE, APOB, MST1, CPN2, FGB, FGA, FGG, C9, ARG1, PLG, FGL1*
Curated gene sets	NABA ECM GLYCOPROTEINS	25	12	8.78E-06	0.01	*BGLAP, MFAP5, GAS6, PXDN, TNFAIP6, IGFBP2, ADIPOQ, FGB, FGA, FGG, THBS4, FGL1*
Curated gene sets	NABA CORE MATRISOME	30	13	1.47E-05	0.02	*BGLAP, PRG2, MFAP5, GAS6, PXDN, TNFAIP6, IGFBP2, ADIPOQ, FGB, FGA, FGG, THBS4, FGL1*
Curated gene sets	YAO HOXA10 TARGETS VIA PROGESTERONE UP	19	10	1.86E-05	0.02	*NBL1, LEPR, NCAM1, GAS6, APOE, REG1B, ADIPOQ, GEM, ENPP2, LCN2*
BioCarta	BIOCARTA FIBRINOLYSIS PATHWAY	8	6	6.67E-05	0.02	*F2, CPB2, FGB, FGA, FGG, PLG*
Computational gene sets	CAR IGFBP1	17	9	4.75E-05	0.02	*C4BPA, MBL2, F2, APOA1, CPB2, APOB, C9, ARG1, PLG*
Computational gene sets	MODULE 88	161	36	6.81E-05	0.02	*TNFRSF1B, EPHA2, C4BPA, RBP4, MDK, F2, PRG2, LRRC32, APOA1, CPB2, SERPINA1, HP, GRN, MPO, FSTL3, EFNA2, APOE, APOB, REG1B, REG1A, IL1R1, IGFBP2, MST1, CPN2, FGB, FGA, ARG1, PLG, GPNMB, AOC1, DEFB1, FGL1, ENPP2, CCL21, LCN2, L1CAM*
Chemical and Genetic pertubation	YAO HOXA10 TARGETS VIA PROGESTERONE UP	19	10	1.86E-05	0.02	*NBL1, LEPR, NCAM1, GAS6, APOE, REG1B, ADIPOQ, GEM, ENPP2, LCN2*
Wikipathways	Blood Clotting Cascade	8	6	6.67E-05	0.03	*F2, F7, FGB, FGA, FGG, PLG*

Category	Gene Set	Total genes	Number of overlapping genes	P-value	FDR adjusted P-value	Overlapping genes
Wikipathways	Human Complement System	64	19	9.40E-05	0.03	*CRP, SELL, C4BPA, CD55, CR2, CR1, CD46, MBL2, CD59, APOA1, C1QBP, PLAUR, CD93, ADIPOQ, FGB, FGA, FGG, C9, PLG*
BioCarta	BIOCARTA AMI PATHWAY	9	6	0.00018	0.03	*F2, F7, FGB, FGA, FGG, PLG*
Computational gene sets	MODULE 131	15	8	0.000122	0.03	*CD59, F2, PLAUR, MST1, FGB, FGA, FGG, PLG*
Cancer modules	MODULE 88	161	36	6.81E-05	0.03	*TNFRSF1B, EPHA2, C4BPA, RBP4, MDK, F2, PRG2, LRRC32, APOA1, CPB2, SERPINA1, HP, GRN, MPO, FSTL3, EFNA2, APOE, APOB, REG1B, REG1A, IL1R1, IGFBP2, MST1, CPN2, FGB, FGA, ARG1, PLG, GPNMB, AOC1, DEFB1, FGL1, ENPP2, CCL21, LCN2, L1CAM*
Cancer modules	MODULE 131	15	8	0.000122	0.03	*CD59, F2, PLAUR, MST1, FGB, FGA, FGG, PLG*
Cancer modules	MODULE 23	108	26	0.000237	0.03	*TNFRSF1B, C4BPA, RBP4, F2, CTSF, APOA1, NPFF, CPB2, F7, SERPINA1, HP, TIMP2, EFNA2, APOE, APOB, IGFBP2, MST1, CPN2, FGB, FGA, ARG1, PLG, DEFB1, FGL1, SEMA4D, ROR2*
Canonical Pathways	BIOCARTA FIBRINOLYSIS PATHWAY	8	6	6.67E-05	0.04	*F2, CPB2, FGB, FGA, FGG, PLG*
Curated gene sets	APPEL IMATINIB RESPONSE	17	9	4.75E-05	0.04	*CTSD, CD300A, APOE, CTSZ, HEXB, GM2A, NEU1, GUSB, CTSB*
Chemical and Genetic pertubation	APPEL IMATINIB RESPONSE	17	9	4.75E-05	0.04	*CTSD, CD300A, APOE, CTSZ, HEXB, GM2A, NEU1, GUSB, CTSB*
Computational gene sets	MODULE 23	108	26	0.000237	0.04	*TNFRSF1B, C4BPA, RBP4, F2, CTSF, APOA1, NPFF, CPB2, F7, SERPINA1, HP, TIMP2, EFNA2, APOE, APOB, IGFBP2, MST1, CPN2, FGB, FGA, ARG1, PLG, DEFB1, FGL1, SEMA4D, ROR2*
BioCarta	BIOCARTA EXTRINSIC PATHWAY	7	5	0.000426	0.04	*F2, F7, FGB, FGA, FGG*
Wikipathways	Selenium Micronutrient Network	33	12	0.000247	0.04	*CRP, F2, APOA1, F7, VIMP, MPO, APOB, GGT2, FGB, FGA, FGG, PLG*
Curated gene sets	BIOCARTA FIBRINOLYSIS PATHWAY	8	6	6.67E-05	0.05	*F2, CPB2, FGB, FGA, FGG, PLG*

Table S5.4: Gene sets in which inflammatory proteins colocalising with CHD are significantly enriched.

Category	Gene Set	Total genes	Number of overlapping genes	P-value	FDR Adjusted P-value	Overlapping genes
GWAScatalog	Blood protein levels	457	103	4.04E-11	7.33E-08	*NBL1, LRP8, PGM1, IL6R, FCRL1, CD5L, CRP, SLAMF6, CD48, FCGR3B, CREG1, SELP, SELL, SELE, PIGR, CD55, PSAP, IFIT2, CTSD, SAA1, SERPING1, CTSF, MMP8, APOA1, LAG3, C1RL, CPB2, LGALS3, LGMN, CXCL16, GP1BA, ATP1B2, ACE, ACE, CD300A, CD300C, TIMP2, CD7, SECTM1, COLEC12, CLEC4G, CD209, CCL25, ICAM1, ICAM4, ICAM5, APOE, KLK7, LILRB1, LILRB4, NCR1, APOB, CD8A, IL1R2, IL1R1, IL1RL1, CST3, BPIFA2, CTSZ, CECR1, TDGF1, CAMP, CD200, CD200R1, ADIPOQ, CPN2, KDR, CXCL6, CXCL10, CXCL11, MANBA, FGG, C9, IL6ST, CSF1R, IL12B, BTNL8, BTN3A3, MICA, LTA, TNF, LTB, HSPA1A, NEU1, CFB, CFB, TAPBP, TREML2, PLA2G7, PLG, IL6, GPNMB, SEMA3C, EPO, DEFB1, DEFA5, CTSB, SFRP1, GGH, LY96, B4GALT1, CCL21, SEMA4D, TLR4, GSN*
GO cc	GO CELL SURFACE	266	62	1.08E-06	0.001	*SDC3, MPL, TGFBR3, VCAM1, IL6R, FCRL1, SELP, SELL, CD55, CR1, CD46, APOA1, LAG3, PLA2G1B, LGALS3, IGHD, B2M, SEMA7A, GP1BA, C1QBP, ITGA2B, ITGB3, ITGB3, ACE, ACE, TIMP2, CD209, ICAM1, LILRB1, CD8A, IL1R1, IL1RL1, CD93, CTSZ, TDGF1, WNT5A, CD200, CD200R1, TIGIT, ADIPOQ, TFRC, CXCL10, AIMP1, FGB, FGA, FGG, IL6ST, CSF1R, IL12B, BTNL8, BTN3A3, MICA, TNF, TREML2, PLG, AZGP1, EPO, SHH, ADAM9, SFRP1, B4GALT1, CTSV, TLR4, L1CAM*
Wikipathways	Human Complement System	64	23	3.34E-06	0.002	*CRP, SELP, SELL, SELE, CD55, CR2, CR1, CD46, SERPING1, APOA1, C1QBP, ITGA2B, ITGB3, ITGB3, ICAM2, ICAM1, CD93, ADIPOQ, FGB, FGA, FGG, C9, CFB, CFB, PLG*
GO cc	GO EXTRACELLULAR MATRIX	126	34	2.28E-05	0.01	*SDC3, TGFBR3, PSAP, CTSD, MDK, PRG2, SERPING1, CTSF, MMP8, APOA1, MFAP5, LGALS3, SEMA7A, GP1BA, COL1A1, TIMP2, COLEC12, ICAM1, APOE, COL3A1, IHH, CTSZ, WNT5A, ADIPOQ, CPN2, FGB, FGA, FGG, PLG, AZGP1, SHH, CTSB, SFRP1, L1CAM*
GO cc	GO EXTERNAL SIDE OF PLASMA MEMBRANE	137	36	2.40E-05	0.01	*TGFBR3, VCAM1, IL6R, SELP, SELL, LAG3, IGHD, B2M, SEMA7A, GP1BA, ITGA2B, ACE, ACE, CD209, ICAM1, LILRB1, CD8A, IL1R1, IL1RL1, CD200R1, TFRC, CXCL10, FGB, FGA, FGG, IL6ST, IL12B, BTNL8, BTN3A3, MICA, TNF, PLG, AZGP1, ADAM9, B4GALT1, CTSV, TLR4*
Wikipathways	Selenium Micronutrient Network	33	14	3.79E-05	0.01	*CRP, SAA1, APOA1, IFNG, VIMP, ICAM1, APOB, GGT2, FGB, FGA, FGG, TNF, PLG, IL6*

Category	Gene Set	Total genes	Number of overlapping genes	P-value	FDR Adjusted P-value	Overlapping genes
GO cc	GO COLLAGEN CONTAINING EXTRACELLULAR MATRIX	103	29	4.23E-05	0.01	*SDC3, PSAP, CTSD, MDK, PRG2, SERPING1, CTSF, MMP8, APOA1, MFAP5, LGALS3, SEMA7A, COL1A1, TIMP2, ICAM1, APOE, COL3A1, CTSZ, WNT5A, ADIPOQ, FGB, FGA, FGG, PLG, AZGP1, SHH, CTSB, SFRP1, L1CAM*
GO bp	GO EXTRACELLULAR STRUCTURE ORGANIZATION	100	31	2.32E-06	0.02	*VCAM1, MMP8, APOA1, MFAP5, CPB2, GAS6, ITGA2B, ITGB3, ITGB3, COL1A1, ICAM2, TIMP2, ICAM1, ICAM4, ICAM5, APOE, KLK7, APOB, COL3A1, IHH, CST3, KDR, FGB, FGA, FGG, TNF, PLA2G7, PLG, IL6, SDCBP, B4GALT1, CTSV*
Reactome	REACTOME EXTRACELLULAR MATRIX ORGANIZATION	73	24	1.22E-05	0.02	*SDC3, VCAM1, CTSD, MMP8, MFAP5, ITGA2B, ITGB3, ITGB3, COL1A1, ICAM2, TIMP2, PTPRS, ICAM1, ICAM4, ICAM5, KLK7, COL3A1, KDR, FGB, FGA, FGG, PLG, CTSB, ADAM9, CTSV*
Hallmark gene sets	HALLMARK COAGULATION	64	19	0.000486	0.02	*S100A13, CTSE, SERPING1, MMP8, APOA1, CPB2, LGMN, CRIP2, GP1BA, ITGB3, ITGB3, FGA, FGG, C9, CFB, CFB, PLG, CTSB, ADAM9, CTSV, GSN*
Canonical Pathways	REACTOME EXTRACELLULAR MATRIX ORGANIZATION	73	24	1.22E-05	0.03	*SDC3, VCAM1, CTSD, MMP8, MFAP5, ITGA2B, ITGB3, ITGB3, COL1A1, ICAM2, TIMP2, PTPRS, ICAM1, ICAM4, ICAM5, KLK7, COL3A1, KDR, FGB, FGA, FGG, PLG, CTSB, ADAM9, CTSV*
GO cc	GO LYSOSOMAL LUMEN	33	13	0.000182	0.04	*SDC3, PSAP, CTSD, CTSF, GNS, LGMN, APOB, MANBA, HEXB, NEU1, GUSB, CTSB, CTSV*
GO cc	GO LOW DENSITY LIPOPROTEIN PARTICLE	6	5	0.000243	0.04	*APOA1, VIMP, APOE, APOB, PLA2G7*
GO cc	GO SIDE OF MEMBRANE	169	39	0.000257	0.04	*TGFBR3, VCAM1, IL6R, SELP, SELL, BIRC2, LAG3, LRRK2, IGHD, B2M, SEMA7A, GP1BA, ITGA2B, ACE, ACE, CD209, ICAM1, LILRB1, CD8A, IL1R1, IL1RL1, CD200R1, TFRC, CXCL10, FGB, FGA, FGG, IL6ST, IL12B, BTNL8, BTN3A3, MICA, TNF, TAPBP, PLG, AZGP1, ADAM9, B4GALT1, CTSV, TLR4*
GO mf	GO SIGNALING RECEPTOR BINDING	484	91	4.04E-05	0.04	*NBL1, TGFBR3, VCAM1, S100A13, IL6R, SEMA4A, CRP, TLR5, RAB4A, PSAP, SAA1, MDK, IL18BP, APOA1, LAG3, NPFF, GDF11, IL23A, IFNG, KITLG, NR1H4, PLA2G1B, EFNB2, GAS6, LGALS3, IGHD, SEMA7A, IGF1R, CXCL16, C1QBP, ITGB3, ITGB3, ACE, ACE, ICAM2, TIMP2, SECTM1, CCL25, ICAM1, ICAM4, ICAM5, APOE, LILRB1, APOB, CD8A, IL1R1, GRB14, COL3A1, MSTN, STAT1, IHH, CECR1, TDGF1, WNT5A, TIGIT, ADIPOQ, KDR, CXCL6, CXCL10, CXCL11, AIMP1, FGB, FGA, FGG, CCL28, IL6ST, IL12B, BTNL8, BTN3A3, MICA, LTA, TNF, LTB, HSPA1A, TAPBP, PLG, IL6, GPNMB, GUSB, SEMA3C, EPO, SHH, DEFB1, ADAM9, SFRP1, SDCBP,*

Category	Gene Set	Total genes	Number of overlapping genes	P-value	FDR Adjusted P-value	Overlapping genes
						LY96, CCL19, CCL21, SEMA4D, TNFSF8, TLR4, FIGF
GO mf	GO CHEMOREPELLENT ACTIVITY	5	5	4.56E-05	0.04	*SEMA4A, APOA1, SEMA7A, SEMA3C, SEMA4D*
KEGG	KEGG LYSOSOME	34	13	0.00026	0.05	*CTSE, PSAP, CTSD, CTSF, GNS, LGMN, CTSZ, MANBA, HEXB, NEU1, GUSB, CTSB, CTSV*
Hallmark gene sets	HALLMARK COMPLEMENT	87	22	0.001963	0.05	*S100A13, CD55, CR2, CR1, CD46, CTSD, SERPING1, MMP8, LGALS3, LGMN, GP1BA, TIMP2, C9, GZMA, HSPA1A, CFB, CFB, PLA2G7, PLG, IL6, CTSB, ADAM9, CTSV*

Table S6.1: Search strategy used for the systematic literature search on incretin GWAS studies.

	Search Terms
#7	#1 AND #2 AND #3 AND #4 AND #5 AND #6
#6	English[lang]
#5	"humans"[MeSH Terms]
#4	("2007/06/01"[PDAT] : "2019/11/04"[PDAT])
#3	(Clinical Trial[ptyp] OR Meta-Analysis[ptyp] OR systematic[sb] OR Journal Article[ptyp] OR Review[ptyp])
#2	((((((((((((((((("genome-wide association study"[MeSH Terms] OR ("genome-wide"[All Fields] AND "association"[All Fields] AND "study"[All Fields]) OR "genome-wide association study"[All Fields] OR "gwas"[All Fields]) OR "GWA study"[All Fields]) OR "Genome-wide association study"[All Fields]) OR "Genome wide association analysis"[All Fields]) OR "Genome wide association scan"[All Fields]) OR "Genome wide association"[All Fields]) OR "Genome-wide association"[All Fields]) OR "Genome wide"[All Fields]) OR ("Socioaffect Neurosci Psychol"[Journal] OR "snp"[All Fields])) OR "Single nucleotide polymorphism"[All Fields]) OR "Genetic variant"[All Fields]) OR "Genetic variants"[All Fields]) OR "Genetic variation"[All Fields]) OR "Genetic variations"[All Fields]) OR "Genetic mutation"[All Fields]) OR "Genetic mutations"[All Fields]) OR "Genetic mutant"[All Fields]) OR "Genetic mutants"[All Fields]) OR "genome-wide association study"[MeSH Terms])
#1	((((((((((((((((((((((((((((((((((("glucagon-like peptide 1"[MeSH Terms] OR "glucagon-like peptide 1"[All Fields] OR "glp 1"[All Fields]) OR GLP1[All Fields]) OR "Glucagon-like peptide-1"[All Fields]) OR "Glucagon-like peptide 1"[All Fields]) OR "Glucagon-like peptide-I"[All Fields]) OR "Glucagon-like peptide I"[All Fields]) OR GIP[All Fields]) OR "Glucose-dependent insulinotropic polypeptide"[All Fields]) OR "Glucose dependent insulinotropic polypeptide"[All Fields]) OR "Glucose-dependent insulinotropic peptide"[All Fields]) OR "Glucose dependent insulinotropic peptide"[All Fields]) OR "Gastric inhibitory polypeptide"[All Fields]) OR "Gastric inhibitory peptide"[All Fields]) OR (Gastroinhibitory[All Fields] AND ("peptides"[MeSH Terms] OR "peptides"[All Fields] OR "polypeptide"[All Fields]))) OR (Gastroinhibitory[All Fields] AND ("peptides"[MeSH Terms] OR "peptides"[All Fields] OR "peptide"[All Fields]))) OR ("glucagon-like peptide 2"[MeSH Terms] OR "glucagon-like peptide 2"[All Fields] OR "glp 2"[All Fields])) OR GLP2[All Fields]) OR "Glucagon-like peptide-2"[All Fields]) OR "Glucagon-like peptide-2"[All Fields]) OR "Glucagon-like peptide-II"[All Fields]) OR "Glucagon-like peptide-II"[All Fields]) OR ("secretin"[MeSH Terms] OR "secretin"[All Fields])) OR SCT[All Fields]) OR ("oxyntomodulin"[MeSH Terms] OR "oxyntomodulin"[All Fields])) OR OXM[All Fields]) OR ("glucagon"[MeSH Terms] OR "glucagon"[All Fields])) OR GCG[All Fields]) OR ("proglucagon"[MeSH Terms] OR "proglucagon"[All Fields] OR "preproglucagon"[All Fields])) OR "Peptide YY"[All Fields]) OR "Peptide tyrosine tyrosine"[All Fields]) OR PYY[All Fields]) OR ("gastrins"[MeSH Terms] OR "gastrins"[All Fields] OR "gastrin"[All Fields])) OR GAST[All Fields]) OR ("ghrelin"[MeSH Terms] OR "ghrelin"[All Fields])) OR GHRL[All Fields]) OR ("glicentin"[MeSH Terms] OR "glicentin"[All Fields])) OR "gastric inhibitory polypeptide"[MeSH Terms]) OR "gastric inhibitory polypeptide"[MeSH Terms]) OR "peptide yy"[MeSH Terms]) OR "secretin"[MeSH Terms]) OR "ghrelin"[MeSH Terms]) OR "glucagon"[MeSH Terms]) OR "glicentin"[MeSH Terms]) OR "glucagon-like peptide 1"[MeSH Terms]) OR "glucagon-like peptide 2"[MeSH Terms]) OR "oxyntomodulin"[MeSH Terms]) OR "incretins"[MeSH Terms])

Table S7.1: Summary of the participating studies.

Outcome type	Outcome	Cases overall, N	Non-cases (for case-control studies) or participants (for continuous trait studies) overall, N	Participating study	PubMed ID for cohort description
Disease outcomes	Type 2 diabetes	74,124	842,006	DIAMANTE	30297969
	Type 2 diabetes (BMI adjusted)	74,124	842,006	DIAMANTE	30297969
	Coronary heart disease	34,541	261,984	CARDIoGRAMplusC4D, UK Biobank	28714975, 25826379
	Any Stroke	40,585	406,111	MEGASTROKE	29531354
	Any Ischemic Stroke	34,217			
	Cardioembolic Stroke	7,193			
	Large Artery Stroke	4,373			
	Small Vessel Stroke	5,386			
	Abdominal Aortic Aneurysm	1,094	366,492	UK Biobank	31756303
	Atrial Fibrillation	16,945	350,641		
	Aortic Valve Stenosis	2,244	365,342		
	Coronary Artery Disease	29,278	338,308		
	Deep Vein Thrombosis	9,454	358,132		
	Haemorrhagic Stroke (all)	1,981	365,605		
	Heart Failure	6,712	360,874		
	Ischaemic Cerebrovascular Disease (all)	8,084	359,502		
	Pulmonary Embolism	6,148	361,438		
	Peripheral Vascular Disease	3,415	364,171		
	Thoracic Aortic Aneurysm	347	367,239		
	Transient Ischaemic Attack	3,962	363,624		
	Intracerebral Haemorrhage	1,064	366,522		
	Subarachnoid Haemorrhage	1,084	366,502		
	Ischaemic Stroke	4,602	362,984		
	Ischaemic stroke plus haemorrhagic stroke plus unknown stroke (but not TIA)	9,652	357,934		
	Venous Thromboembolism (all)	14,097	353,489		
Glycaemic outcomes	Fasting glucose (BMI adjusted)		51,750	MAGIC	22581228
	Non-fasted plasma glucose		413,905	UK Biobank; InterAct	25826379
	2-hr glucose (BMI adjusted)		41,888	MAGIC	20081857
	Fasting insulin (BMI adjusted)		51,750	MAGIC	22581228
	Corrected insulin response		5,318	MAGIC	24699409
	HbA1C		451,782	UK Biobank; InterAct	25826379
Cardiovascular and lipid-related outcomes	Apolipoprotein A1		412,328	UK Biobank; InterAct	25826379
	High-density lipoprotein		450,957		
	Apolipoprotein B		448,859		
	Low-density lipoprotein		375,774		
	Lipoprotein A		406,825		
	Total cholesterol		377,031		

Outcome type	Outcome	Cases overall, N	Non-cases (for case-control studies) or participants (for continuous trait studies) overall, N	Participating study	PubMed ID for cohort description
	Triglycerides		450,625		
	C-reactive protein		465,067		
Anthropometric outcomes	BMI		738,628	GIANT, UK Biobank	25673413; 25826379
	Hip circumference		568,765		
	Hip circumference (BMI adjusted)		633,860		
	Waist circumference		654,577		
	Waist circumference (BMI adjusted)		654,253		
	Waist-to-hip ratio		636,672		
	Waist-to-hip ratio (BMI adjusted)		636,282		
Additional biomarker outcomes	Albumin		415,714	UK Biobank; InterAct	25826379
	Alkaline phosphatase		450,743		
	Alanine aminotransferase		452,291		
	Aspartate transaminase		450,594		
	Bilirubin		448,652		
	Calcium		414,173		
	Creatinine		451,942		
	Gamma-glutamyl transpeptidase		450,745		
	Urate		451,665		
Regional adiposity outcomes (measured by bio-impedance)	Android fat mass		435,387	UK Biobank	25826379
	Arms fat mass				
	Gynoid fat mass				
	Legs fat mass				
	Peripheral fat mass				
	Subcutaneous fat mass				
	Total fat mass				
	Trunk fat mass				
	Visceral fat mass				
	Appendicular lean mass				
	Android lean mass				
	Arms lean mass				
	Gynoid lean mass				
	Legs lean mass				
	Total lean mass				
	Trunk lean mass				
Plasma proteins	4,979 proteins		10,708	Fenland	27841877
Metabolites	1,008 metabolites		11,539	EPIC-Norfolk	10466767
GIP measures	Fasting and 2hr GIP		7,828	MDC and PPP-Botnia	29093273

Table S7.2: Clusters of colocalised traits identified by the main analysis across the permutations of prior 2 and the regional and alignment thresholds.

Locus	Colocalised traits	PP coloc	Candidate variant	PP explained	N variants	Prior 2	Regional and alignment threshold
GIPR	LDL, CHD, HDL, Total Cholesterol, Lipoprotein A, ApoB	1	rs7412	1	425	0.02	0.5
GIPR	Glucose, HbA1c, ApoA1, WHRadjBMI, Waist circumference adjBMI, WHR	0.7034	rs4420638	1	425	0.02	0.5
GIPR	GIP SOMAmer 16292_288, Fasting GIP, 2hr GIP, BMI, Hip circumference, Waist circumference, 2hr Glucose adjBMI	0.9707	rs11672660	0.9994	425	0.02	0.5
GIPR	Hip circumference adjBMI, T2D	0.9717	rs10408179	1	425	0.02	0.5
GIPR	LDL, CHD, HDL, Total Cholesterol, Lipoprotein A, ApoB	1	rs7412	1	425	0.02	0.6
GIPR	Glucose, HbA1c, ApoA1, WHRadjBMI, Waist circumference adjBMI, WHR	0.7034	rs4420638	1	425	0.02	0.6
GIPR	GIP SOMAmer 16292_288, Fasting GIP, 2hr GIP, BMI, Hip circumference, Waist circumference, 2hr Glucose adjBMI	0.9707	rs11672660	0.9994	425	0.02	0.6
GIPR	Hip circumference adjBMI, T2D	0.9717	rs10408179	1	425	0.02	0.6
GIPR	LDL, CHD, HDL, Total Cholesterol, Lipoprotein A, ApoB	1	rs7412	1	425	0.02	0.7
GIPR	Glucose, HbA1c, ApoA1, WHRadjBMI, Waist circumference adjBMI, WHR	0.7034	rs4420638	1	425	0.02	0.7
GIPR	GIP SOMAmer 16292_288, Fasting GIP, 2hr GIP, BMI, Hip circumference, Waist circumference, 2hr Glucose adjBMI	0.9707	rs11672660	0.9994	425	0.02	0.7
GIPR	Hip circumference adjBMI, T2D	0.9717	rs10408179	1	425	0.02	0.7
GIPR	LDL, CHD, HDL, Total Cholesterol, Lipoprotein A, ApoB	1	rs7412	1	425	0.02	0.8
GIPR	HbA1c, ApoA1, WHRadjBMI, Waist circumference adjBMI, WHR	0.9994	rs4420638	1	425	0.02	0.8
GIPR	GIP SOMAmer 16292_288, Fasting GIP, 2hr GIP, BMI, Glucose, Hip circumference, Waist circumference, 2hr Glucose adjBMI	0.9683	rs11672660	0.9995	425	0.02	0.8
GIPR	Hip circumference adjBMI, T2D	0.9717	rs10408179	1	425	0.02	0.8
GIPR	LDL, CHD, HDL, Total Cholesterol, Lipoprotein A, ApoB	1	rs7412	1	425	0.02	0.9
GIPR	HbA1c, ApoA1, WHRadjBMI, Waist circumference adjBMI, WHR	0.9994	rs4420638	1	425	0.02	0.9
GIPR	GIP SOMAmer 16292_288, Fasting GIP, 2hr GIP, BMI, Glucose, Hip circumference, Waist circumference, 2hr Glucose adjBMI	0.9683	rs11672660	0.9995	425	0.02	0.9
GIPR	Hip circumference adjBMI, T2D	0.9717	rs10408179	1	425	0.02	0.9
GIPR	LDL, CHD, HDL, Total Cholesterol, Lipoprotein A, ApoB	1	rs7412	1	425	0.01	0.5
GIPR	Glucose, HbA1c, ApoA1, WHRadjBMI, Waist circumference adjBMI, WHR	0.5466	rs4420638	1	425	0.01	0.5
GIPR	GIP SOMAmer 16292_288, Fasting GIP, 2hr GIP, BMI, Hip circumference, Waist circumference, 2hr Glucose adjBMI	0.9441	rs11672660	0.9994	425	0.01	0.5
GIPR	Hip circumference adjBMI, T2D	0.945	rs10408179	1	425	0.01	0.5
GIPR	LDL, CHD, HDL, Total Cholesterol, Lipoprotein A, ApoB	1	rs7412	1	425	0.01	0.6
GIPR	HbA1c, ApoA1, WHRadjBMI, Waist circumference adjBMI, WHR	0.9989	rs4420638	1	425	0.01	0.6
GIPR	GIP SOMAmer 16292_288, Fasting GIP, 2hr GIP, BMI, Glucose, Hip circumference, Waist circumference, 2hr Glucose adjBMI	0.9397	rs11672660	0.9995	425	0.01	0.6
GIPR	Hip circumference adjBMI, T2D	0.945	rs10408179	1	425	0.01	0.6

Locus	Colocalised traits	PP coloc	Candidate variant	PP explained	N variants	Prior 2	Regional and alignment threshold
GIPR	LDL, CHD, HDL, Total Cholesterol, Lipoprotein A, ApoB	1	rs7412	1	425	0.01	0.7
GIPR	HbA1c, ApoA1, WHRadjBMI, Waist circumference adjBMI, WHR	0.9989	rs4420638	1	425	0.01	0.7
GIPR	GIP SOMAmer 16292_288, Fasting GIP, 2hr GIP, BMI, Glucose, Hip circumference, Waist circumference, 2hr Glucose adjBMI	0.9397	rs11672660	0.9995	425	0.01	0.7
GIPR	Hip circumference adjBMI, T2D	0.945	rs10408179	1	425	0.01	0.7
GIPR	LDL, CHD, HDL, Total Cholesterol, Lipoprotein A, ApoB	1	rs7412	1	425	0.01	0.8
GIPR	HbA1c, ApoA1, WHRadjBMI, Waist circumference adjBMI, WHR	0.9989	rs4420638	1	425	0.01	0.8
GIPR	GIP SOMAmer 16292_288, Fasting GIP, 2hr GIP, BMI, Glucose, Hip circumference, Waist circumference, 2hr Glucose adjBMI	0.9397	rs11672660	0.9995	425	0.01	0.8
GIPR	Hip circumference adjBMI, T2D	0.945	rs10408179	1	425	0.01	0.8
GIPR	LDL, CHD, HDL, Total Cholesterol, Lipoprotein A, ApoB	1	rs7412	1	425	0.01	0.9
GIPR	HbA1c, ApoA1, WHRadjBMI, Waist circumference adjBMI, WHR	0.9989	rs4420638	1	425	0.01	0.9
GIPR	GIP SOMAmer 16292_288, Fasting GIP, 2hr GIP, BMI, Glucose, Hip circumference, Waist circumference, 2hr Glucose adjBMI	0.9397	rs11672660	0.9995	425	0.01	0.9
GIPR	Hip circumference adjBMI, T2D	0.945	rs10408179	1	425	0.01	0.9
GIPR	LDL, CHD, HDL, Total Cholesterol, Lipoprotein A, ApoB	1	rs7412	1	425	0.001	0.5
GIPR	HbA1c, ApoA1, WHRadjBMI, Waist circumference adjBMI, WHR	0.9889	rs4420638	1	425	0.001	0.5
GIPR	GIP SOMAmer 16292_288, Fasting GIP, 2hr GIP, BMI, Glucose, Hip circumference, Waist circumference, 2hr Glucose adjBMI	0.5855	rs11672660	0.9995	425	0.001	0.5
GIPR	Hip circumference adjBMI, T2D	0.6322	rs10408179	1	425	0.001	0.5
GIPR	LDL, CHD, HDL, Total Cholesterol, Lipoprotein A, ApoB	1	rs7412	1	425	0.001	0.6
GIPR	HbA1c, ApoA1, WHRadjBMI, Waist circumference adjBMI, WHR	0.9889	rs4420638	1	425	0.001	0.6
GIPR	GIP SOMAmer 16292_288, Fasting GIP, 2hr GIP, BMI, Glucose, Hip circumference, Waist circumference, 2hr Glucose adjBMI	0.5855	rs11672660	0.9995	425	0.001	0.6
GIPR	Hip circumference adjBMI, T2D	0.6322	rs10408179	1	425	0.001	0.6
GIPR	LDL, CHD, HDL, Total Cholesterol, Lipoprotein A, ApoB	1	rs7412	1	425	0.001	0.7
GIPR	HbA1c, ApoA1, WHRadjBMI, Waist circumference adjBMI, WHR	0.9889	rs4420638	1	425	0.001	0.7
GIPR	GIP SOMAmer 16292_288, Fasting GIP, 2hr GIP, BMI, Glucose, Hip circumference, Waist circumference, 2hr Glucose adjBMI	0.5855	rs11672660	0.9995	425	0.001	0.7
GIPR	LDL, CHD, HDL, Total Cholesterol, Lipoprotein A, ApoB	1	rs7412	1	425	0.001	0.8
GIPR	HbA1c, ApoA1, WHRadjBMI, Waist circumference adjBMI, WHR	0.9889	rs4420638	1	425	0.001	0.8
GIPR	GIP SOMAmer 16292_288, Fasting GIP, 2hr GIP, BMI, Hip circumference, Waist circumference	0.7977	rs11672660	0.9994	425	0.001	0.8
GIPR	Glucose, Hip circumference adjBMI	0.8674	rs725660	0.7079	425	0.001	0.8
GIPR	LDL, CHD, HDL, Total Cholesterol, Lipoprotein A, ApoB	1	rs7412	1	425	0.001	0.9
GIPR	HbA1c, ApoA1, WHRadjBMI, Waist circumference adjBMI, WHR	0.9889	rs4420638	1	425	0.001	0.9

Locus	Colocalised traits	PP coloc	Candidate variant	PP explained	N variants	Prior 2	Regional and alignment threshold
GIPR	BMI, Waist circumference	0.9524	rs1800437	0.6839	425	0.001	0.9
GIPR	Fasting GIP, Hip circumference	0.921	rs11672660	0.9998	425	0.001	0.9

Abbreviations: PP, Posterior probability; N, Number; LDL, Low-density lipoprotein; CHD, Coronary heart disease; HDL, High-density lipoprotein; ApoB, Apolipoprotein B; ApoA1, Apolipoprotein A1; HbA1c, Glycated haemoglobin; WHR, Waist-to-hip ratio; adjBMI, Adjusted for BMI; BMI, Body mass index; GIP, Gastric inhibitory polypeptide

Table S7.3: Clusters of colocalised traits identified by the sensitivity analysis across the permutations of prior 2 and the regional and alignment thresholds.

Locus	Colocalised traits	PP coloc	Candidate variant	PP explained	N variants	Prior 2	Regional and alignment threshold
GIPR	LDL, CHD, HDL, Total Cholesterol, Lipoprotein A, ApoB	1	rs7412	1	5016	0.02	0.5
GIPR	HbA1c, ApoA1, WHRadjBMI, Waist circumference adjBMI, WHR, T2D	0.8471	rs429358	1	5016	0.02	0.5
GIPR	BMI, Waist circumference	0.9975	rs1800437	0.6839	5016	0.02	0.5
GIPR	Triglycerides, Hip circumference adjBMI	0.983	rs5117	0.9328	5016	0.02	0.5
GIPR	GIP SOMAmer 16292_288, Hip circumference, 2hr Glucose adjBMI	0.8999	rs11672660	0.9996	5016	0.02	0.5
GIPR	LDL, CHD, HDL, Total Cholesterol, Lipoprotein A, ApoB	1	rs7412	1	5016	0.02	0.6
GIPR	HbA1c, ApoA1, WHRadjBMI, Waist circumference adjBMI, WHR, T2D	0.8471	rs429358	1	5016	0.02	0.6
GIPR	BMI, Waist circumference	0.9975	rs1800437	0.6839	5016	0.02	0.6
GIPR	Triglycerides, Hip circumference adjBMI	0.983	rs5117	0.9328	5016	0.02	0.6
GIPR	GIP SOMAmer 16292_288, Hip circumference, 2hr Glucose adjBMI	0.8999	rs11672660	0.9996	5016	0.02	0.6
GIPR	LDL, CHD, HDL, Total Cholesterol, Lipoprotein A, ApoB	1	rs7412	1	5016	0.02	0.7
GIPR	HbA1c, ApoA1, WHRadjBMI, Waist circumference adjBMI, WHR, T2D	0.8471	rs429358	1	5016	0.02	0.7
GIPR	BMI, Waist circumference	0.9975	rs1800437	0.6839	5016	0.02	0.7
GIPR	Triglycerides, Hip circumference adjBMI	0.983	rs5117	0.9328	5016	0.02	0.7
GIPR	GIP SOMAmer 16292_288, Hip circumference, 2hr Glucose adjBMI	0.8999	rs11672660	0.9996	5016	0.02	0.7
GIPR	LDL, CHD, HDL, Total Cholesterol, Lipoprotein A, ApoB	1	rs7412	1	5016	0.02	0.8
GIPR	HbA1c, ApoA1, WHRadjBMI, Waist circumference adjBMI, WHR, T2D	0.8471	rs429358	1	5016	0.02	0.8
GIPR	BMI, Waist circumference	0.9975	rs1800437	0.6839	5016	0.02	0.8
GIPR	Triglycerides, Hip circumference adjBMI	0.983	rs5117	0.9328	5016	0.02	0.8
GIPR	GIP SOMAmer 16292_288, Hip circumference, 2hr Glucose adjBMI	0.8999	rs11672660	0.9996	5016	0.02	0.8
GIPR	LDL, CHD, HDL, Total Cholesterol, Lipoprotein A, ApoB	1	rs7412	1	5016	0.02	0.9
GIPR	HbA1c, ApoA1, WHRadjBMI, Waist circumference adjBMI, WHR	0.9651	rs429358	1	5016	0.02	0.9
GIPR	BMI, Waist circumference	0.9975	rs1800437	0.6839	5016	0.02	0.9
GIPR	Triglycerides, Hip circumference adjBMI	0.983	rs5117	0.9328	5016	0.02	0.9
GIPR	GIP SOMAmer 16292_288, Hip circumference, 2hr Glucose adjBMI	0.8999	rs11672660	0.9996	5016	0.02	0.9
GIPR	LDL, CHD, HDL, Total Cholesterol, Lipoprotein A, ApoB	1	rs7412	1	5016	0.01	0.5
GIPR	HbA1c, ApoA1, WHRadjBMI, Waist circumference adjBMI, WHR, T2D	0.7384	rs429358	1	5016	0.01	0.5
GIPR	BMI, Waist circumference	0.995	rs1800437	0.6839	5016	0.01	0.5
GIPR	Triglycerides, Hip circumference adjBMI	0.9665	rs5117	0.9328	5016	0.01	0.5
GIPR	GIP SOMAmer 16292_288, Hip circumference, 2hr Glucose adjBMI	0.815	rs11672660	0.9996	5016	0.01	0.5
GIPR	LDL, CHD, HDL, Total Cholesterol, Lipoprotein A, ApoB	1	rs7412	1	5016	0.01	0.6
GIPR	HbA1c, ApoA1, WHRadjBMI, Waist circumference adjBMI, WHR, T2D	0.7384	rs429358	1	5016	0.01	0.6
GIPR	BMI, Waist circumference	0.995	rs1800437	0.6839	5016	0.01	0.6

Locus	Colocalised traits	PP coloc	Candidate variant	PP explained	N variants	Prior 2	Regional and alignment threshold
GIPR	Triglycerides, Hip circumference adjBMI	0.9665	rs5117	0.9328	5016	0.01	0.6
GIPR	GIP SOMAmer 16292_288, Hip circumference, 2hr Glucose adjBMI	0.815	rs11672660	0.9996	5016	0.01	0.6
GIPR	LDL, CHD, HDL, Total Cholesterol, Lipoprotein A, ApoB	1	rs7412	1	5016	0.01	0.7
GIPR	HbA1c, ApoA1, WHRadjBMI, Waist circumference adjBMI, WHR, T2D	0.7384	rs429358	1	5016	0.01	0.7
GIPR	BMI, Waist circumference	0.995	rs1800437	0.6839	5016	0.01	0.7
GIPR	Triglycerides, Hip circumference adjBMI	0.9665	rs5117	0.9328	5016	0.01	0.7
GIPR	GIP SOMAmer 16292_288, Hip circumference, 2hr Glucose adjBMI	0.815	rs11672660	0.9996	5016	0.01	0.7
GIPR	LDL, CHD, HDL, Total Cholesterol, Lipoprotein A, ApoB	1	rs7412	1	5016	0.01	0.8
GIPR	HbA1c, ApoA1, WHRadjBMI, Waist circumference adjBMI, WHR	0.9334	rs429358	1	5016	0.01	0.8
GIPR	BMI, Waist circumference	0.995	rs1800437	0.6839	5016	0.01	0.8
GIPR	Triglycerides, Hip circumference adjBMI	0.9665	rs5117	0.9328	5016	0.01	0.8
GIPR	GIP SOMAmer 16292_288, Hip circumference, 2hr Glucose adjBMI	0.815	rs11672660	0.9996	5016	0.01	0.8
GIPR	LDL, CHD, HDL, Total Cholesterol, Lipoprotein A, ApoB	1	rs7412	1	5016	0.01	0.9
GIPR	HbA1c, ApoA1, WHRadjBMI, Waist circumference adjBMI, WHR	0.9334	rs429358	1	5016	0.01	0.9
GIPR	BMI, Waist circumference	0.995	rs1800437	0.6839	5016	0.01	0.9
GIPR	Triglycerides, Hip circumference adjBMI	0.9665	rs5117	0.9328	5016	0.01	0.9
GIPR	GIP SOMAmer 16292_288, Hip circumference	0.9509	rs11672660	0.9996	5016	0.01	0.9
GIPR	LDL, CHD, HDL, Total Cholesterol, Lipoprotein A, ApoB	1	rs7412	1	5016	0.001	0.5
GIPR	HbA1c, ApoA1, WHRadjBMI, Waist circumference adjBMI, WHR	0.5827	rs429358	1	5016	0.001	0.5
GIPR	BMI, Waist circumference	0.9524	rs1800437	0.6839	5016	0.001	0.5
GIPR	Triglycerides, Hip circumference adjBMI	0.7428	rs5117	0.9328	5016	0.001	0.5
GIPR	GIP SOMAmer 16292_288, Hip circumference	0.6348	rs11672660	0.9996	5016	0.001	0.5
GIPR	LDL, CHD, HDL, Total Cholesterol, Lipoprotein A, ApoB	1	rs7412	1	5016	0.001	0.6
GIPR	HbA1c, ApoA1, WHRadjBMI, WHR	0.8888	rs429358	1	5016	0.001	0.6
GIPR	BMI, Waist circumference	0.9524	rs1800437	0.6839	5016	0.001	0.6
GIPR	Triglycerides, Hip circumference adjBMI	0.7428	rs5117	0.9328	5016	0.001	0.6
GIPR	GIP SOMAmer 16292_288, Hip circumference	0.6348	rs11672660	0.9996	5016	0.001	0.6
GIPR	LDL, CHD, HDL, Total Cholesterol, Lipoprotein A, ApoB	1	rs7412	1	5016	0.001	0.7
GIPR	HbA1c, ApoA1, WHRadjBMI, WHR	0.8888	rs429358	1	5016	0.001	0.7
GIPR	BMI, Waist circumference	0.9524	rs1800437	0.6839	5016	0.001	0.7
GIPR	Triglycerides, Hip circumference adjBMI	0.7428	rs5117	0.9328	5016	0.001	0.7
GIPR	GIP SOMAmer 16292_288, Hip circumference	0.6348	rs11672660	0.9996	5016	0.001	0.7
GIPR	LDL, CHD, HDL, Total Cholesterol, Lipoprotein A, ApoB	1	rs7412	1	5016	0.001	0.8
GIPR	HbA1c, ApoA1, WHRadjBMI, WHR	0.8888	rs429358	1	5016	0.001	0.8
GIPR	BMI, Waist circumference	0.9524	rs1800437	0.6839	5016	0.001	0.8
GIPR	LDL, CHD, HDL, Total Cholesterol, Lipoprotein A, ApoB	1	rs7412	1	5016	0.001	0.9

Locus	Colocalised traits	PP coloc	Candidate variant	PP explained	N variants	Prior 2	Regional and alignment threshold
GIPR	HbA1c, ApoA1, WHR	0.9999	rs429358	1	5016	0.001	0.9
GIPR	BMI, Waist circumference	0.9524	rs1800437	0.6839	5016	0.001	0.9

Abbreviations: PP, Posterior probability; N, Number; LDL, Low-density lipoprotein; CHD, Coronary heart disease; HDL, High-density lipoprotein; ApoB, Apolipoprotein B; ApoA1, Apolipoprotein A1; HbA1c, Glycated haemoglobin; WHR, Waist-to-hip ratio; adjBMI, Adjusted for BMI; BMI, Body mass index; GIP, Gastric inhibitory polypeptide

Table S7.4: Association of E354 with the traits it was associated with at nominal significance from the 2SMR analysis, after conditioning on independent SNPs for each trait. Estimates of 2-hour glucose, Total cholesterol and BMI were not included in this table as the independent signal selection showed that E354 was one of the independent variants.

	2SMR result		Conditional result		Independent variant	
Trait	**Beta (SE)**	**P-value**	**Beta (SE)**	**P-value**	**Conditioned on[a]**	**LD with rs1800437[b]**
T2D	-0.03 (0.007)	$7x10^{-5}$	-0.03 (0.008)	$4x10^{-4}$	rs3810291	0.001
T2DadjBMI	-0.07 (0.009)	$2x10^{-14}$	-0.02 (0.009)	0.04	rs2238689	0.363
CHD	0.03 (0.007)	$2x10^{-6}$	0.01 (0.007)	0.06	rs1964272	0.269
SVS	-0.08 (0.029)	0.009	-0.04 (0.029)	0.12	rs1964272	0.269
Non-fasted plasma glucose	0.02 (0.003)	$3x10^{-8}$	0.01 (0.003)	0.05	rs1964272	0.269
HbA1c	-0.01 (0.003)	$1x10^{-7}$	-0.0003 (0.003)	0.92	rs9676912	0.356
ApoA1	0.01 (0.003)	$3x10^{-6}$	0.002 (0.003)	0.37	rs2238689	0.363
HDL	0.02 (0.003)	$7x10^{-9}$	0.003 (0.003)	0.31	rs2238689	0.363
ApoB	0.02 (0.002)	$5x10^{-13}$	0.01 (0.002)	$2x10^{-5}$	rs7412	0.004
LDL	0.02 (0.003)	$2x10^{-16}$	0.016 (0.003)	$1x10^{-8}$	rs7412	0.004
Triglycerides	-0.01 (0.003)	$2x10^{-5}$	-0.01 (0.003)	$5x10^{-5}$	rs4803936	0.001
CRP	-0.01 (0.002)	0.02	-0.004 (0.002)	0.07	rs7412	0.004
Albumin	-0.01 (0.003)	$6x10^{-6}$	-0.01 (0.003)	0.001	rs35114617	0.061
Creatinine	-0.02 (0.002)	$1x10^{-11}$	-0.02 (0.002)	$3x10^{-11}$	rs7412	0.004
QPCTL	-0.07 (0.016)	$9x10^{-6}$	0.01 (0.016)	0.48	rs1964272	0.269
Secretoglobin family 3A member 1	-0.08 (0.017)	$6x10^{-7}$	-0.04 (0.017)	0.01	rs61703905	0.1

Abbreviations: SE, Standard error; T2D, Type 2 diabetes; adjBMI, Adjusted for BMI; CHD, Coronary heart disease; SVS, Small vessel stroke; HbA1c, Glycated haemoglobin; Apo, Apolipoprotein; HDL, High-density lipoprotein; LDL, Low-density lipoprotein; CRP, C-reactive protein; QPCTL, Glutaminyl-peptide cyclotransferase like

a. The independent variant showing the greatest attenuation of the E354 association estimate with the respective trait
b. LD estimates are in R^2 and are quoted from 5 European populations in the LDlink database v4.1.0

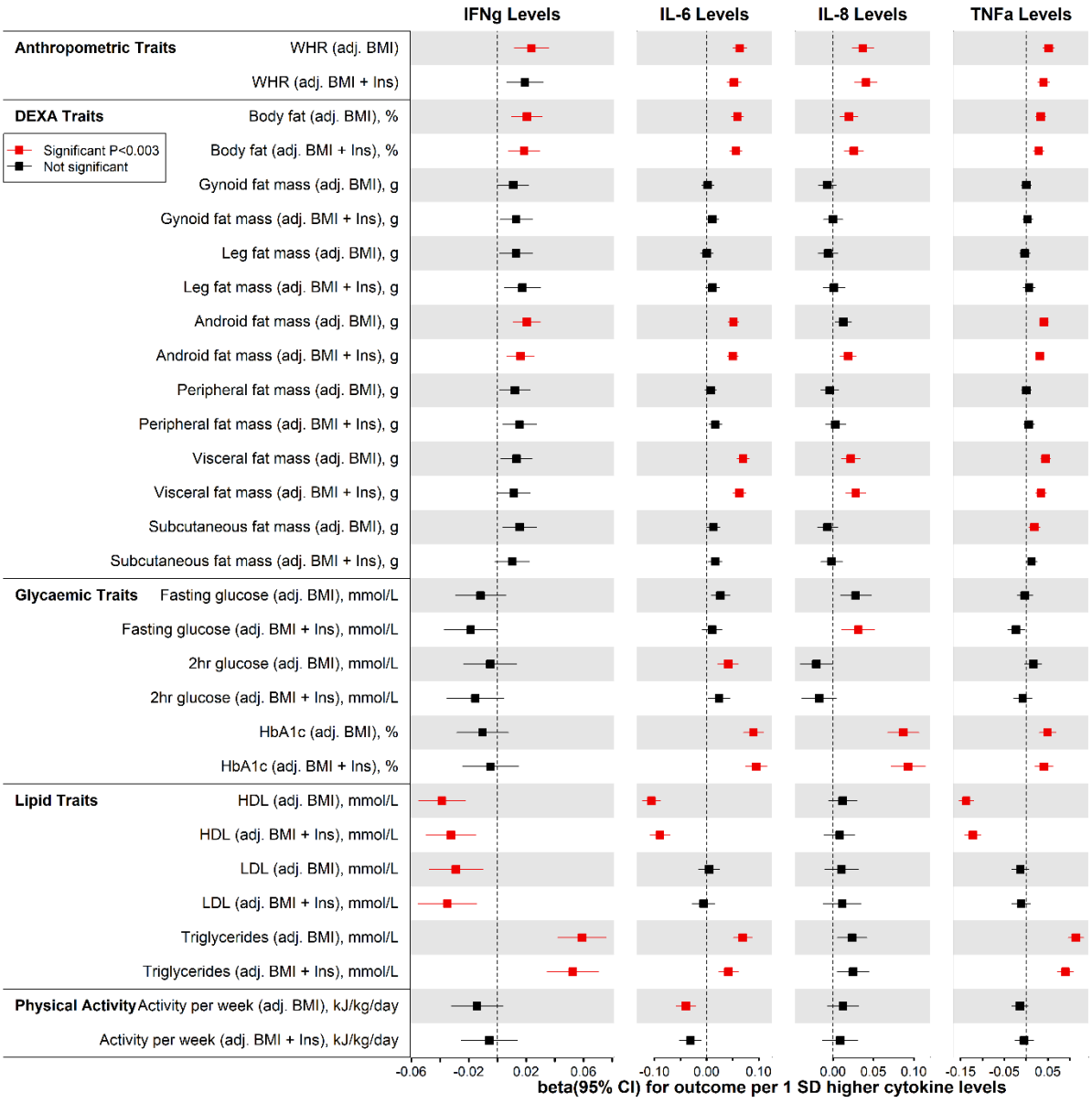
Table S7.5: Association of other previously identified fasting GIP variants with CHD. The association of rs2287019 with CHD was not considered, as it is in high LD with E354.

Variant	Chr:pos	EA	EAF	Beta	SE	P-value	Cases	Controls
rs17681684	17:9792768	A	0.3082	-0.0074	0.0057	0.1925	34,541	261,984

Abbreviations: Chr, Chromosome; pos, Position; EA, Effect allele; EAF, Effect allele frequency; SE, Standard error

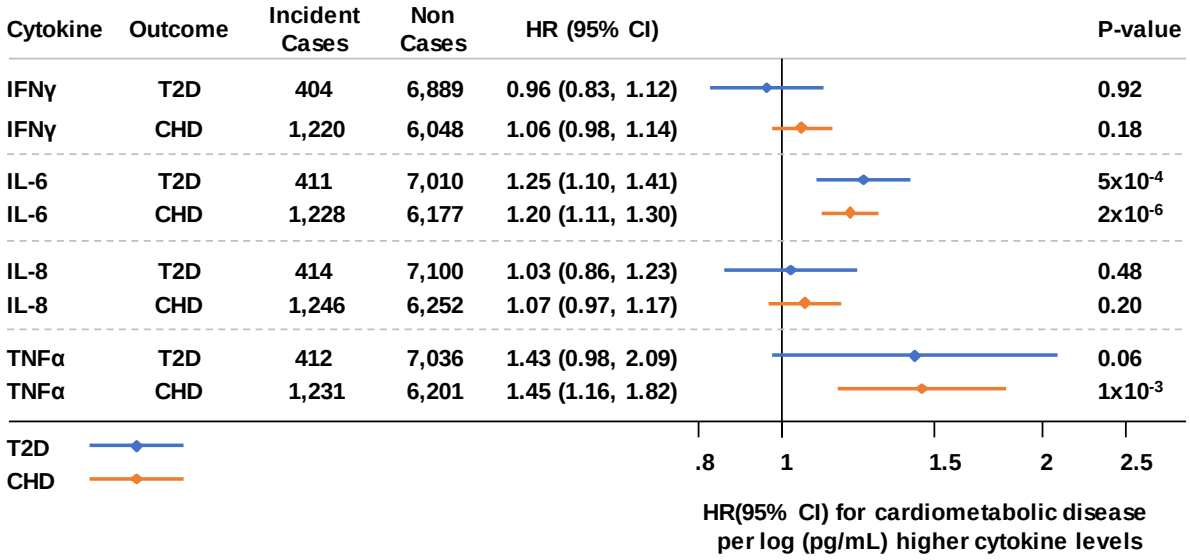
Supplementary Figures

Figure S2.1: Associations of IFNγ, IL-6, IL-8 and TNFα with cardiometabolic risk factors per 1 SD higher cytokine levels, adjusting for fasting insulin levels. Risk factors are grouped into anthropometric, DEXA, glycaemic, lipid and physical activity categories. All traits are adjusted for BMI, additional adjustments for fasting insulin levels are marked as (adj. BMI + Ins) in the trait name. Traits that are significantly associated (P<0.003) with a respective cytokine are coloured in red, non-significant results are in black. Effect estimates are expressed per 1 SD higher cytokine levels.



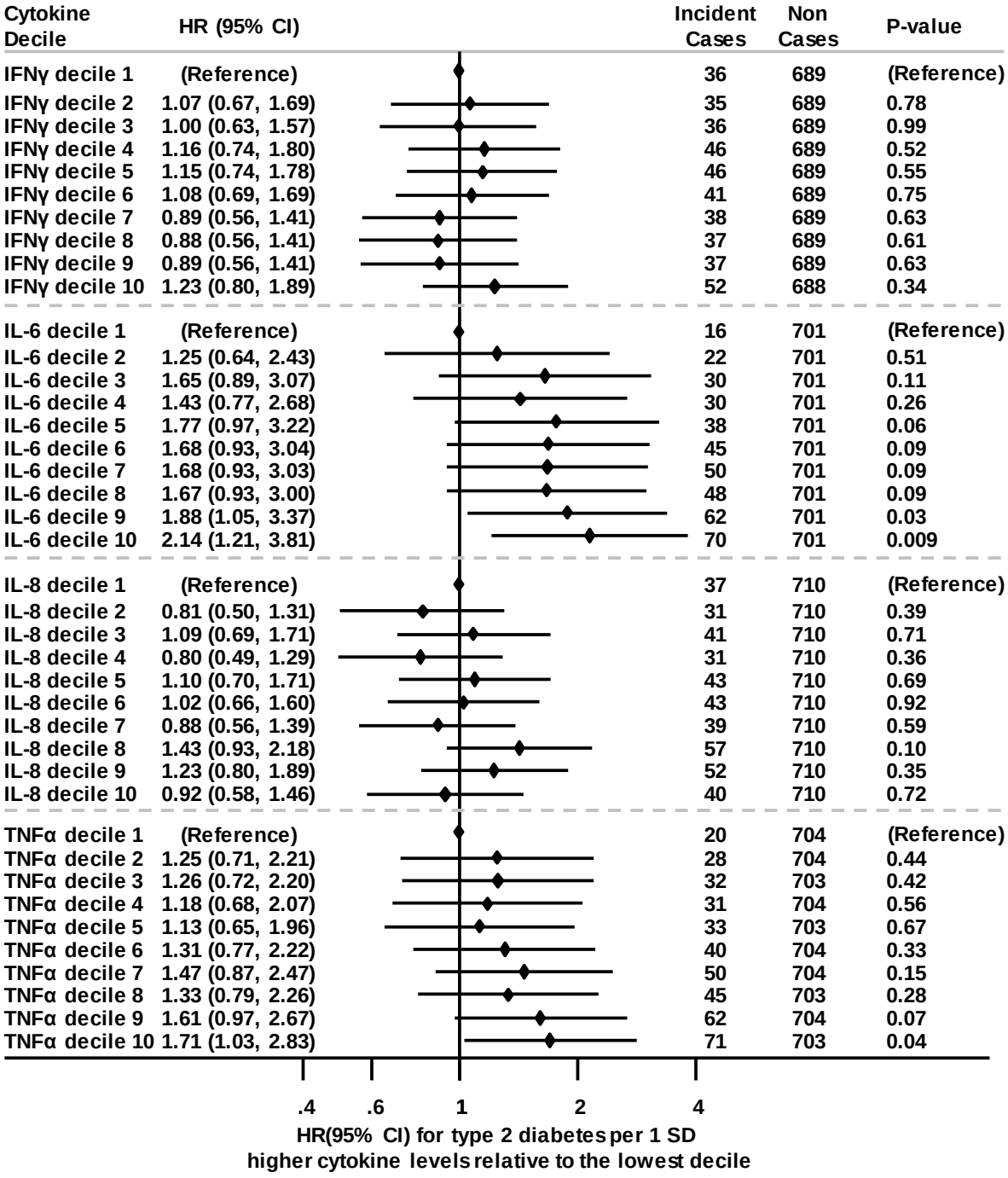
Abbreviations: BMI, Body mass index; adj., Adjusted for; WHR, Waist-to-hip ratio; mmol/L, millimoles per Litre; HbA1c, Glycated haemoglobin; HDL, High-density lipoprotein; LDL, Low-density lipoprotein; g, Grams; Avg., Average; kJ, Kilojoules; Kg, Kilograms; IFNγ, Interferon gamma, IL-6, Interleukin-6; IL-8, Interleukin-8; TNFα, Tumour necrosis factor alpha.

Figure S2.2: Association of natural log transformed cytokine levels with incident cardiometabolic diseases in EPIC-Norfolk. Extensively adjusted results for T2D and CHD are shown in blue and orange respectively. **T2D specific covariates:** age, sex, BMI, WHR, ethnicity, education level, family history of diabetes, smoking status, units of alcohol per week and physical activity. **CHD specific covariates:** age, sex, BMI, WHR, education level, smoking status, physical activity, systolic blood pressure, prevalent diabetes and total cholesterol.



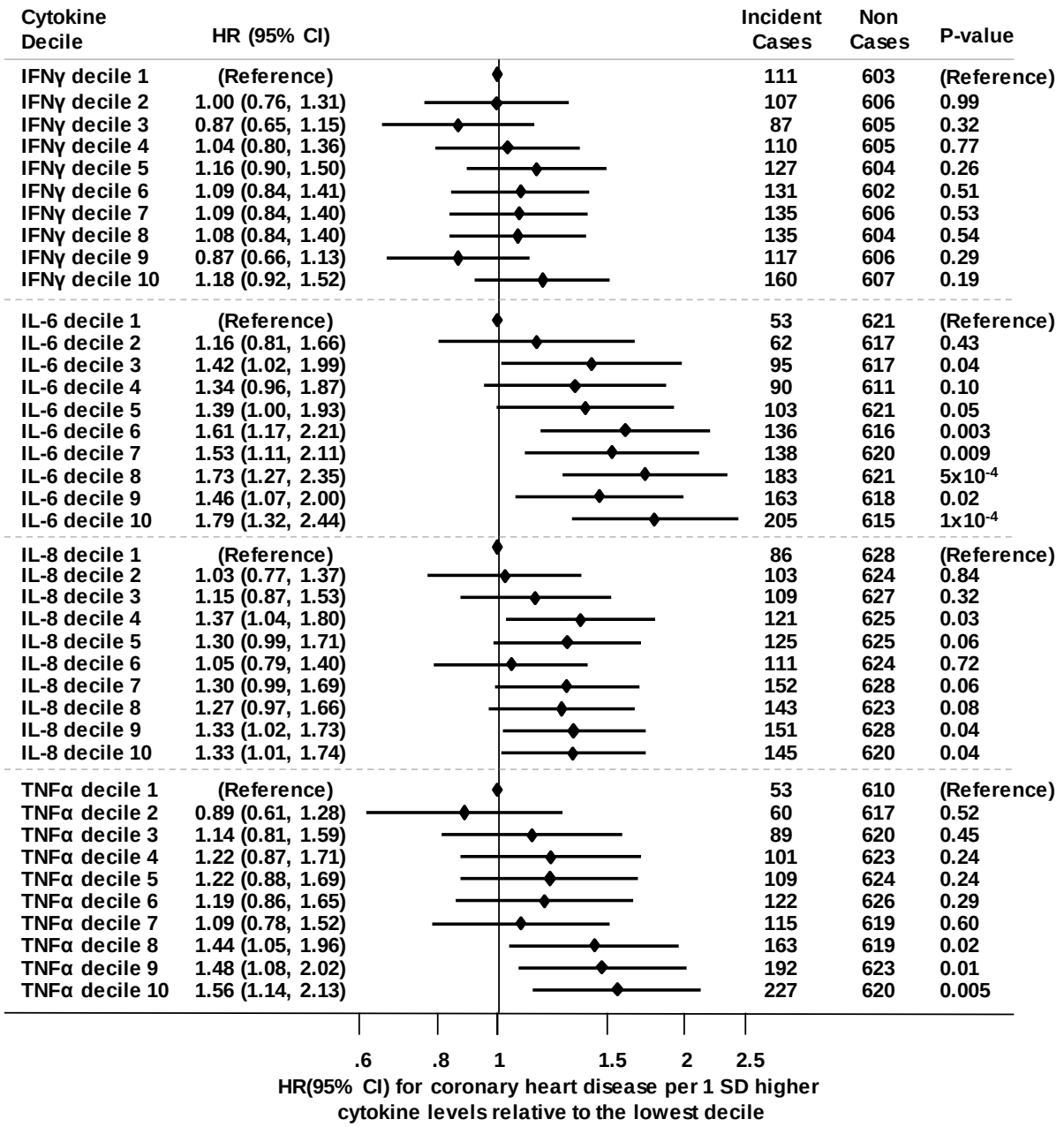
Abbreviations: HR, Hazard ratio; SD, Standard deviation; BMI, Body mass index; WHR, Waist-to-hip ratio; T2D, Type 2 diabetes; CHD, Coronary heart disease; IFNγ, Interferon gamma, IL-6, Interleukin-6; IL-8, Interleukin-8; TNFα, Tumour necrosis factor alpha.

Figure S2.3: Association of deciles of inverse-rank normal transformed cytokine levels with incident T2D in EPIC-Norfolk. Extensively adjusted results for T2D are shown. **T2D specific covariates:** age, sex, BMI, WHR, ethnicity, education level, family history of diabetes, smoking status, units of alcohol per week and physical activity.



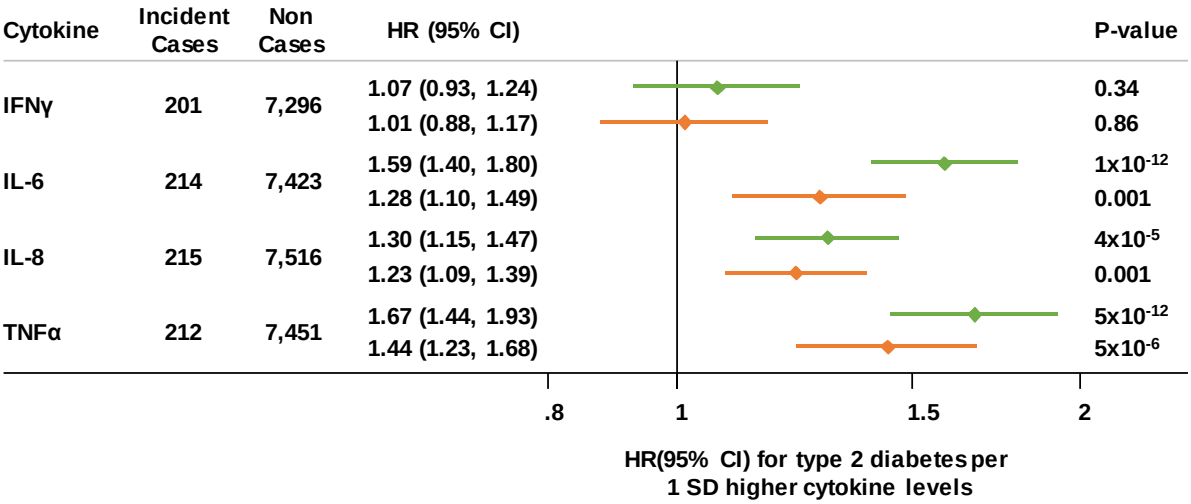
Abbreviations: HR, Hazard ratio; SD, Standard deviation; BMI, Body mass index; WHR, Waist-to-hip ratio; T2D, Type 2 diabetes; IFNγ, Interferon gamma, IL-6, Interleukin-6; IL-8, Interleukin-8; TNFα, Tumour necrosis factor alpha.

Figure S2.4: Association of deciles of inverse-rank normal transformed cytokine levels with incident CHD in EPIC-Norfolk. Extensively adjusted results for T2D are shown. **CHD specific covariates:** age, sex, BMI, WHR, education level, smoking status, physical activity, systolic blood pressure, prevalent diabetes and total cholesterol.



Abbreviations: HR, Hazard ratio; SD, Standard deviation; BMI, Body mass index; WHR, Waist-to-hip ratio; CHD, Coronary heart disease; IFNγ, Interferon gamma, IL-6, Interleukin-6; IL-8, Interleukin-8; TNFα, Tumour necrosis factor alpha.

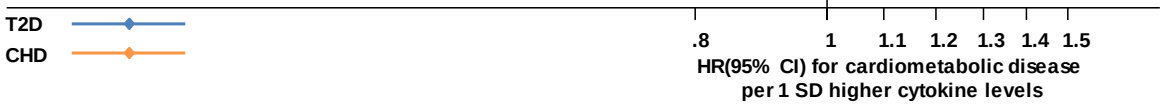
Figure S2.5: Association of cytokine levels with incident T2D, defined using an established definition[248] of incident T2D. Association of inverse-rank normal transformed cytokine levels with incident T2D (defined using an established definition[248] for incident T2D) in EPIC-Norfolk. Results for T2D adjusted for age and sex only are shown in green. Extensively adjusted models are shown in orange. **T2D specific covariates:** age, sex, BMI, WHR, ethnicity, education level, family history of diabetes, smoking status, units of alcohol per week and physical activity.



Abbreviations: HR, Hazard ratio; SD, Standard deviation; BMI, Body mass index; WHR, Waist-to-hip ratio; T2D, Type 2 diabetes; IFNγ, Interferon gamma, IL-6, Interleukin-6; IL-8, Interleukin-8; TNFα, Tumour necrosis factor alpha.

Figure S2.6: Association of inverse-rank normal transformed cytokine levels with incident cardiometabolic diseases in EPIC-Norfolk. Extensively adjusted results for T2D and CHD are shown in blue and orange respectively. In Model 1, all T2D and CHD covariates aside from BMI and WHR were adjusted for. Subsequent models added BMI or WHR separately to the covariates to estimate their independent effects.

Cytokine	Outcome	Adjustment	Incident Cases	Non Cases	HR (95% CI)	P-value
IFNγ	T2D	Model 1	402	6,822	1.03 (0.93, 1.14)	0.59
IFNγ	T2D	Model 1 + BMI	402	6,822	1.00 (0.90, 1.11)	0.96
IFNγ	T2D	Model 1 + WHR	402	6,822	0.99 (0.90, 1.10)	0.89
IFNγ	T2D	Model 1 + BMI + WHR	402	6,822	0.98 (0.89, 1.09)	0.77
IFNγ	CHD	Model 1	1,220	6,048	1.05 (0.99, 1.11)	0.13
IFNγ	CHD	Model 1 + BMI	1,220	6,048	1.04 (0.98, 1.10)	0.19
IFNγ	CHD	Model 1 + WHR	1,220	6,048	1.04 (0.98, 1.10)	0.20
IFNγ	CHD	Model 1 + BMI + WHR	1,220	6,048	1.04 (0.98, 1.10)	0.22
IL-6	T2D	Model 1	409	6,944	1.38 (1.25, 1.52)	$1x10^{-10}$
IL-6	T2D	Model 1 + BMI	409	6,944	1.21 (1.09, 1.34)	$6x10^{-4}$
IL-6	T2D	Model 1 + WHR	409	6,944	1.26 (1.14, 1.40)	$1x10^{-5}$
IL-6	T2D	Model 1 + BMI + WHR	409	6,944	1.18 (1.06, 1.31)	0.003
IL-6	CHD	Model 1	1,228	6,177	1.19 (1.12, 1.27)	$1x10^{-8}$
IL-6	CHD	Model 1 + BMI	1,228	6,177	1.16 (1.09, 1.24)	$2x10^{-6}$
IL-6	CHD	Model 1 + WHR	1,228	6,177	1.18 (1.11, 1.25)	$2x10^{-7}$
IL-6	CHD	Model 1 + BMI + WHR	1,228	6,177	1.16 (1.09, 1.24)	$3x10^{-6}$
IL-8	T2D	Model 1	412	7,032	1.08 (0.98, 1.20)	0.11
IL-8	T2D	Model 1 + BMI	412	7,032	1.07 (0.97, 1.18)	0.20
IL-8	T2D	Model 1 + WHR	412	7,032	1.07 (0.97, 1.18)	0.19
IL-8	T2D	Model 1 + BMI + WHR	412	7,032	1.06 (0.96, 1.17)	0.24
IL-8	CHD	Model 1	1,246	6,252	1.08 (1.02, 1.14)	0.01
IL-8	CHD	Model 1 + BMI	1,246	6,252	1.07 (1.01, 1.13)	0.02
IL-8	CHD	Model 1 + WHR	1,246	6,252	1.07 (1.01, 1.13)	0.02
IL-8	CHD	Model 1 + BMI + WHR	1,246	6,252	1.07 (1.01, 1.13)	0.02
TNFα	T2D	Model 1	410	6,969	1.29 (1.16, 1.43)	$3x10^{-6}$
TNFα	T2D	Model 1 + BMI	410	6,969	1.18 (1.05, 1.31)	0.004
TNFα	T2D	Model 1 + WHR	410	6,969	1.21 (1.08, 1.35)	$6x10^{-4}$
TNFα	T2D	Model 1 + BMI + WHR	410	6,969	1.16 (1.04, 1.29)	0.01
TNFα	CHD	Model 1	1,231	6,201	1.19 (1.11, 1.26)	$9x10^{-8}$
TNFα	CHD	Model 1 + BMI	1,231	6,201	1.16 (1.09, 1.24)	$2x10^{-6}$
TNFα	CHD	Model 1 + WHR	1,231	6,201	1.17 (1.10, 1.25)	$7x10^{-7}$
TNFα	CHD	Model 1 + BMI + WHR	1,231	6,201	1.16 (1.09, 1.24)	$4x10^{-6}$

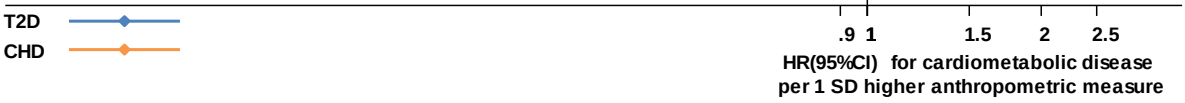


Model 1
CHD: Age, sex, education level, smoking status, avg. physical activity per week, systolic blood pressure, prevalent T2D, total cholesterol
T2D: Age, sex, education level , ethnicity, family history of T2D, smoking status, avg. physical activity per week, avg. units of alcohol per week

Abbreviations: HR, Hazard ratio; SD, Standard deviation; BMI, Body mass index; WHR, Waist-to-hip ratio; T2D, Type 2 diabetes; CHD, Coronary heart disease; IFNγ, Interferon gamma, IL-6, Interleukin-6; IL-8, Interleukin-8; TNFα, Tumour necrosis factor alpha.

Figure S2.7: Association of BMI, WHR and WHRadjBMI with incident cardiometabolic diseases in EPIC-Norfolk. Extensively adjusted results for T2D and CHD are shown in blue and orange respectively. In Model 1, all T2D and CHD covariates aside from BMI and WHR were adjusted for. Subsequent models added IL-6 or TNFα separately to the covariates to estimate their independent effects.

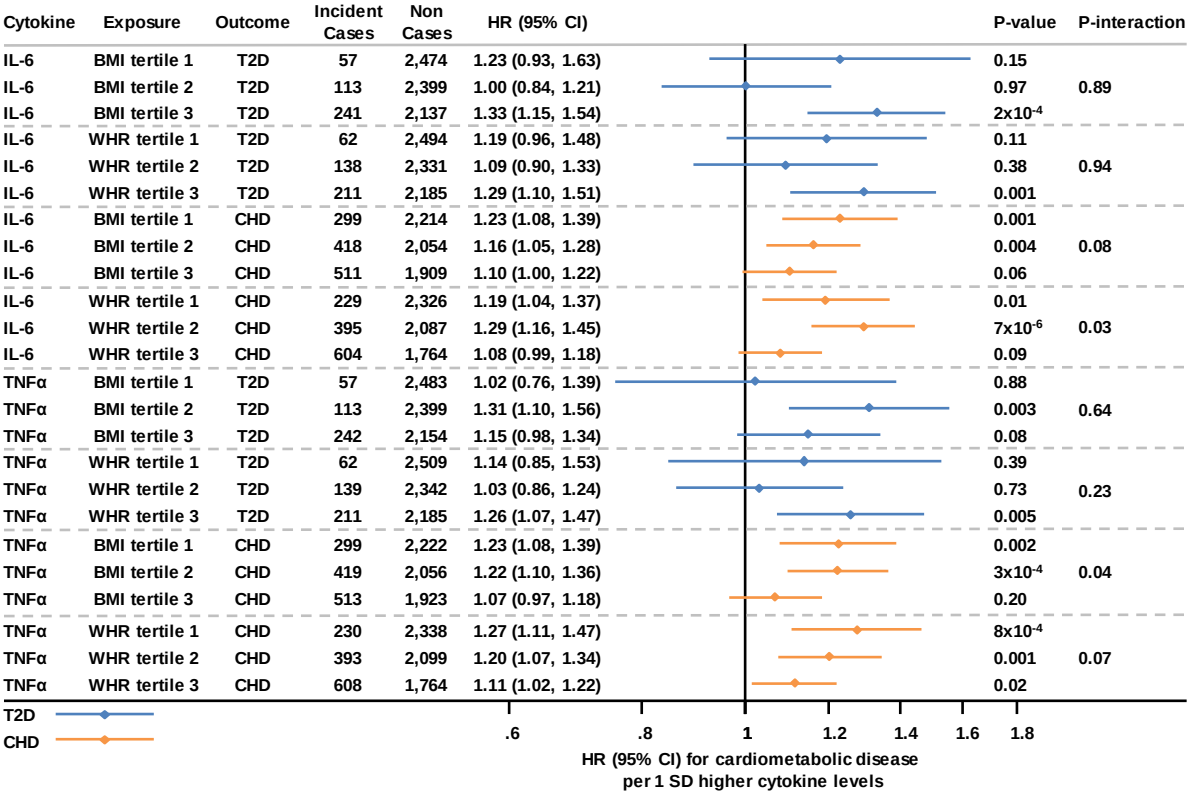
Anthropometric Measure	Outcome	Adjustment	Incident Cases	Non Cases	HR (95% CI)	P-value
BMI	T2D	Model 1	407	6,880	1.92 (1.72, 2.15)	$1x10^{-30}$
BMI	T2D	Model 1 + IL-6	407	6,880	1.84 (1.65, 2.07)	$2x10^{-26}$
BMI	T2D	Model 1 + TNFα	407	6,880	1.88 (1.68, 2.11)	$4x10^{-28}$
BMI	T2D	Model 1 + IL-6 + TNFα	407	6,880	1.83 (1.63, 2.05)	$3x10^{-25}$
BMI	CHD	Model 1	1,213	6,125	1.17 (1.09, 1.25)	$4x10^{-6}$
BMI	CHD	Model 1 + IL-6	1,213	6,125	1.13 (1.06, 1.21)	$3x10^{-4}$
BMI	CHD	Model 1 + TNFα	1,213	6,125	1.14 (1.07, 1.22)	$7x10^{-5}$
BMI	CHD	Model 1 + IL-6 + TNFα	1,213	6,125	1.12 (1.05, 1.20)	$9x10^{-4}$
WHR	T2D	Model 1	407	6,880	2.13 (1.85, 2.46)	$3x10^{-25}$
WHR	T2D	Model 1 + IL-6	407	6,880	2.01 (1.74, 2.32)	$2x10^{-21}$
WHR	T2D	Model 1 + TNFα	407	6,880	2.08 (1.80, 2.40)	$3x10^{-23}$
WHR	T2D	Model 1 + IL-6 + TNFα	407	6,880	2.00 (1.73, 2.31)	$9x10^{-21}$
WHR	CHD	Model 1	1,213	6,125	1.17 (1.07, 1.28)	$7x10^{-4}$
WHR	CHD	Model 1 + IL-6	1,213	6,125	1.13 (1.03, 1.24)	0.008
WHR	CHD	Model 1 + TNFα	1,213	6,125	1.14 (1.04, 1.25)	0.005
WHR	CHD	Model 1 + IL-6 + TNFα	1,213	6,125	1.12 (1.02, 1.23)	0.02
WHRadjBMI	T2D	Model 1	371	6,606	1.62 (1.37, 1.91)	$1x10^{-8}$
WHRadjBMI	T2D	Model 1 + IL-6	371	6,606	1.58 (1.34, 1.86)	$4x10^{-8}$
WHRadjBMI	T2D	Model 1 + TNFα	371	6,606	1.61 (1.36, 1.90)	$2x10^{-8}$
WHRadjBMI	T2D	Model 1 + IL-6 + TNFα	371	6,606	1.58 (1.34, 1.86)	$5x10^{-8}$
WHRadjBMI	CHD	Model 1	1,213	6,125	1.08 (0.98, 1.20)	0.11
WHRadjBMI	CHD	Model 1 + IL-6	1,213	6,125	1.07 (0.97, 1.18)	0.20
WHRadjBMI	CHD	Model 1 + TNFα	1,213	6,125	1.07 (0.97, 1.18)	0.20
WHRadjBMI	CHD	Model 1 + IL-6 + TNFα	1,213	6,125	1.06 (0.96, 1.17)	0.26



Model 1:
CHD: Age, sex, education level, smoking status, avg. physical activity per week, systolic blood pressure, prevalent T2D, total cholesterol
T2D: Age, sex, education level, ethnicity, family history of T2D, smoking status, avg. units of alcohol per week, avg. physical activity per week

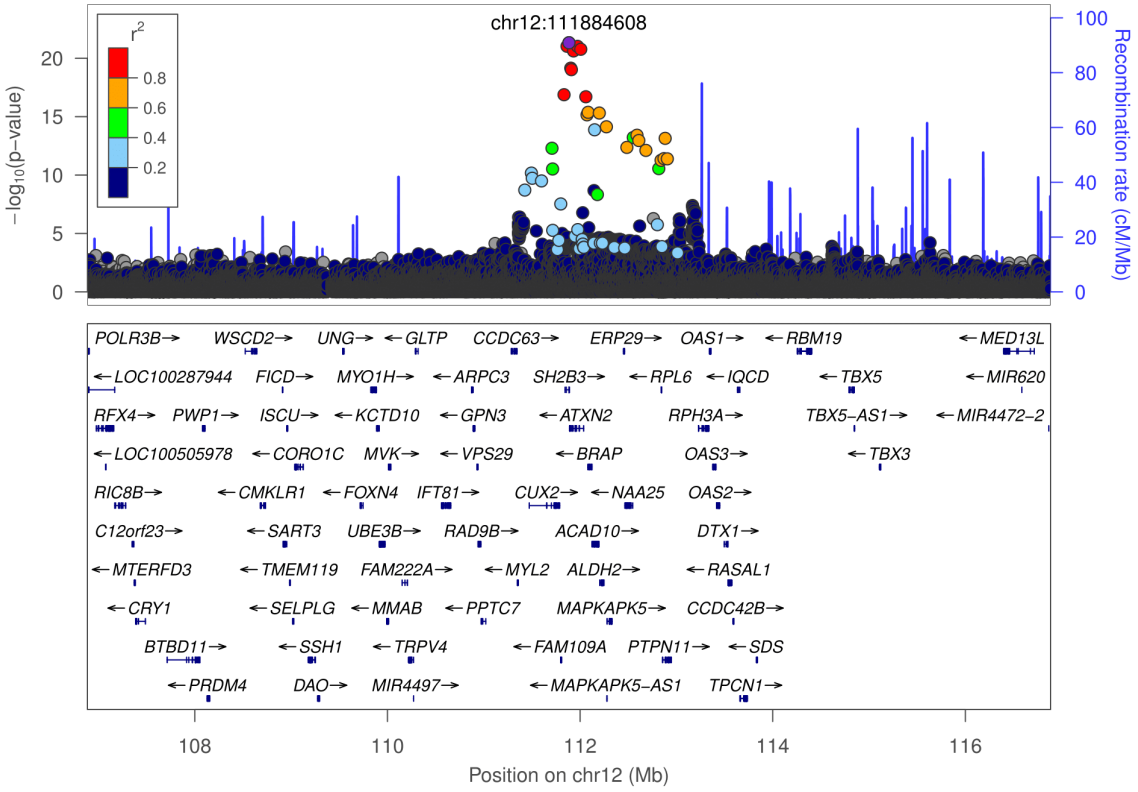
Abbreviations: HR, Hazard ratio; SD, Standard deviation; BMI, Body mass index; WHR, Waist-to-hip ratio; WHRadjBMI, Waist-to-hip ratio adjusted for BMI; T2D, Type 2 diabetes; CHD, Coronary heart disease; IL-6, Interleukin-6; TNFα, Tumour necrosis factor alpha.

Figure S2.8: Interaction between cytokine levels and anthropometric traits on incident cardiometabolic diseases. Shown is the interaction between inverse-rank normal transformed cytokine levels and anthropometric traits on the association with incident cardiometabolic diseases in EPIC-Norfolk. Extensively adjusted results for T2D and CHD are shown in blue and orange respectively within tertiles of either BMI or WHR. **T2D specific covariates:** age, sex, BMI, WHR, ethnicity, education level, family history of diabetes, smoking status, units of alcohol per week and physical activity. **CHD specific covariates:** age, sex, BMI, WHR, education level, smoking status, physical activity, systolic blood pressure, prevalent diabetes and total cholesterol.



Abbreviations: HR, Hazard ratio; SD, Standard deviation; BMI, Body mass index; WHR, Waist-to-hip ratio; T2D, Type 2 diabetes; CHD, Coronary heart disease; IL-6, Interleukin-6; TNFα, Tumour necrosis factor alpha.

Figure S3.1: Regional association plot depicting the 10Mb region of chromosome 12 associated with TNFα. This is a 5Mb region either side of rs3184504 (chr12:111884608) in *SH2B3* used for independent signal selection and conditional analysis. The extensive LD in the region is shown, variants in very high LD ($R^2 \geq 0.8$) with rs3184504 are shown in red. Plot drawn using LocusZoom v1.2[278].



Abbreviations: chr12:111884608, rs3184504; TNFα, Tumour necrosis factor α; LD, Linkage disequilibrium; *SH2B3*, SH2B Adaptor Protein 3; Mb, Megabase.

Figure S3.2: Quantile-quantile plots for the respective cytokine meta-analyses. The inflation of association test statistics compared to what is expected under the null hypothesis. The IL-6 and TNFα plots include only the Fenland and EPIC-Norfolk datasets. Data from the CoLaus population study is available for these two cytokines but was not included for this analysis. **A.** IFNγ, **B.** IL-6, **C.** IL-8, **D.** TNFα.

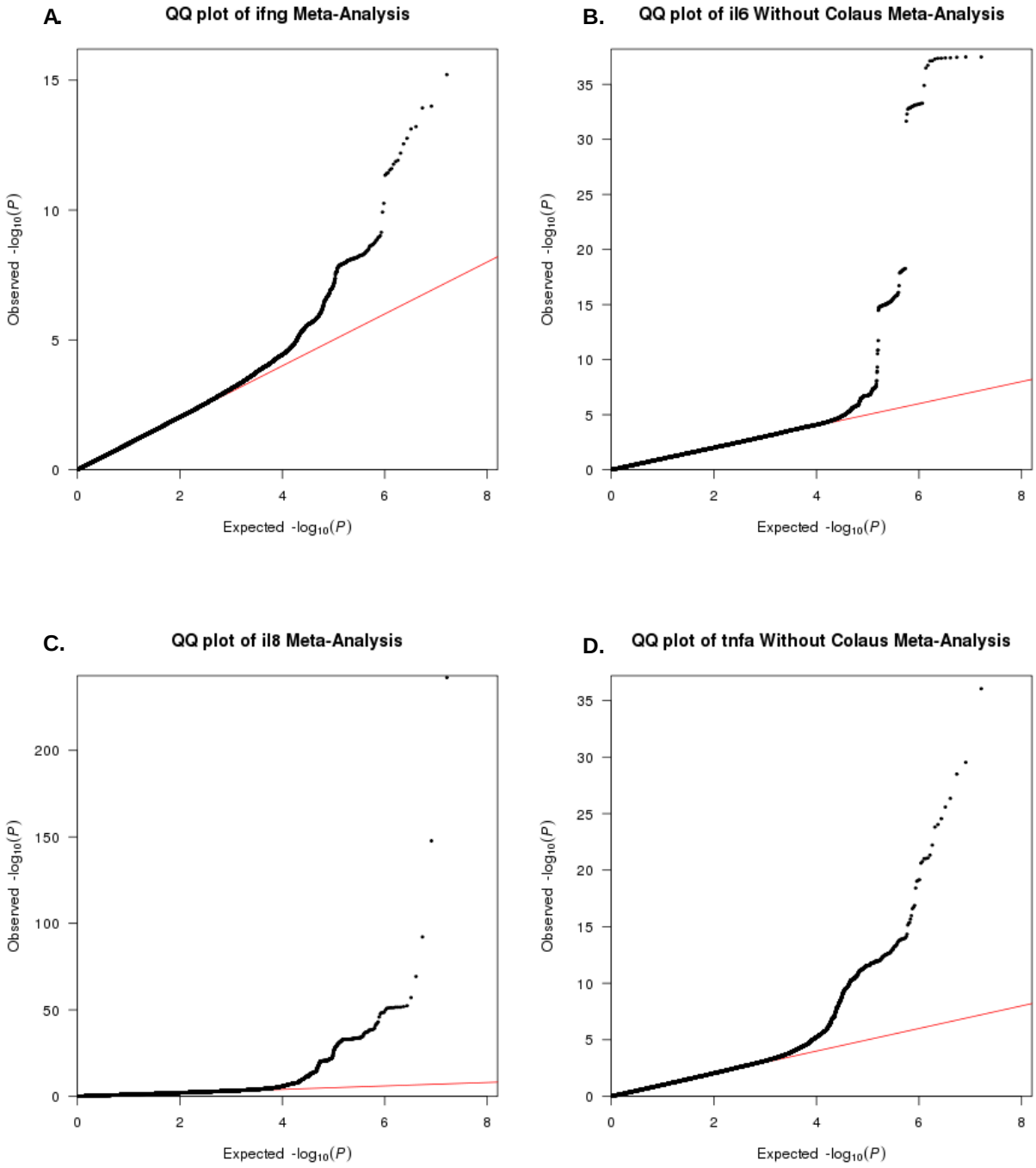


Figure S4.1: Design of the observational and genetic mediation analyses. The cohorts in which each coefficient was estimated are reported in the inner part of each panel. **Panel A** illustrates the design of the observational mediation analysis, estimated in the EPIC-Norfolk cohort. **Panels B and C** illustrate the design of the genetic mediation analyses, for cardiometabolic diseases and fasting insulin levels respectively estimated in the EPIC-Norfolk, Fenland UK Biobank and MAGIC cohorts.

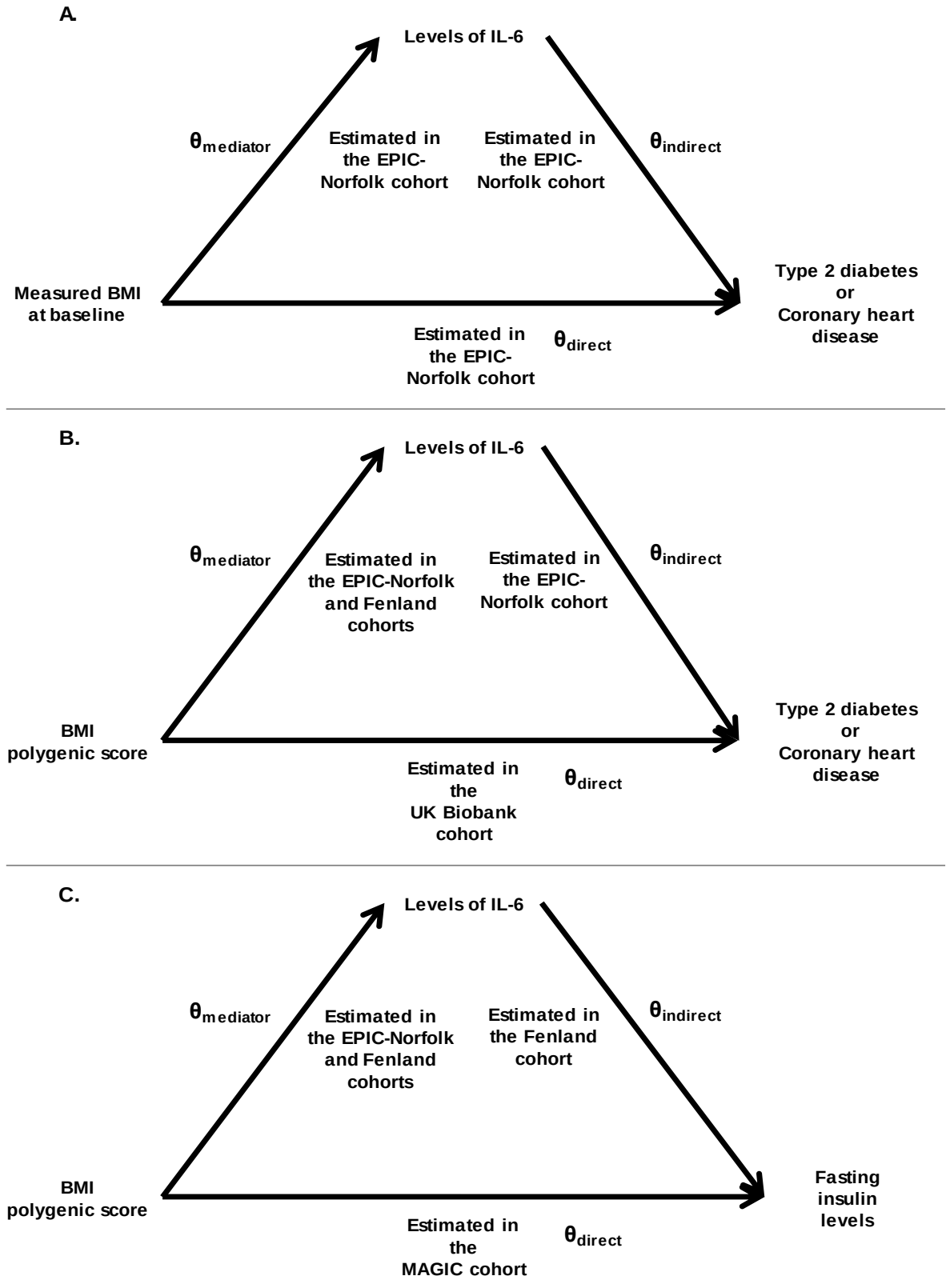
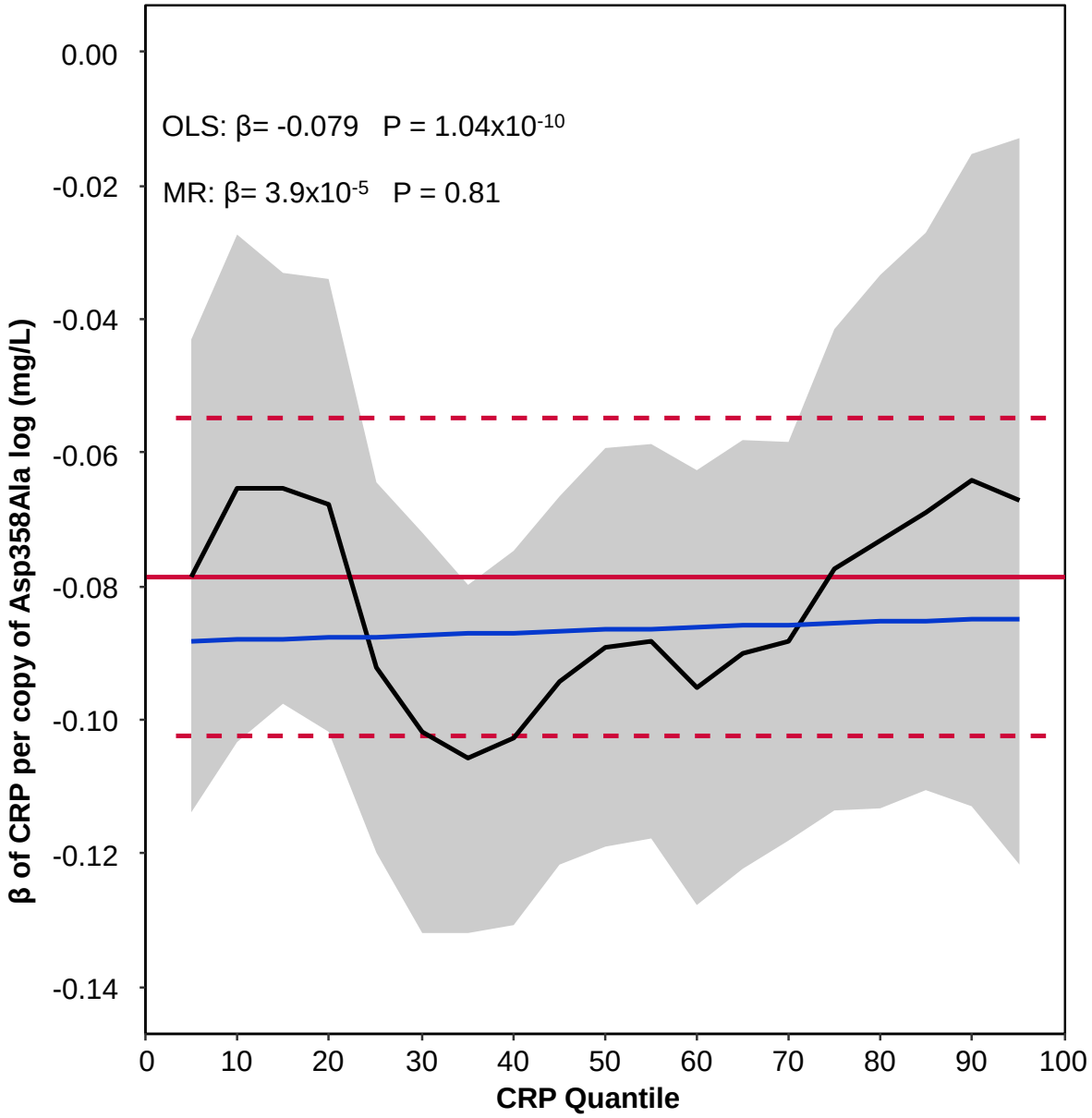
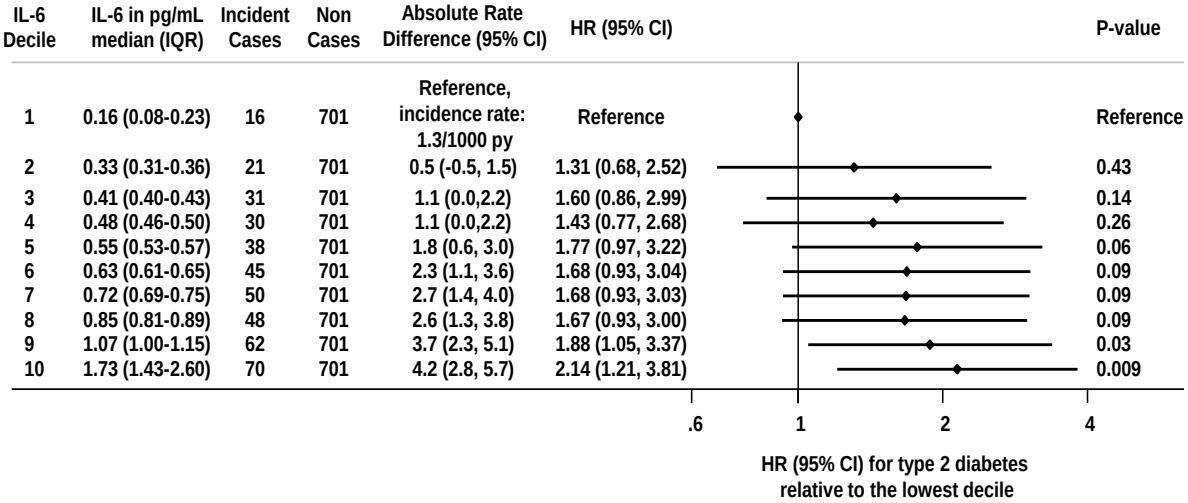


Figure S4.2: Conditional quantile regression of CRP levels in EPIC-Norfolk. The association of 358Ala with ln-CRP levels was estimated in EPIC-Norfolk. The ordinary least squares (OLS) regression line is shown in red with the corresponding 95% confidence interval denoted by the dashed red lines. The OLS beta estimate and p-value are provided within the plot. The beta estimates from conditional quantile regression per ln-CRP quantile are shown by the black line and the corresponding 95% confidence interval is shown by the grey ribbon. The blue line shows the meta-regression line. The meta-regression estimate, and P-value are provided within the plot. No statistically significant difference in the effect of Asp358Ala across CRP quantiles is shown by the meta-regression model.



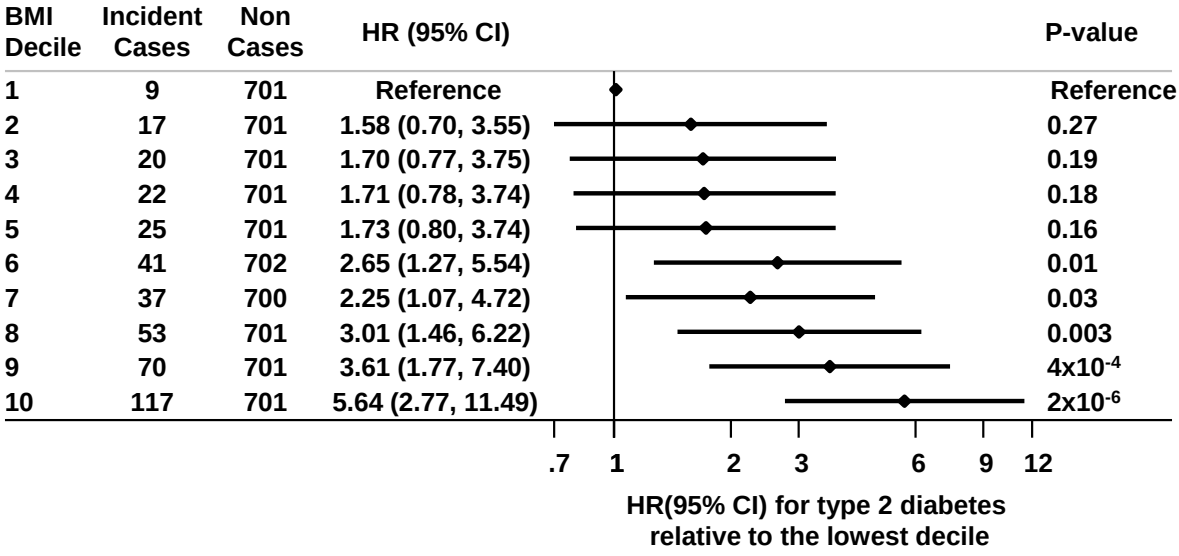
Abbreviations: OLS, Ordinary least squares; MR, Meta-regression; CRP, C-reactive protein; mg, milligrams; L, Litre; Asp, Aspartic acid; Ala, Alanine.

Figure S4.3: Association of IL-6 levels in deciles with incident type 2 diabetes in EPIC-Norfolk. Estimates were adjusted for age, sex, BMI, WHR, ethnicity, education level, family history of type 2 diabetes, smoking status, average units of alcohol per week and average physical activity per week. Hazard ratios for type 2 diabetes were estimated relative to the lowest decile.



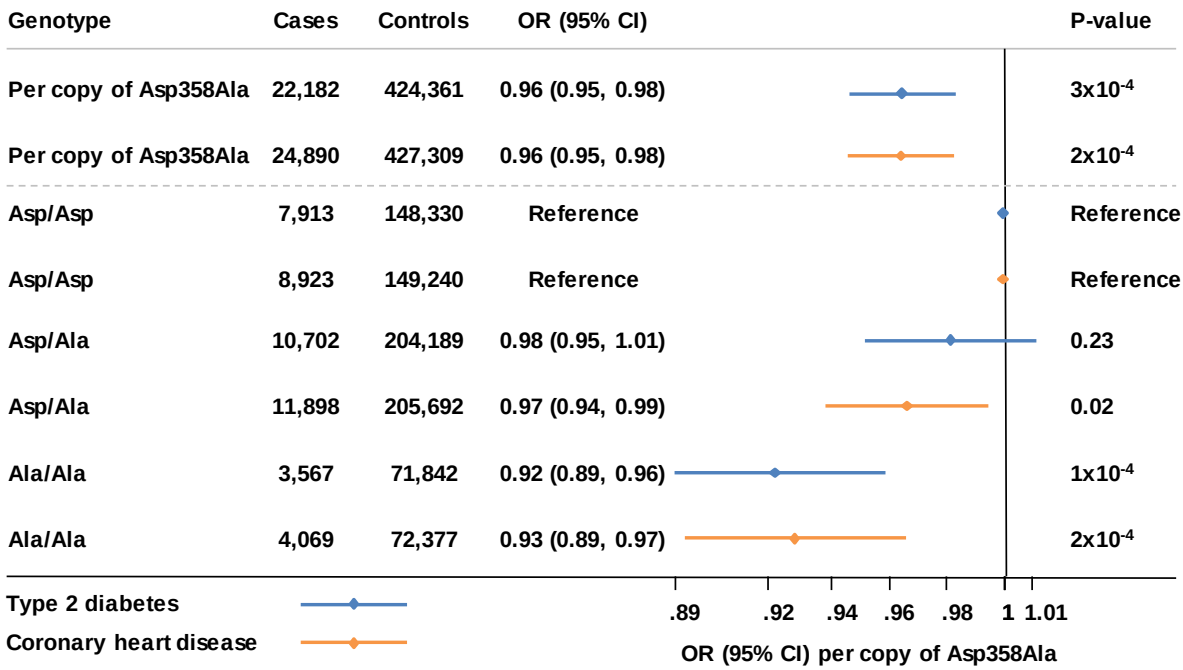
Abbreviations: HR, Hazard ratio; CI, Confidence interval; BMI, Body mass index; WHR, Waist-to-hip ratio; T2D, Type 2 diabetes; Avg., Average; pg/mL, picograms per millilitre; IL-6, Interleukin 6.

Figure S4.4: Association of BMI deciles with incident type 2 diabetes in EPIC-Norfolk. The association of BMI deciles with incident type 2 diabetes was estimated using multivariable Cox regression in the EPIC-Norfolk study. Age, sex, IL-6 levels, WHR, ethnicity, education level, family history of type 2 diabetes, smoking status, average physical activity per week, and average units of alcohol per week were covariates in the models. Hazard ratios for type 2 diabetes were estimated per kg/m^2 higher BMI, relative to the lowest decile.



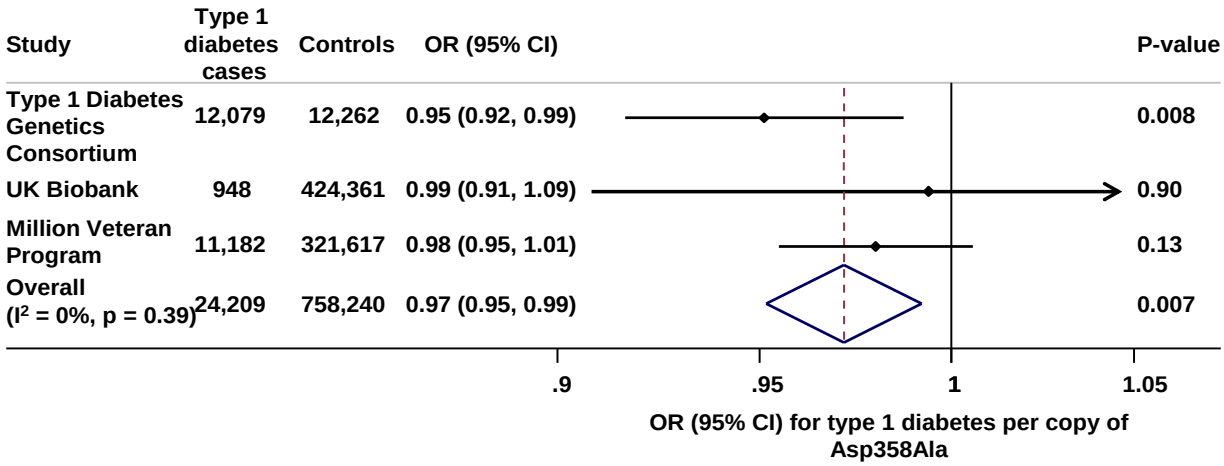
Abbreviations: HR, Hazard ratio; CI, Confidence interval; BMI, Body mass index; WHR, Waist-to-hip ratio; IL-6, Interleukin 6.

Figure S4.5: Association of Asp358Ala with type 2 diabetes and coronary heart disease by genotype in UK Biobank. Type 2 diabetes results are shown in blue and those of coronary heart disease in orange. Multivariable logistic regression models were adjusted for age, sex, genotyping chip, the first 40 principal components and a kinship matrix. Odds ratios were estimated per copy of Asp358Ala relative to the wild-type Asp/Asp genotype. No statistical evidence of deviation from an additive association was observed for type 2 diabetes ($P_{non\text{-}linearity}$ = 0·13) or coronary heart disease ($P_{non\text{-}linearity}$ = 0·84).



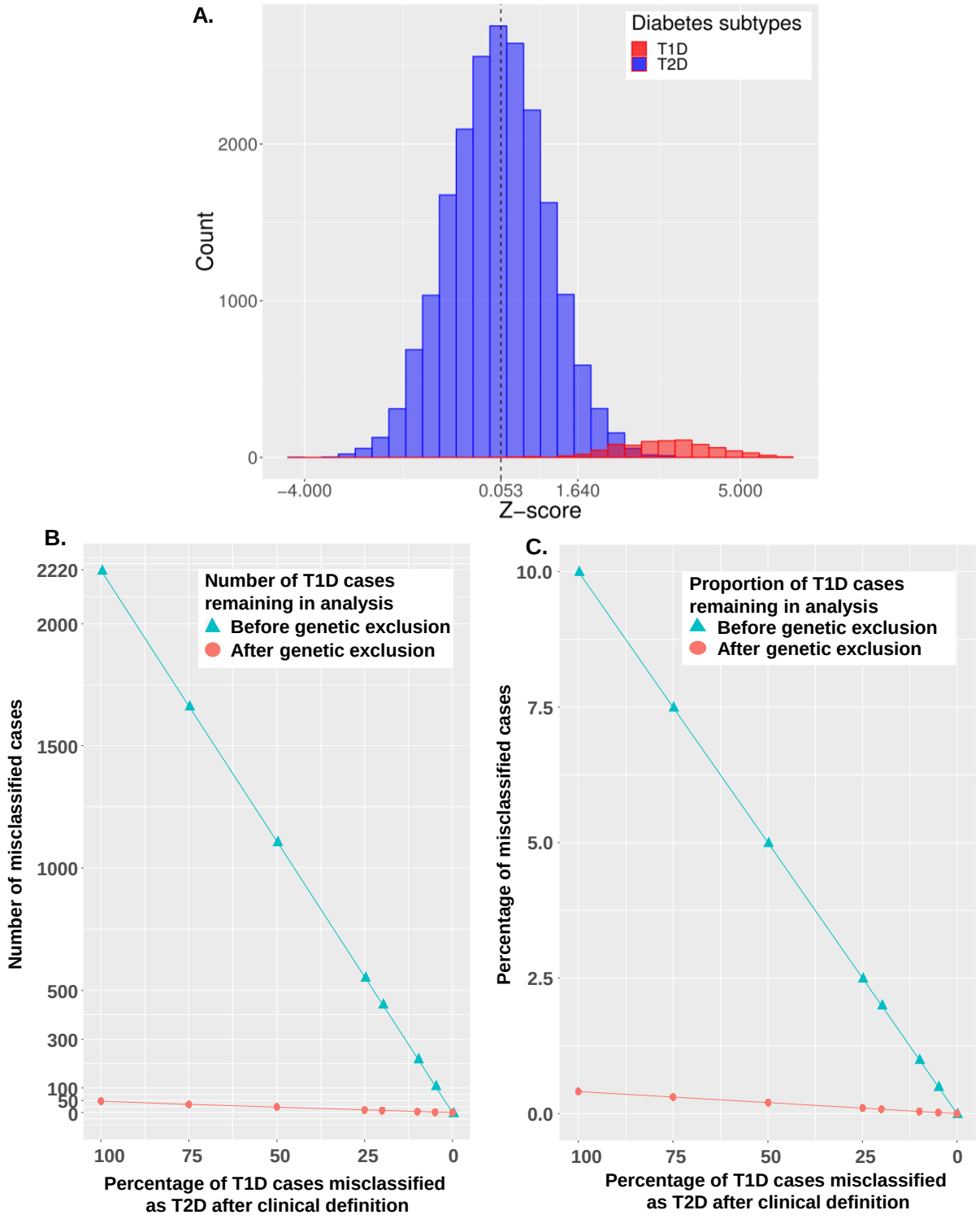
Abbreviations: T2D, Type 2 diabetes; CHD, Coronary heart disease; OR, Odds ratio; CI, Confidence interval; Asp, Aspartic acid; Ala, Alanine.

Figure S4.6: Association of Asp358Ala with type 1 diabetes. In UK Biobank, the association of Asp358Ala with type 1 diabetes was estimated using multivariable logistic regression adjusted for age, sex, genotyping chip, the first 40 principal components and a kinship matrix. The result was meta-analysed with summary statistics from the type 1 diabetes genetics consortium and the Million Veteran Program respectively. Odds ratios for type 1 diabetes were estimated per copy of Asp358Ala. It should be noted that despite the use of the largest available studies and the sufficient power of this analysis, there may too few studies to make definitive statistical conclusions of the association between Asp358Ala and type 1 diabetes.



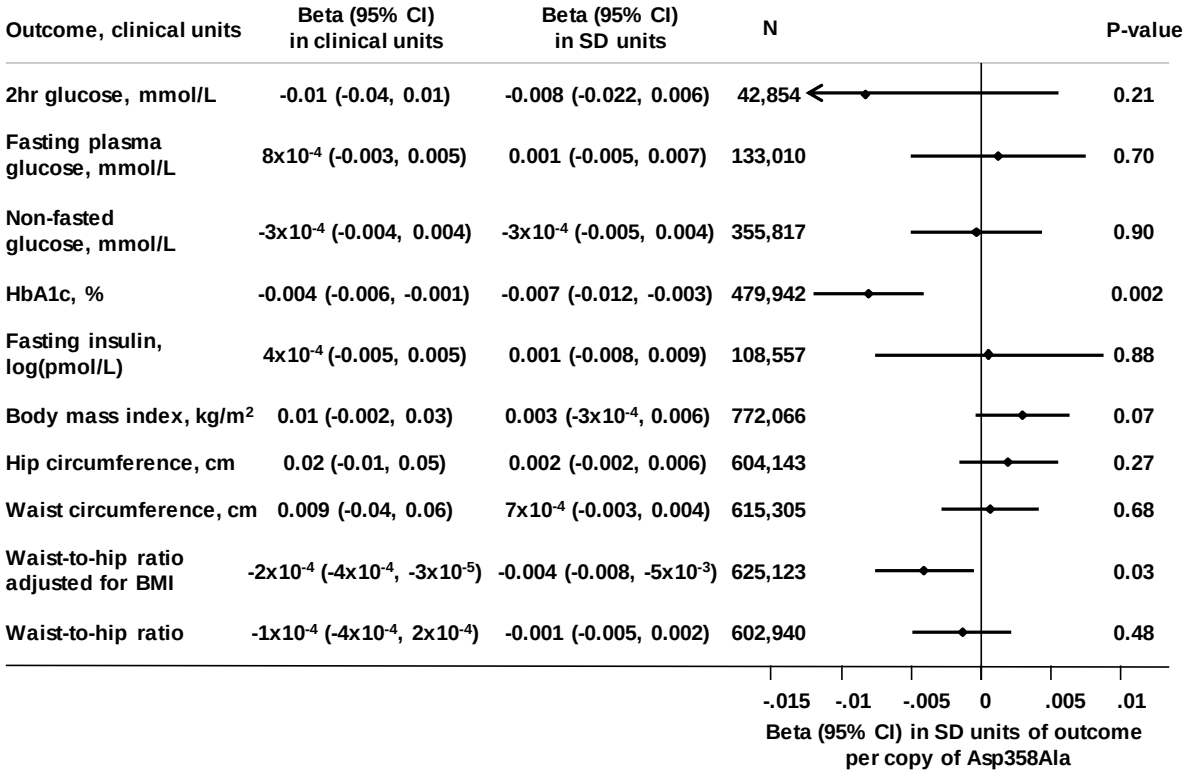
Abbreviations: OR, Odds ratio; CI, Confidence interval; Asp, Aspartic acid; Ala, Alanine.

Figure S4.7: Excluding type 1 diabetes cases using a 29-variant polygenic risk score. Panel A. Theoretical distribution of the polygenic score in diabetes cases·[350] Type 2 diabetes cases are shown in blue and type 1 diabetes cases in red. A z-score of 0·053 (actual value in UK Biobank, 13·04) denotes the median of the overall diabetes distribution (**Table 2**). **Panels B and C** show the number and percentage, respectively, of type 1 diabetes cases misclassified as type 2 diabetes that would be included in the analysis before and after applying the polygenic score exclusion. These simulations assume a 10% prevalence of type 1 diabetes amongst people with diabetes in the UK Biobank study[350] and simulate percentage misclassification ranging from 100% to 0%.



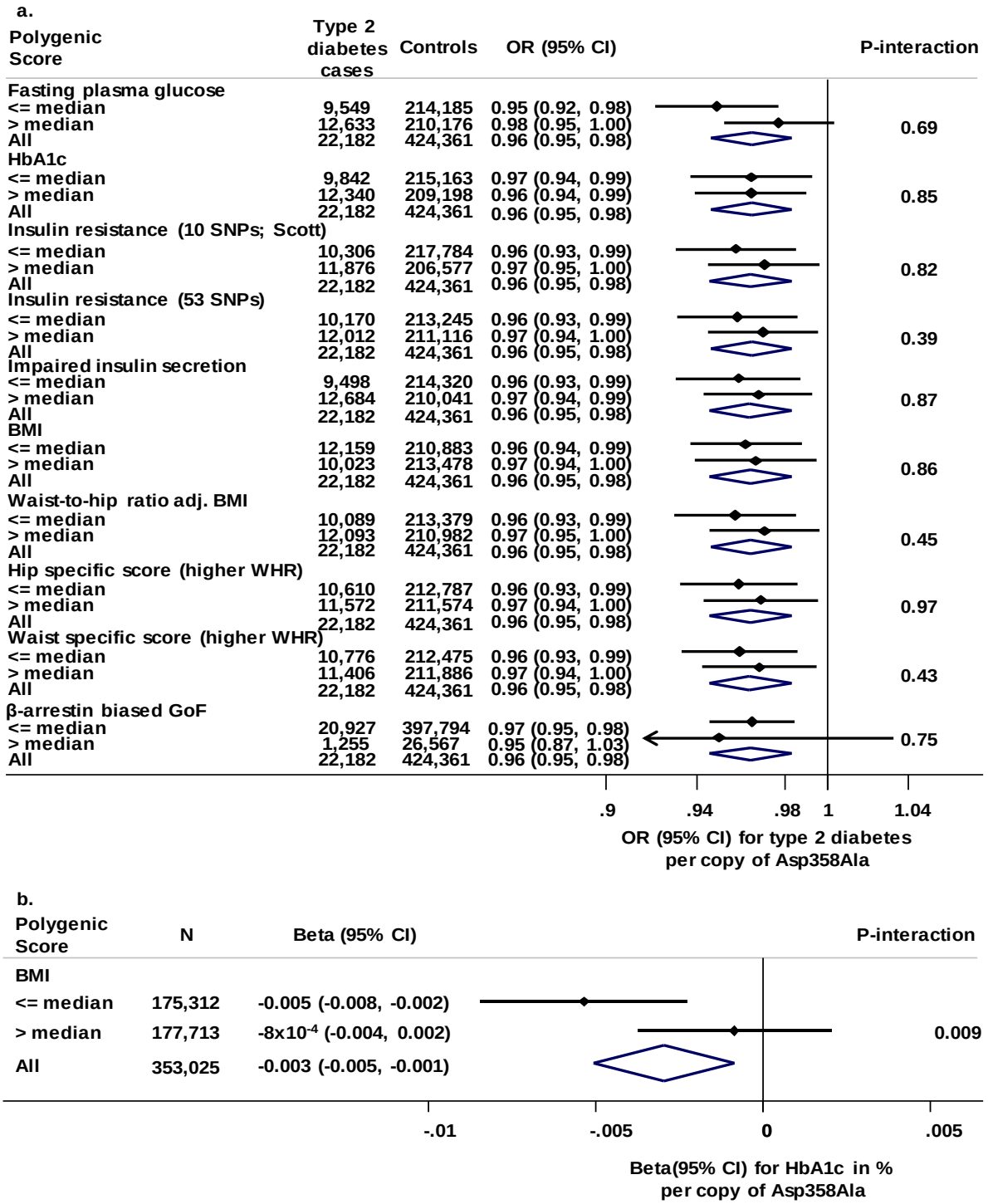
Abbreviations: T1D, Type 1 diabetes; T2D, Type 2 diabetes.

Figure S4.8: Association of Asp358Ala with continuous metabolic traits. A Bonferroni corrected significance threshold of P<·005 (·05 / 10 continuous traits) was used.



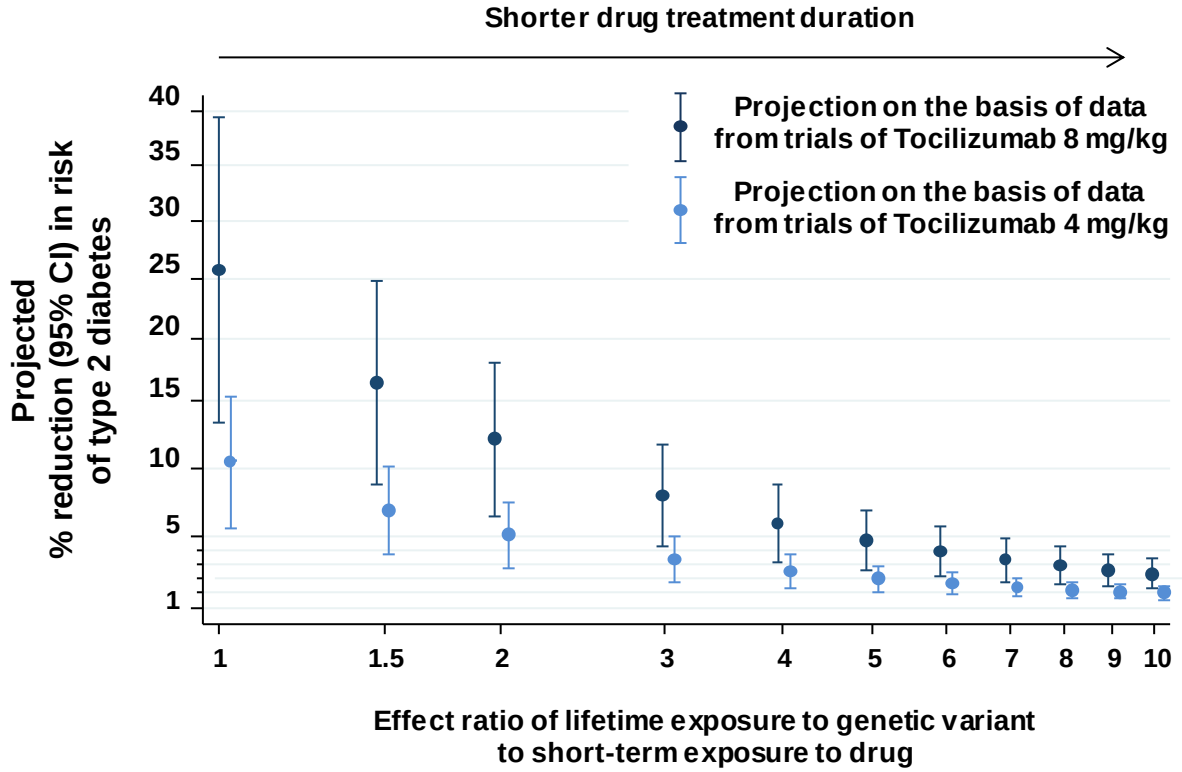
Abbreviations: HbA1c, glycated haemoglobin; BMI, Body mass index; WHR, Waist-to-hip ratio; WHRadjBMI, WHR adjusted for BMI; mmol, millimoles; L, Litre; pmol, picomoles; Kg, Kilograms; m, meters; cm, centimetres Asp, Aspartic acid; Ala, Alanine; N, Number of participants.

Figure S4.9: Interaction of Asp358Ala with type 2 diabetes genetic risk factors on type 2 diabetes. Panel a. The association of *IL6R* Asp358Ala with type 2 diabetes was estimated in individuals above or below the median value for 10 polygenic scores capturing genetic predisposition to risk factors for type 2 diabetes. The association with type 2 diabetes from the main analysis is shown below each score for comparison. A Bonferroni corrected significance threshold of P<·005 (·05/10 tests for heterogeneity) for the heterogeneity P-value was used. **Panel b.** The association of *IL6R* Asp358Ala with HbA1c levels was estimated in individuals above or below the median value for a polygenic score capturing genetic predisposition to higher BMI. The association with HbA1c in UK Biobank is shown for comparison. Estimates for HbA1c are expressed in % per copy of Asp358Ala.



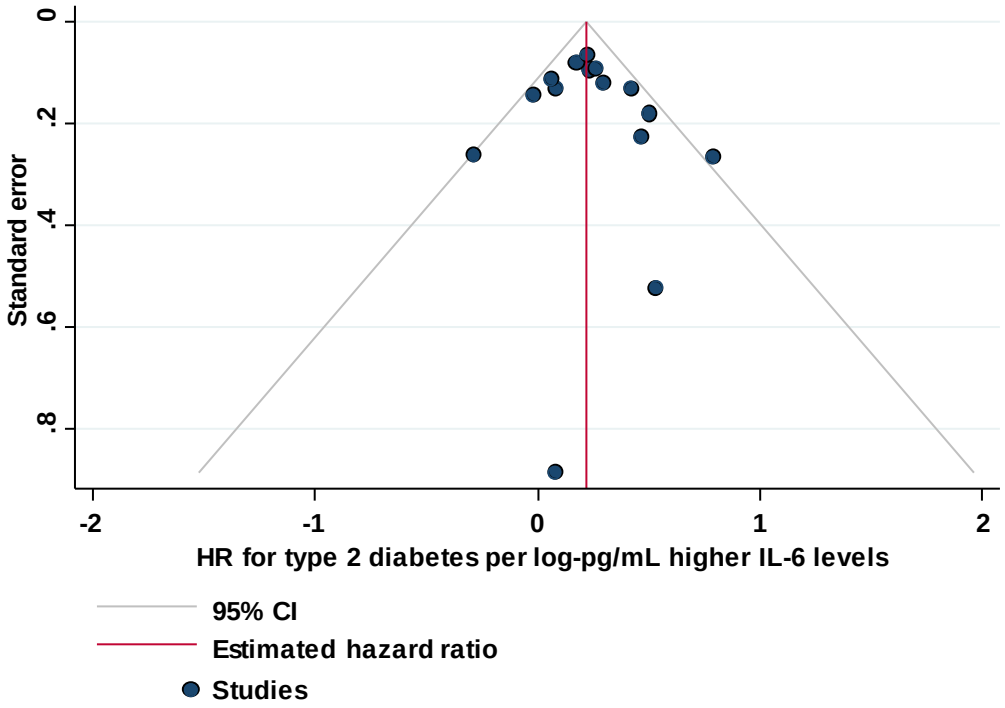
Abbreviations: HbA1c, glycated haemoglobin; SNPs, Single nucleotide polymorphisms; BMI, Body mass index; adj., adjusted for; GoF, Gain-of-function; OR, Odds ratio

Figure S4.10: Projection of the range of possible effects of *IL6R* blocking therapy on the risk of type 2 diabetes in a primary prevention setting. The ratio between lifetime exposure (genetic association) to short-term exposure (clinical trial) to lower IL-6R-mediated inflammation is on the x-axis. Lower ratio values correspond to estimates of longer treatment duration. The projected % reduction in diabetes risk, projected based on data from clinical trials of different dosages of tocilizumab, is represented on the y-axis.



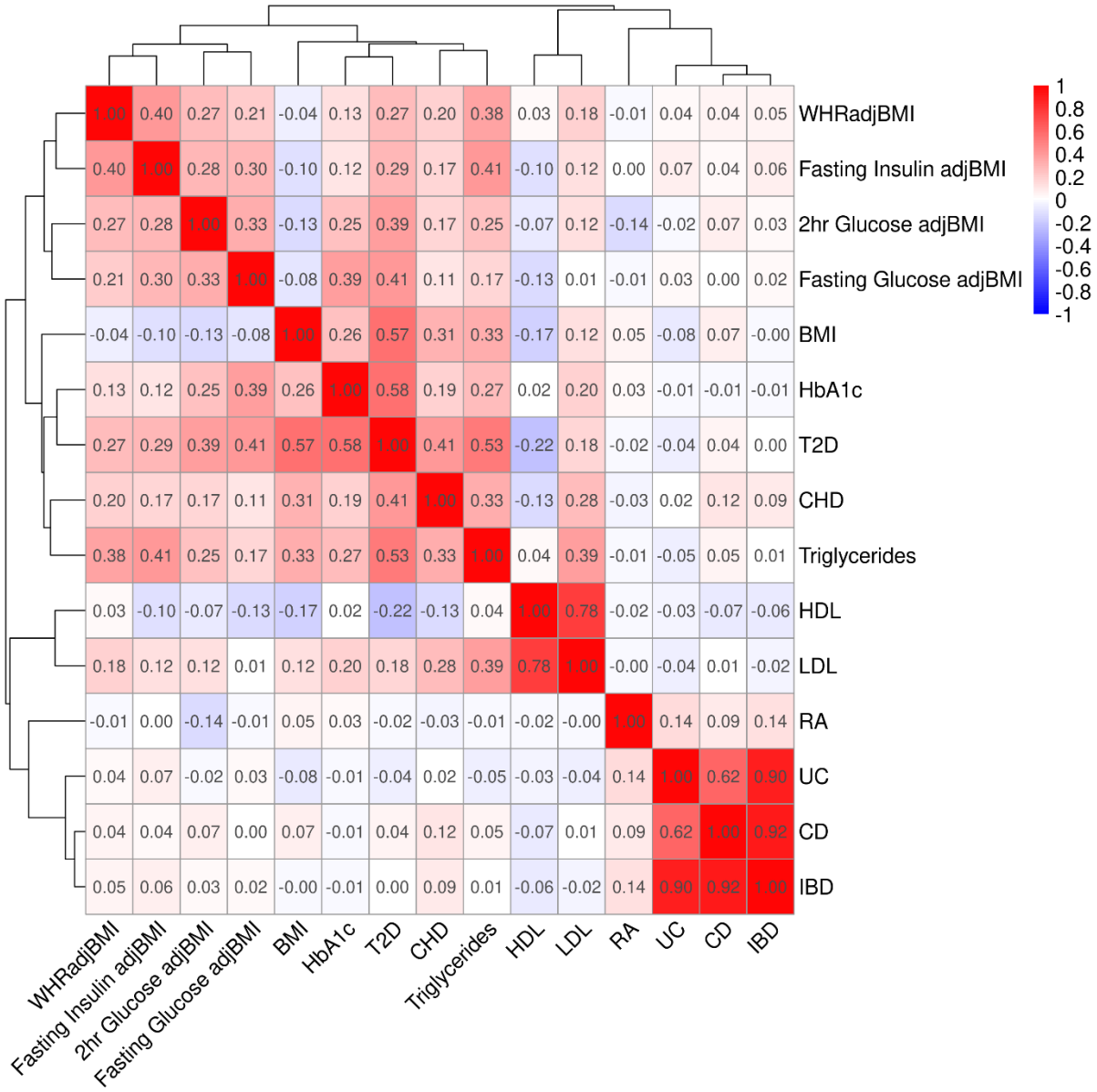
Abbreviations: CI, Confidence interval; mg, milligrams; kg, kilograms

Figure S4.11: Funnel plot for publication bias assessment. Illustration of the relationship between the hazard ratio for type 2 diabetes per log-pg/mL higher IL-6 levels and standard error for each study in included in the systematic review of prospective studies. Each study is represented by a dot and the estimated hazard ratio is depicted by the red line. The 95% confidence interval is shown by the grey contours.



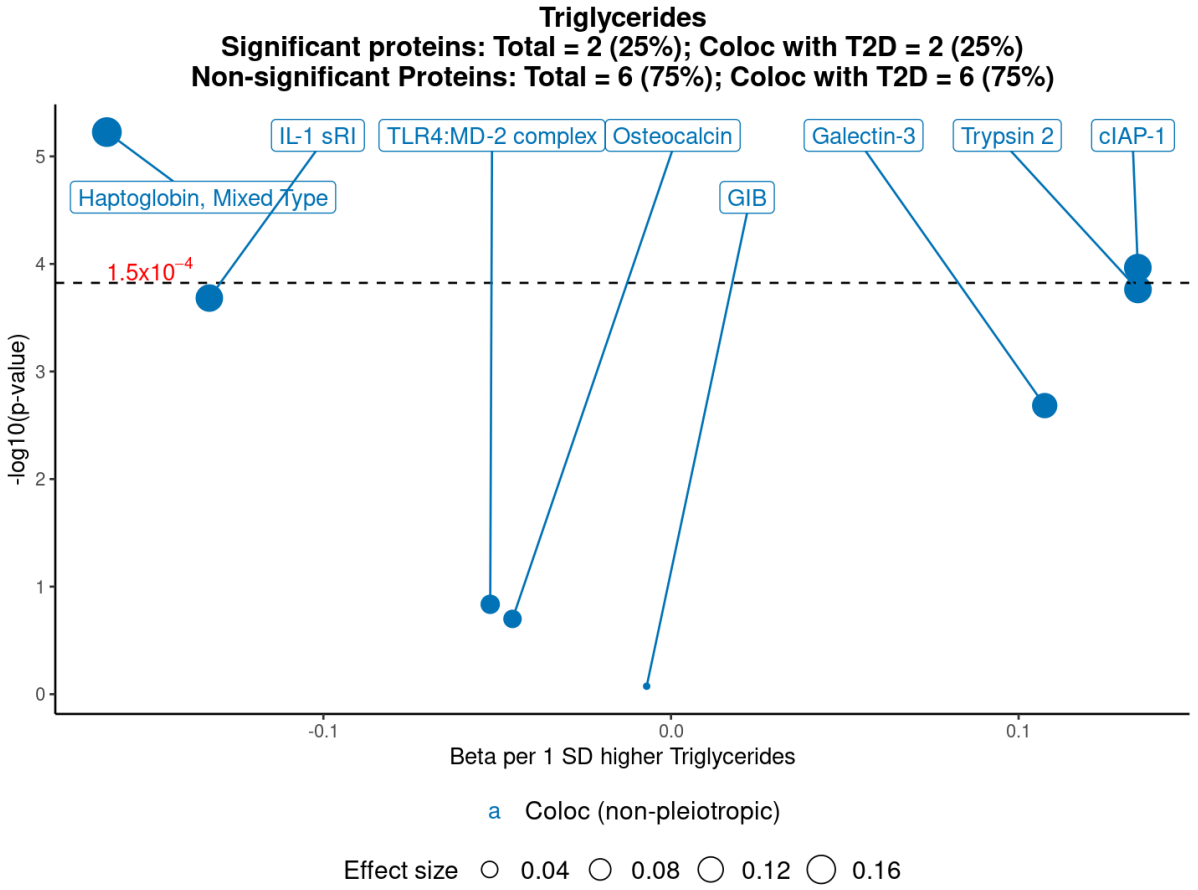
Abbreviations: CI, Confidence interval

Figure S5.1: Correlation heatmap illustrating the pairwise genetic correlation between autoimmune diseases and cardiometabolic traits. Positive correlations are shown in red whereas inverse correlations are shown in blue. Colour saturation denotes the strength of the correlation between traits.



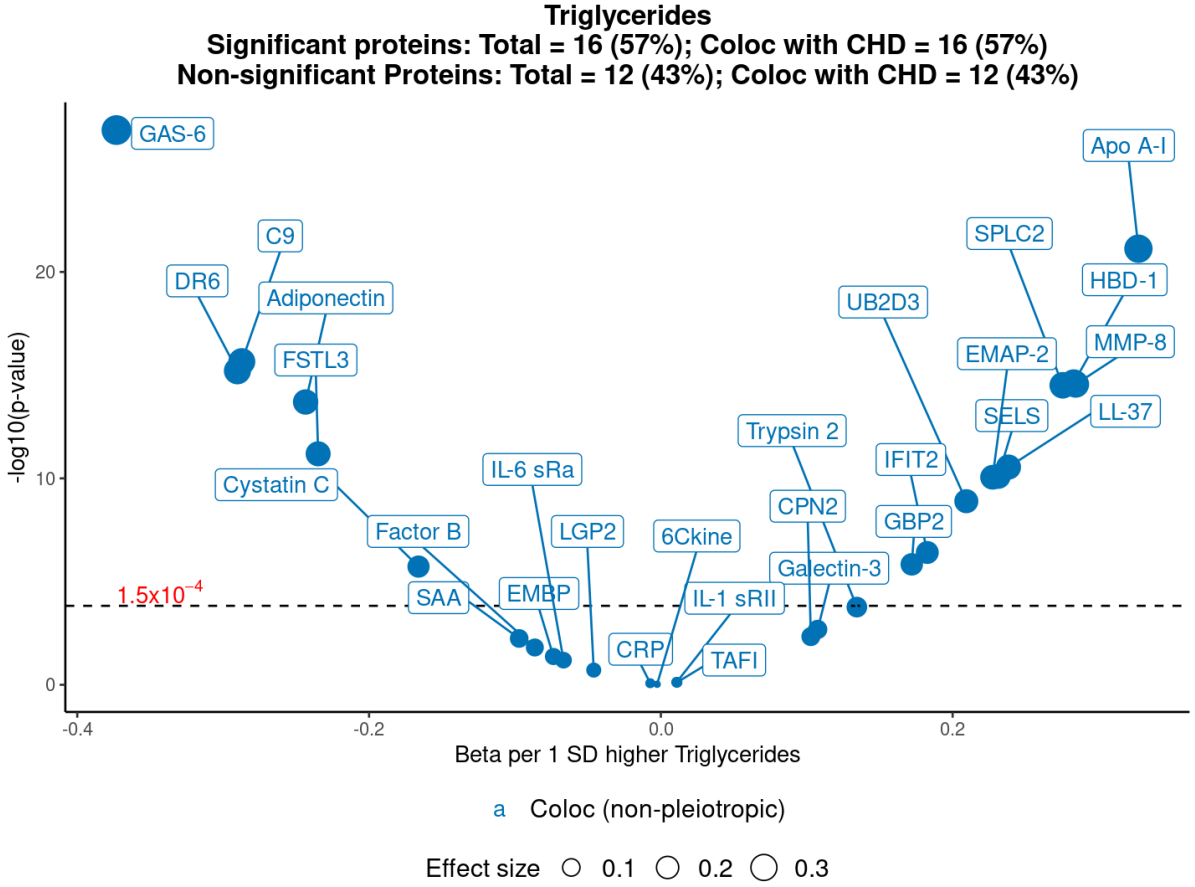
Abbreviations: BMI, Body mass index; HbA1c, Glycated haemoglobin; T2D, Type 2 diabetes; CHD, Coronary heart disease; HDL, High density lipoprotein; LDL, Low-density lipoprotein; RA, Rheumatoid arthritis; UC, Ulcerative colitis; IBD, Inflammatory bowel disease

Figure S5.2: Volcano plot of the association between a polygenic score for higher triglyceride levels with levels of the inflammatory proteins which colocalised with T2D at non-pleiotropic loci. The number of proteins significantly associated ($P<1.5x10^{-4}$) with the score is shown. Markers for each protein are sized according to their association with the triglycerides score and coloured according to their colocalisation with T2D.



Abbreviations: Coloc, Colocalisation; SD, Standard deviation; T2D, Type 2 diabetes

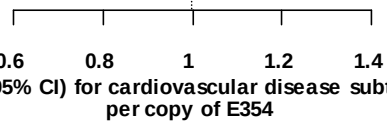
Figure S5.3: Volcano plot of the association between a polygenic score for higher triglyceride levels with levels of the inflammatory proteins which colocalised with CHD at non-pleiotropic loci. The number of proteins significantly associated ($P<1.5x10^{-4}$) with the score is shown. Markers for each protein are sized according to their association with the triglycerides score and coloured according to their colocalisation with CHD.



Abbreviations: Coloc, Colocalisation; SD, Standard deviation; CHD, Coronary heart disease

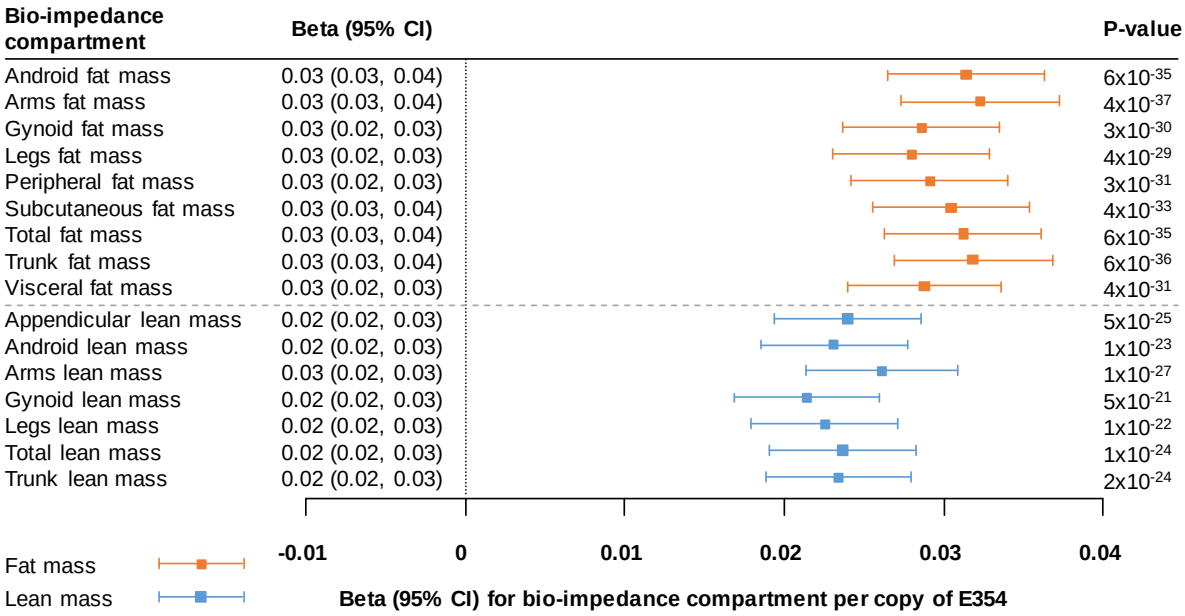
Figure S7.1: Association of E354 and cardiovascular disease sub-types in UK Biobank. Cardiovascular disease sub-types were defined in UK Biobank and tested for association with E354 using multivariable logistic regression adjusting for age, sex and 10 principal components[535]. Estimates for each disease are expressed per copy of E354. A Bonferroni corrected significance threshold of P<0.0029 was used.

Cardiovascular disease subtype	Cases	Controls	OR (95% CI)	P-value
Ischaemic, haemorrhagic and unknown stroke (but not TIA)	9,652	357,934	1.00 [0.97, 1.04]	0.87
Ischaemic Cerebrovascular Disease (all)	8,084	359,502	1.00 [0.96, 1.04]	0.91
Ischaemic Stroke	4,602	362,984	1.00 [0.95, 1.05]	0.97
Transient Ischaemic Attack	3,962	363,624	1.00 [0.94, 1.05]	0.91
Haemorrhagic Stroke (all)	1,981	365,605	0.97 [0.89, 1.05]	0.43
Intracerebral Haemorrhage	1,064	366,522	0.98 [0.88, 1.09]	0.75
Subarachnoid Haemorrhage	1,084	366,502	0.98 [0.88, 1.09]	0.73
Abdominal Aortic Aneurysm	1,094	366,492	1.19 [1.07, 1.30]	0.003
Thoracic Aortic Aneurysm	347	367,239	0.86 [0.68, 1.04]	0.1
Venous Thromboembolism (all)	14,097	353,489	1.03 [1.00, 1.06]	0.03
Deep Vein Thrombosis	9,454	358,132	1.05 [1.01, 1.08]	0.02
Pulmonary Embolism	6,148	361,438	1.04 [0.99, 1.08]	0.13
Peripheral Vascular Disease	3,415	364,171	1.05 [0.99, 1.11]	0.13
Aortic Valve Stenosis	2,244	365,342	0.93 [0.86, 1.00]	0.05
Atrial Fibrillation	16,945	350,641	1.00 [0.97, 1.03]	0.89
Heart Failure	6,712	360,874	1.01 [0.97, 1.05]	0.65
Adjusted for age, sex, 10 PCs				



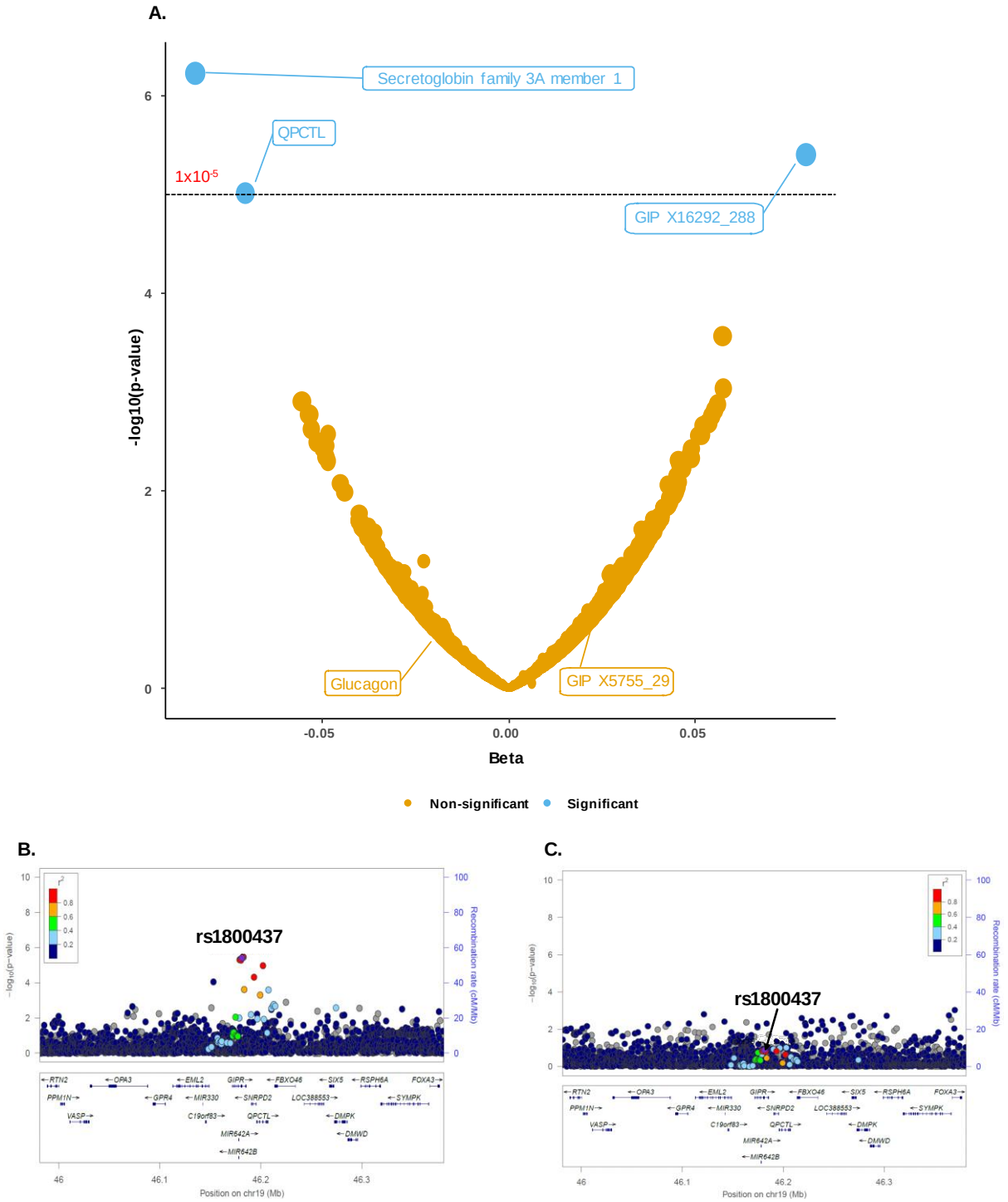
Abbreviations: TIA, Transient ischaemic attack; PC, principal component; OR, Odds ratio; CI, Confidence interval

Figure S7.2: Associations between E354 and regional adiposity compartments in 435,387 participants measured by bio-impedance. Fat mass in each compartment is shown in orange and lean mass in blue. Estimates for each compartment are in SD per copy of E354 (rs1800437).



Abbreviations: CI, Confidence interval

Figure S7.3: Associations between E354 and human protein levels. Panel A. Volcano plot showing the associations between E354 and 4,979 human protein levels. The dashed line indicates the Bonferroni significance threshold P ≤ $1x10^{-5}$. The point size for each protein is proportional to its effect size. Significant protein associations with E354 are shown in blue, non-significant proteins are shown in yellow. Associations with significant proteins and proteins of interest are labelled. Two SOMAmers from the SOMAscan® 4k assay target GIP levels, both are labelled. **Panels B & C.** Regional association plots depicting the E354 (rs1800437) association with both GIP SOMAmers, X16292_288 and X5755_29 respectively.



Abbreviations: QPCTL, Glutaminyl-peptide cyclotransferase like; GIP, Gastric inhibitory polypeptide.

Figure S7.4: Associations between E354 and human metabolite levels. Volcano plot showing the associations between E354 and the levels of 1,008 human plasma metabolites. The dashed line indicates the Bonferroni significance threshold $P \leq 5x10^{-5}$. The point size for each protein is proportional to its effect size. Metabolites are coloured according to their metabolite class. Significant metabolite associations with E354 are labelled in orange.

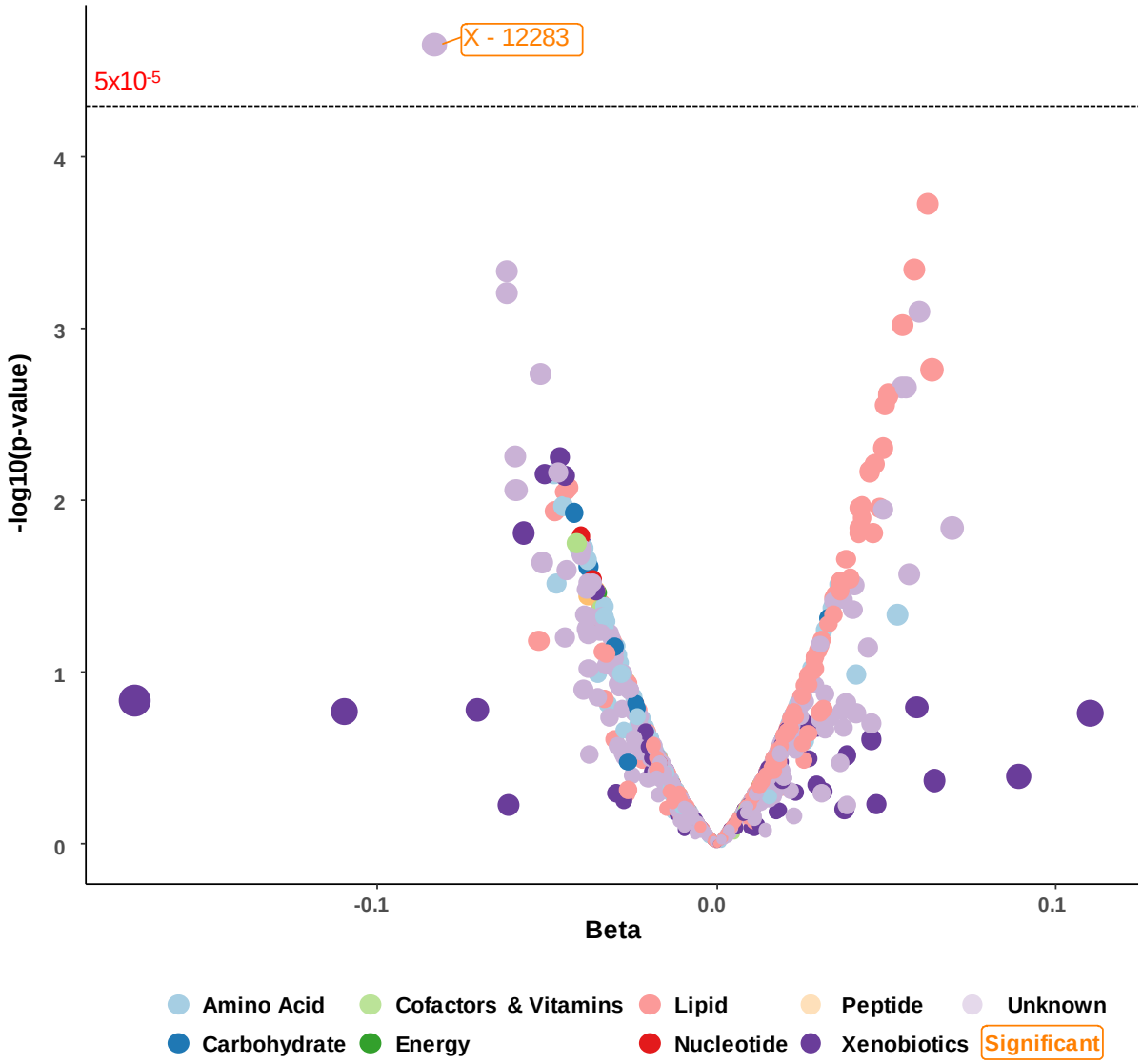
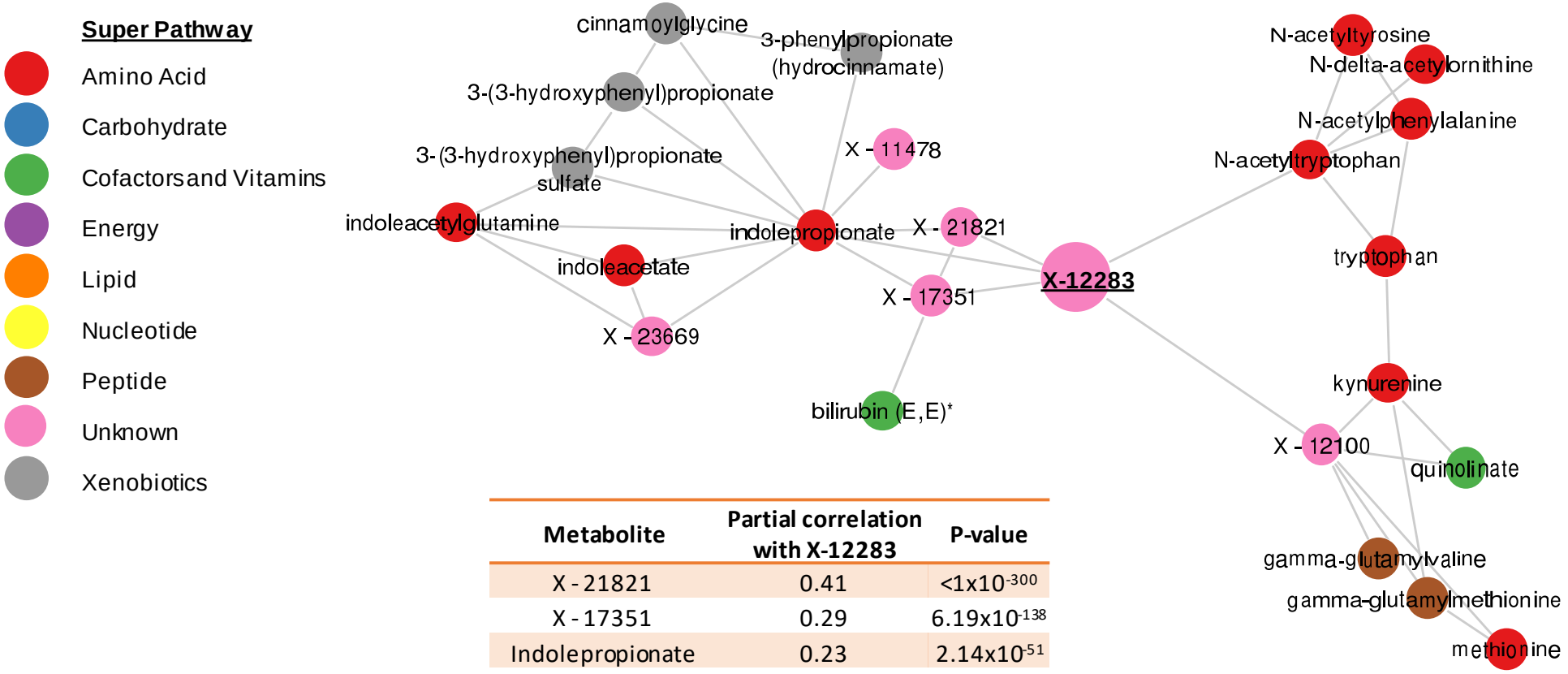


Figure S7.5: Gaussian graphical model illustrating the partial correlation network in 11,966 participants between X-12283 and first and second order connections most correlated with X-12283. Metabolites directly connected with X-12283 represent first order connections, others are second order connections. Metabolites clustered closest to X-12283 are more strongly correlated. Metabolite nodes are coloured by their super pathway. The table outlines the 5 metabolites with a first order connection to X-12283 and shows their partial correlation coefficients and related P-values.



Metabolite	Partial correlation with X-12283	P-value
X - 21821	0.41	$<1x10^{-300}$
X - 17351	0.29	$6.19x10^{-138}$
Indolepropionate	0.23	$2.14x10^{-51}$
N-acetyltryptophan	0.16	$2.49x10^{-33}$
X - 12100	0.15	$1.05x10^{-44}$

Figure S7.6: Stacked regional association plot showing the cluster of cardiovascular-related traits which colocalise near the *GIPR* locus. The purple diamond represents the rs7412 variant, a missense variant in *APOE*. variant markers are coloured by their LD with rs7412, with red indicating LD (R^2 > 0.8).

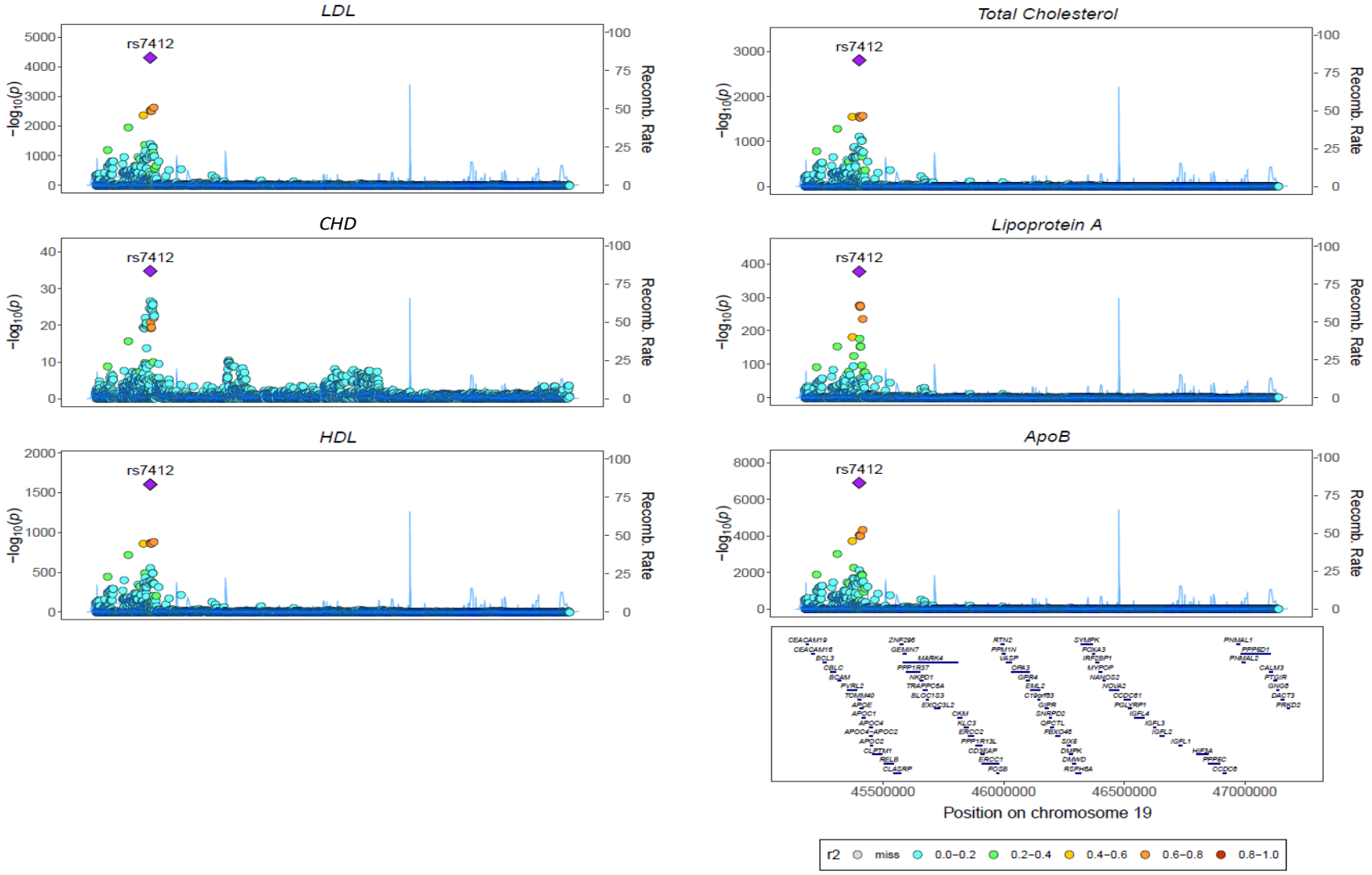
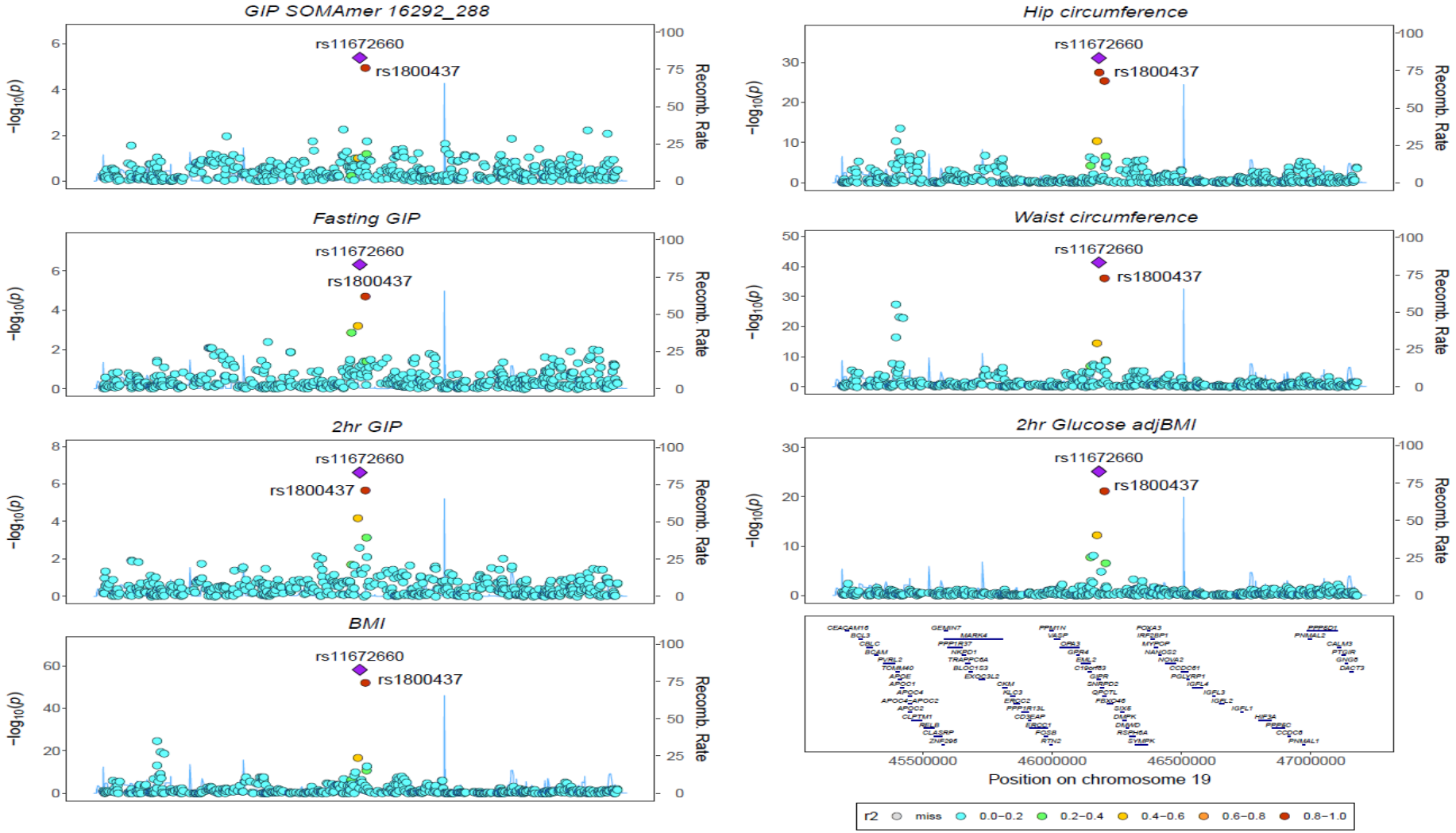
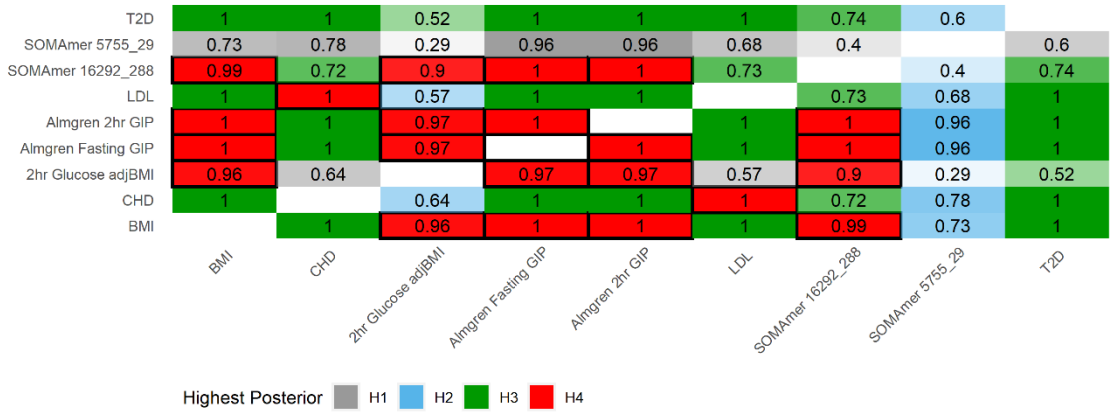


Figure S7.7: Regional association plot illustrating the cluster of traits which colocalise with the GIP measures at the *GIPR* locus. The purple diamond represents the rs11672660 variant which is in perfect LD with rs1800437 (E354). variant markers are coloured by their LD with rs11672660, with red indicating LD (R^2 > 0.8). Fasting and 2-hour GIP levels are from the MDC cohort of Almgren *et al.* 2017[508].



Abbreviations: GIPR, Gastric inhibitory polypeptide receptor; LD, Linkage disequilibrium; adj, Adjusted for; BMI, Body mass index; HbA1c, Glycated haemoglobin

Figure S7.8: Heatmap matrix depicting the pairwise colocalisation between cardiometabolic traits at the *GIPR* locus. The largest pairwise colocalisation estimate between fasting GIP measures from SomaLogic, fasting and 2-hour GIP measures from Almgren *et al*. 2017[508], 2hr glucose adjusted for BMI ,BMI, LDL, CHD and T2D is shown. Each colocalisation hypothesis is coloured differently with the colour saturation referring to the evidential strength. Trait-pairs with significant posterior probability estimates of colocalisation were outlined in black. To discriminate between H1 and H2 hypotheses, traits along the X-axis were used as "Trait 1" in the analysis and traits listed on the Y-axis were used as "Trait 2".



Abbreviations: H, Hypothesis; BMI, Body mass index; CHD, Coronary heart disease; GIP, Gastric inhibitory polypeptide; 2hr, 2-hour; LDL, Low-density lipoprotein; T2D, Type 2 diabetes

Figure S7.9: Matrix illustrating the LD between each of the independent CHD variants and rs1800437 (E354) estimated using 5 European populations in LDlink[590]. Pairwise R^2 values between variants are shown in red in the lower triangle, whereas D' values are shown in blue in the top triangle. Colour saturation represents the strength of the LD estimate between two variants. The LD between rs1800437 and rs1964272 (R^2 = 0.27) is depicted in light pink, whereas the very low LD between rs1800437 and the other CHD variants are shown as blank spaces.

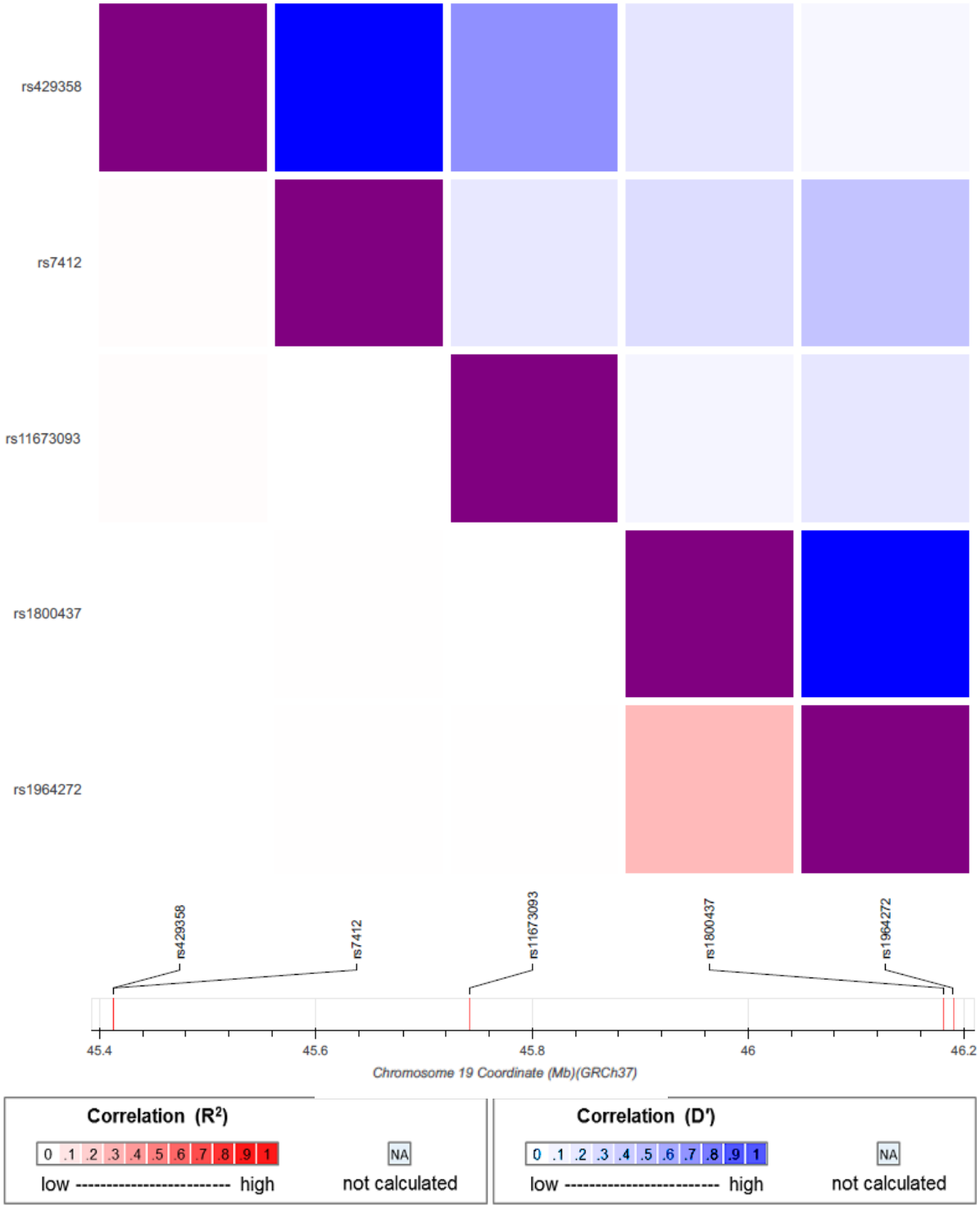
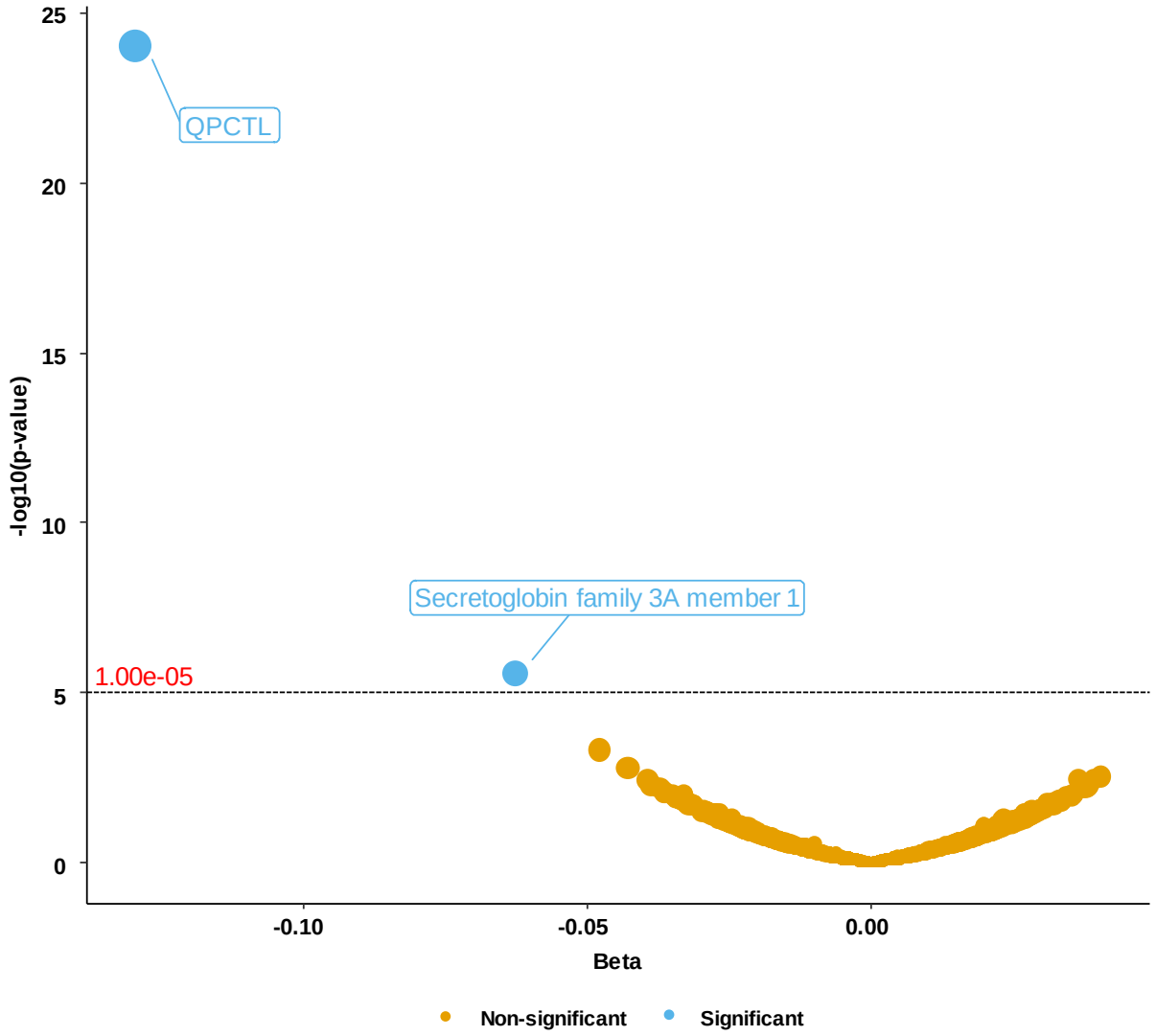


Figure S7.10: Volcano plot showing the associations between rs1964272 and 4,979 human protein levels. The dashed line indicates the Bonferroni significance threshold P ≤ $1x10^{-5}$. The point size for each protein is proportional to its effect size. Significant protein associations with rs1964272 are shown in blue, non-significant proteins are shown in yellow.



Abbreviations: QPCTL, Glutaminyl-peptide cyclotransferase like