The association between adiposity and platelet function: implications for cardiovascular disease

Lucy Goudswaard
Bristol Platelet Group / Team Timpson
Physiology, Pharmacology and Neuroscience / Population Health Sciences
Faculty of Life Sciences
University of Bristol

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of PhD in the Faculty of Life Sciences

Physiology, Pharmacology and Neuroscience / Population Health Sciences, March 2022, Word Count:

Abstract

Variation in adiposity is associated with thrombotic outcomes, but mechanisms leading from this exposure to disease are unclear. The overall aims were to assess the effect of adiposity on platelet function (key cells involved in thrombosis) and identify potential mechanisms.

Data from INTERVAL (a UK study of blood donors), COPTIC (Bristol-based cardiac surgery cohort) and DiRECT (Diabetes REmission Clinical Trial in Glasgow/Newcastle) were used. The effect of body mass index (BMI) on platelet traits measured by Sysmex XN-1000 was assessed in INTERVAL (N=33,388) using Mendelian randomization (MR). The association between platelet traits and platelet function was explored using linear regression in COPTIC (N=655). Bariatric patient (BMI >40kg/m2) platelet function was also compared with controls (BMI 18.5-25kg/m2). MR evidence in INTERVAL suggested that higher BMI raises immature platelet count (IPC - newly produced platelets). Results from COPTIC indicated that greater IPC is associated with greater platelet aggregation. Validation work in a specifically designed study of bariatric patients and healthy weight controls is ongoing.

To identify the involvement of the circulating proteome in adiposity-related platelet changes, the causal effect of BMI on >3600 proteins measured by SomaLogic was estimated using MR (N=2,737). To complement this, the effect of weight loss on the serum proteome was examined using linear regression in the DiRECT cohort (N=265). Adiposity had a substantial effect on the plasma proteome: higher BMI raised levels of leptin and fumarylacetoacetase while reducing sex hormone binding-globulin levels.

These analyses suggest that higher BMI raises immature platelet count (newly produced platelets), which are pro-thrombotic. The proteomic

analyses from different study designs point towards potential proteins which may influence platelet production and function which require further investigation. Future analyses would include developing approaches to compare protein profiles across cohorts. Overall, these analyses provide novel mechanistic insight into how higher BMI causes disease.

Acknowledgements

First of all, I would like to say a huge thanks to my supervisors Professor Ingeborg Hers and Professor Nicholas Timpson for their guidance and advice throughout my PhD. I have felt supported when I have needed it, but also feel like I have been given the independence and academic freedom, which I will take forward into my career. It goes without saying that this work would not have been possible without the British Heart Foundation Integrative Cardiovascular Science PhD programme, along with the funding I received from the University of Bristol alumni.

Thank you to all the members of Bristol Platelet Group and Team Timpson who have been so encouraging, supportive and willing to help me, especially in the early stages of my PhD. Although interrupted by the pandemic, I enjoyed all the pub trips and socials that I attended when they were allowed!

I would also like to thank my family and friends for their support throughout my PhD. To my flat mates of 3.5 years, Tessa and Josie, thanks for the support and motivation to keep going, especially in each lockdown when all we saw was each other. And lastly, thank you to Harry. Whether it be sending gifts for abstracts being accepted, printing out my first publication into a booklet or just offering words of encouragement, he has been a rock through it all!

Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

Signed

Dated

Table of Contents

List of Tables

List of Figures

Preface

Introduction

0.1 Format

This chapter will provide some background information to put the thesis into context. It will provide information on obesity, in particular its prevalence and implications for cardiovascular disease. Possible mechanisms of disease will be introduced, such as the involvement of changes in circulating proteins and blood platelets. At the end of this chapter, the aims of the thesis and organisation of the results chapters will be introduced.

0.2 Obesity on the rise

Obesity, defined as a body mass index (BMI) of greater than 30kg/m^2, has tripled worldwide since 1975. It is currently estimated that 40% of adults in the United States and 26% of adults in the United Kingdom have obesity.[1] As of 2016, the number of individuals globally with obesity surpassed 650 million.[2] The average BMI of the UK adult population is ~27kg/m^2,[3] which is in the 'overweight' category (between 25 and 30 kg/m^2). Overweight is also more common than 'normal-weight' in many other higher-income countries,[4] and prevalence of overweight is increasing in lower-income countries.[5]

BMI is generally used as a proxy for adiposity as there are high correlations between BMI and objective measures of fat mass measured by dual-energy X-ray absorptiometry (DXA),[6] such as total fat mass index and regional fat mass indices. Due to the specialised equipment needed for fat mass measures, BMI is often much more widely available within cohort studies and trials.

0.3 Obesity is a modifiable risk factor for cardiovascular disease

Adiposity is a risk factor for certain non-communicable diseases. Such diseases include Type 2 Diabetes mellitus (T2D), musculoskeletal diseases, certain types of cancer and cardiovascular disease.[7–10] Collectively, treatment for obesity-related diseases is costly for health services.[11,12] Cardiovascular disease is one of the leading causes of death. Cardiovascular disease is a general term for diseases affecting the heart or circulatory system and can be subcategorised further into diseases such as thrombotic disorders including coronary heart disease (CHD) and stroke. Obesity is associated with an increased risk of such cardiovascular events.[8]

0.4 Mendelian randomisation can be used to explore causality

Mendelian randomisation is a technique which uses the random assortment of alleles to estimate the causal association between a modifiable risk factor (such as BMI) and disease outcomes. It can be likened to a natural randomised control trial (RCT) [FIG]. Associations between BMI and disease have been shown through use of prospective observational studies,[8] but more recently, through use of Mendelian randomisation studies.[13–15] These studies have provided evidence for a causal re-

lationship between adiposity and ischaemic heart disease (IHD),[13] coronary heart disease (CHD), T2D[14] and types of cancer including pancreatic cancer.[15] Studies have also found that weight loss improves cardiovascular risk factors [[16]; other refs]. Despite the evidence for a causal association between adiposity and cardiovascular disease, mechanisms of obesity-related disease remain unclear.

0.5 Mechanisms of obesity-related cardiovascular disease

Obesity is caused by an increase in fat deposition and generally occurs when energy intake exceeds energy expenditure. There are two types of fat: white adipose tissue (WAT) and brown adipose tissue (BAT). WAT secretes proteins and peptides which have pro-inflammatory effects.[17,18] Due to these adipose tissue effects, obesity results in a state of chronic low-grade inflammation,[19] with an increase in adipose tumour necrosis factor-alpha (TNF-alpha) and interleukin-6 (IL-6).[18] Efforts have been made to understand how adiposity causes cardiovascular disease, with suggestion that inflammation, increased production of reactive oxygen species and endothelial dysfunction contribute to increased cardiovascular risk.[17, add other refs]

0.6 Plasma markers of obesity

Studies which have explored the molecular footprint of adiposity have tended to focus on the lipidome, such as cholesterol and triglycerides in lipoprotein subtypes, inlcuding low-density lipoproteins (LDL), very-low-density lipoproteins (VLDL) and high-density lipoproteins (HDL),[20,21] where a higher BMI is associated with an increase in LDLs and a decrease in large HDLs. This has been expored in a

causal framework.

As the majority of druggable targets are proteins, it is also valuable to explore the effects of BMI on circulating plasma proteins. It is estimated that 25% of proteins in the human proteome circulate in the blood,[22] including enzymes, protein kinases and transport proteins. Therefore, focusing on effects of adiposity on the circulating proteome may allow for the development of or repurposing of therapeutic interventions to reduce obesity-related disease. Previous studies that have explored the effect of adiposity on circulating proteins have used hypothesis driven approaches.[23] Attempts to evaluate the effect of BMI on the general proteome have been performed using observational analyses.[24] Recent technologies have been developed which allow protein detection in unprecedented scope and detail, for example the SOMAscan by SomaLogic,[25] which can detect over 3,500 unique proteins. Although identifying alterations of individual proteins may provide therapeutic opportunities, assessing the overall protein profile of adiposity may provide more of a mechanistic insight into how adiposity causes disease.

0.7 Platelets are key cells involved in thrombosis

Platelets are enucleate blood cells that are involved in the cessation of bleeding, but can cause thrombosis when activated pathologically. Antiplatelet drugs used to prevent reocurrence of MI. BMI is linked to increased rates of thrombosis, therefore it is likely that platelets become hyperactive. There is conflicting literature surrounding this and cohorts to explore these effects are limited.

0.8 Importance of interdisciplinary methods

Each approach has biases and problems, which means we need to be careful .... A combination of laboratory experiments, clinical studies and epidemiology (including Mendelian randomization and RCTs) in order to investigate mechanisms of disease.

0.9 Thesis aims and objectives

The aims of the thesis are indicated below, with a schematic of the multi-step pathway which this thesis aims to explore (Figure 1). The 3 overall aims are:

1) Explore the effect of body mass index on platelet traits, platelet function and signalling.
2) Characterise the effects of the chemokines macrophage-derived chemokine (MDC) and thymus and activation regulated chemokine (TARC) on platelet function.
3) Determine the effects of adiposity on the plasma proteome.

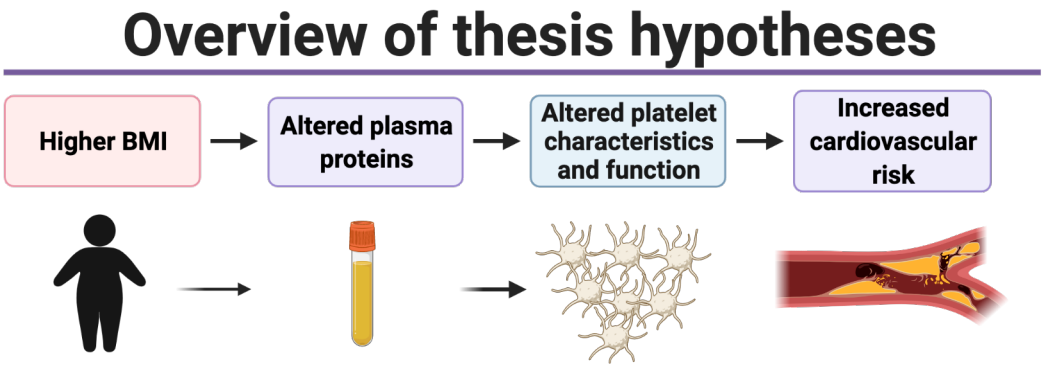

Figure 1: Schematic of the multi-step pathway addressed in the thesis

0.10 Thesis structure

There will be 5 results chapters in this thesis:

1) Higher body mass index raises immature platelet count: evidence from Mendelian randomization analyses (Chapter 1)
2) The effect of obesity on platelet function: a clinical pilot study (Chapter 2)
3) Pathways linking BMI and platelet function: do the chemokines MDC and TARC play a role? (Chapter 3)
4) Effects of adiposity on the human plasma proteome: Observational and Mendelian randomization estimates (Chapter 4)
5) Using a randomised control trial to explore the effects of caloric restriction-induced weight loss on the plasma proteome (Chapter 5)

0.11 Publications from the PhD (in preparation or published)

Goudswaard L. J., Bell J. A., Hughes D. A., Corbin L. J., Walter K., Davey Smith G., Soranzo N., Danesh J., Di Angelantonio E., Ouwehand W. H., Watkins N. A., Roberts D. J., Butterworth A. S., Hers I., Timpson N. J. Effects of adiposity on the human plasma proteome: Observational and Mendelian randomisation estimates (2021) Int J Obes `https://doi.org/10.1038/s41366-021-00896-1`

Goudswaard L. J., Harrison S., Van De Klee D., Chaturvedi N., Lawlor D. A., Davey Smith G., Hughes A.D., Howe L. D. Blood pressure variability and nighttime dipping assessed by 24-hour ambulatory monitoring: Cross-sectional association with cardiac structure in adolescents. (2021) PLOS ONE 16(6): e0253196.

https://doi.org/10.1371/journal.pone.0253196

Goudswaard L. J., Corbin L. J., Burley K. L., Mumford A., Akbari P., Soranzo N., Butterworth A. S., Watkins N. A., Pournaras D. J., Harris J., Timpson N. J., Hers I. Higher body mass index raises immature platelet count: evidence from Mendelian randomization analyses. (2021) medRxiv. doi:10.1101/2021.05.19.2125744325.

Liu, H., Jackson, M. L., **Goudswaard, L. J.**, Moore S. F., Hutchinson J. L, Hers I. Sphingosine-1-phosphate modulates PAR1-mediated human platelet activation in a concentration-dependent biphasic manner. (2021) Sci Rep 11, 15308. https://doi.org/10.1038/s41598-021-94052-4

Sledz, K. M., Moore, S. F., Vijayaragavan, V., Mallah, S., **Goudswaard, L. J.**, Williams, C. M., Hunter, R. W. & Hers, I. Redundant role of ASK1-mediated p38MAPK activation in human platelet function. (2020) Cellular Signalling. 68, 11 p., 109528. doi:10.1016/j.cellsig.2020.109528

0.12 Presentations from the PhD

Platelet Society meeting, 2018, Cambridge, UK. Title: Do the chemokines MDC and TARC contribute to obesity related platelet hyperactivity and cardiovascular disease? (Oral presentation).

IEU Scientific Advisory Board meeting, 2020, Bristol, UK. Title: Effect of adiposity on the human plasma proteome: a Mendelian randomisation study using individual level data (Poster).

International Genetic Epidemiological Society meeting, 2020, Virtual conference. Title: Effects of body mass index on the human proteome: a Mendelian

randomization study using individual level data (Selected for short talk as one of the top ranked posters).

International Society of Thrombosis and Haemostasis meeting, 2020, Virtual conference. Title: Effects of body mass index on the human plasma proteome – potential drivers of obesity-related cardiovascular disease? (Poster).

Platelet society meeting, 2021, Virtual conference. Title: Altered platelet proteome, enhanced basal platelet procoagulant activity, and increased platelet-leukocyte interactions in patients with Covid-19. (Oral presentation)

Platelet society meeting, 2021, Virtual conference. Title: Estimating the causal association between body mass index and platelet properties: evidence for a positive association with immature platelet count (Poster)

British Heart Foundation PhD Student Conference, 2021, Virtual conference. Title: Estimating the association of body mass index and platelet properties: an observational and Mendelian randomization study (Poster)

International Society of Thrombosis and Haemostasis meeting, 2021, Virtual conference. Title: Altered platelet proteome and platelet-leukocyte interactions in patients with COVID-19 (Oral presentation and Travel Award, presented by Prof Hers)

International Society of Thrombosis and Haemostasis meeting, 2021, Virtual conference. Title: Estimating the causal association between body mass index and platelet properties: evidence for a positive association with immature platelet count (Oral presentation and Travel Award)

World Congress of Epidemiology, 2021, Virtual conference. Title: Effects of adiposity on the human proteome: Mendelian randomization study using individual-

level data (Oral presentation)

International Genetic Epidemiology Society, 2021, Virtual conference. Title: Combining Mendelian randomization and randomized control trial study designs to determine effects of adiposity on the plasma proteome (Oral presentation)

Background

0.13 This section

0.14 That section

0.15 Another section

Chapter 1

Higher body mass index raises immature platelet count: evidence from Mendelian randomization analyses

This work is under review at Platelets journal and is on the preprint server MedRXiV.[26] I performed all analyses, created all figures and tables and wrote the manuscript.

1.1 Background

Obesity (body mass index $\geq$ 30 kg/m^2) has nearly tripled worldwide in the last 45 years.[1] This is a major health concern as a higher body mass index (BMI) is an important risk factor for various noncommunicable disorders, including cardio-

vascular disease.[27] Clinical studies and genetic association studies have identified BMI as an independent risk factor for thrombotic disorders, including coronary heart disease and stroke 4-6.[13,14,28]

Platelets are blood cells which are essential for the cessation of bleeding upon injury to a blood vessel and are involved in thrombosis and progression of cardiovascular disease.[29] When pathologically activated, platelets can aggregate to form thrombi thereby occluding arteries and triggering a myocardial infarction or stroke.[29] Platelet hyperactivity can be an indicator of those who may be at an increased risk of thrombosis.[30] There is evidence that people with a higher BMI have hyperactive platelets,[31,32] which could explain why a higher BMI is linked to an increase in thrombosis. Mechanisms of platelet hyperactivity are unclear, but it is possible that changes in platelet numbers, size and immaturity may play a role. Furthermore, an alteration in circulating metabolites or proteins induced by obesity could modulate the activation of platelets.[33]

Platelet function can be directly assessed by various methods; these methods commonly include platelet aggregation experiments and/or using antibodies to detect platelet receptors that are expressed upon activation. One of the limitations of measuring platelet function directly is that these techniques are not widely available or readily standardized. Haematology analyzers are more commonly used to provide full blood counts, providing detailed readouts of platelet characteristics. Some of these characteristics have been reported to be indirect measures of platelet function.[34] For example, an increase in mean platelet volume (MPV) has been reported to be predictive of vascular mortality and ischaemic heart disease.[35] Another commonly measured platelet characteristic is the number of platelets in circulation (platelet count, PLT). An increase in platelet count has been reported to be associated with platelet hyperactivity[34] and there is evidence that higher platelet count is

associated with ischaemic stroke.[36]

Sysmex haematology analyzers are able to provide additional platelet characteristics, such as immature platelet fraction (IPF) and immature platelet count (IPC).[37] Immature platelets (also known as reticulated platelets) are the youngest platelets in circulation and are detected based on having a greater forward scatter (FSC, an indicator of their size) and a greater side fluorescence (SFL, an indicator of their mRNA content).[38] Higher levels of immature platelets are indicative of enhanced platelet production[39] and these platelets are believed to be more prothrombotic than older circulating platelets, with more dense granule release and increased P-selectin expression.[38–40] Higher immature platelet count is associated with adverse cardiovascular outcomes in patients with coronary artery disease[38] and reduced effectiveness of antiplatelet therapies,[40,41] suggesting that hyperactive immature platelets may contribute to adverse vascular events. Immature platelets have also been suggested to be less responsive to antiplatelet therapies such as prasugrel in acute coronary syndrome patients.[40]

As these platelet characteristics provide potential information about platelet hyperactivity and given the increased thrombotic risk seen with adiposity, it is important to assess how adiposity affects platelet properties. Previous studies with modest samples sizes have implemented observational epidemiological methods to explore the effect of BMI on PLT, plateletcrit (PCT) and MPV. There is conflicting observational evidence regarding an association between BMI and PLT,[42,43] with some studies suggesting a positive association between BMI and MPV[44] and other studies reporting no such association.[45] Furthermore, as there is evidence that immature platelet production is increased in patients with metabolic syndrome and type II diabetes,[46,47] it is therefore important to explore whether BMI may be an independent predictor of this trait. It is currently unknown whether the influence of

BMI on platelet properties and function are causal and independent of confounding effects.

In this study, we used data from the INTERVAL prospective cohort (N=33388) to explore the association between BMI and platelet traits. We combined observational and Mendelian randomization (MR) approaches to test the hypothesis that higher BMI leads to changes in platelet characteristics. Although observational studies can demonstrate associations between BMI and platelet characteristics, they cannot determine direct causality. To address the latter, we therefore employed MR, using a genetic risk score derived from single nucleotide polymorphisms (SNPs) associated with higher BMI (Figure 1.1). This allowed estimation of the causal effect of BMI on platelet traits, reducing the effect of confounding factors that are inherent to observational studies. To assess functional implications of BMI-platelet associations, a follow-up analysis was designed to explore the associations between platelet characteristics and whole blood aggregation in a cohort of cardiac surgery patients.

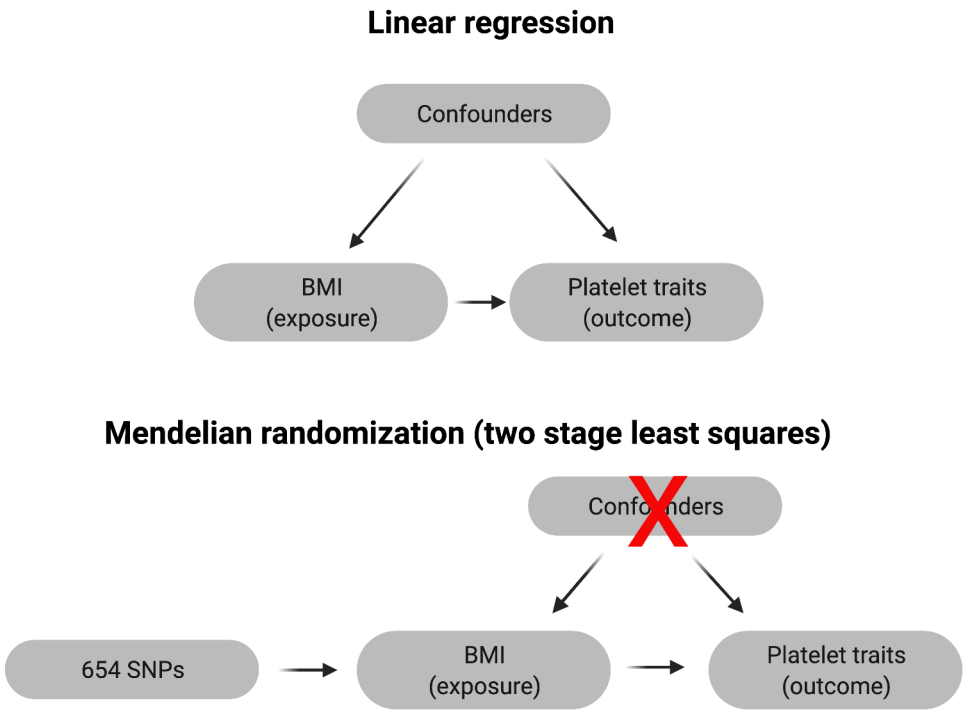

Figure 1.1: Schematic of linear regression and Mendelian randomization analyses. Linear regression assesses the association between BMI (exposure) and platelet traits (outcome), with adjustment for potential confounders. Under the assumptions of Mendelian randomization (MR), the genetic risk score (GRS) for body mass index (BMI) does not associate with confounding variables, and if there is a causal association, the GRS should only associate with platelet outcomes through its association with body mass index.

1.2 Methods

1.2.1 Study population

INTERVAL is a prospective cohort study that initially aimed to test the safety of reducing the time interval between donation of whole blood in 50 000 participants [[48]. Participants were over the age of 18 years, able to provide informed consent

and free from a history of major disease. Participants were recruited between June 11th, 2012 and June 15th, 2014 from 25 National Health Service Blood and Transplant (NHSBT) centres across England. They filled out online questionnaires including self-reported height and weight, smoking status and alcohol consumption. Blood samples were also taken at baseline (before randomization within the study) where full blood counts were obtained. The study was approved by Cambridge East Research Ethics Committee. Permission for data access was provided by the Data Access Committee. Data contains sensitive content and requires permission to use therefore cannot be made publicly available. Access to data needs to be approved by the INTERVAL team www.intervalstudy.org.uk/more-information.

The present study was conducted on up to 33 388 European ancestry participants living in the United Kingdom. These were the participants who had genetic data, had basic phenotype data and full blood count measures (Figure 1.2). Participants were mostly of European descent.

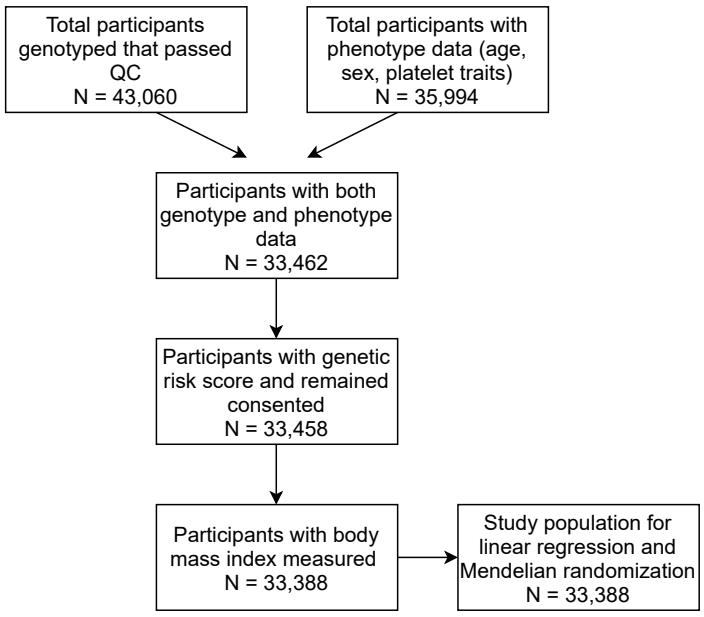

Figure 1.2: A STROBE diagram outlining participants included in the INTERVAL study (N=33388).

1.2.2 Measurement of BMI and covariables

Participants self-reported their weight and height using online questionnaires. BMI was derived from their weight in kilograms by the square of their height in metres (kg/m^2). BMI was rank normal transformed (rntransform() function from "moosefun" package `https://github.com/hughesevoanth/moosefun/`). Available covariables used in the analysis were age, sex, smoking status (in three categories of never, previous and current) and alcohol consumption (in four categories of rarely, less than once a week, 1-2 times a week and 3-5 times a week or most days). These covariables were chosen due to their plausible associations

with both BMI and cardiovascular health,[27] therefore it is important to adjust for these variables in the observational estimates.

1.2.3 Measurement of platelet traits

Blood was taken by venipuncture into 3mL EDTA tubes. Platelet parameters were measured using the Sysmex XN-1000 instrument.[49] This analyzer provides information on cell counts by using a combination of fluorescence (PLT-F) and impedance (I) flow cytometry. The PLT-F channel used a Fluorocell fluorescent dye (oxazine), whereas the impedance method uses electrical resistance to detect platelets. Platelet indices included in the current study, along with the raw units, are provided in Table 1.1. These measurements were pre-adjusted for technical covariates such as time between venipuncture and blood count, as well as instrument drift, seasonal and weekly variation.[49] Adjustment was performed with cyclic, thin-plate and P-splines in a generalized additive model (GAM).[50] Resulting platelet trait values along with their units of measurement are reported in Table 1.1. These data were rank normal transformed to normalize the distribution of each trait. Therefore, each platelet index is measured in normalized standard deviation (SD) units.

Table 1.1: Platelet traits measured by Sysmex XN-1000

Short name	Full name	Derivation of platelet trait	Units	Mean (SD)
P-LCR	Platelet large cell ratio	(platelet number with volume>12 fL)/(platelet number of ordinary volume) x 100	%	36.9 (7.7)
H-IPF	High fluorescence immature platelet fraction	(high fluorescence immature platelet number)/(platelet count) x 100	%	1.3 (1.0)
PLT (F)	Platelet count (PLT-F channel)	Total number of platelets per litre	109/L	255 (57)
PLT (I)	Platelet count (impedance channel)	Total number of platelets per litre	109/L	250 (55)
PCT	Plateletcrit	Volume percentage of blood occupied by platelets	%	0.28 (0.06)
PDW	Platelet distribution width	Standard deviation of platelet volume distribution, estimated as width at 20% modal height of platelet volume histogram	fL	14.0 (1.8)
IPC	Immature platelet count	Immature platelet count	109/L	10.1 (5.0)
IPF	Immature platelet fraction	(Immature platelet number)/(platelet count) x 100	%	4.2 (2.5)
MPV	Mean platelet volume	Mean platelet volume	fL	11.1 (1.0)
SFL	Side fluorescence	Side fluorescence measures the nucleic acid (mRNA) content of the platelet	RFUs	80.3 (5.0)
FSC	Forward scatter	Forward scatter is a measure of the size of the platelet	RFUs	53.0 (6.4)
SSC	Side scatter	Side scatter is a measure of the granularity of the platelet	RFUs	37.9 (2.9)

1.2.4 Genetic data and the association of the genetic risk score (GRS) with BMI

The genotyping of INTERVAL participants was performed using the Affymetrix GeneTitan® Multi-Channel (MC) Instrument. Quality control (QC) was performed as previously described.[49] Imputation was implemented using a combined 1000 Genomes phase-3-UK + 10K reference panel using the Practical Byzantine Fault Tolerance (PBFT) algorithm.[49] A genetic instrument for BMI was created by using 654 genetic variants available of the 656 variants which were independently associated with BMI (P<5x10-8) in a recent meta-analysis of GWAS of around 700,000 individuals of European descent.[51] Of these individuals, 450,000 were from UK Biobank and 250,000 were from the Genetic Investigation of Anthropometric Traits (GIANT) consortium. The weighted GRS was made using PLINK 2.0.[52] The effect alleles and beta coefficients were extracted from the source GWAS. The weighted score was calculated by multiplying the dosage of each effect alleles by its effect estimate, summing these and dividing by the number of SNPs (654). The GRS therefore reflects the average per-SNP effect on BMI per individual.

1.2.5 Statistical analysis

Analyses were performed using R version 3.4.2.[53] To visualize correlation between outcome platelet variables, a correlation matrix was made using the pairwise Pearson correlation coefficients of the rank normal transformed data. A dendrogram was also created to show the hierarchical relationship between platelet traits (`https://github.com/hughesevoanth/iPVs`), where the height at which variables are joined is set at (1 – Pearson correlation coefficient (r)). To explore observational associations between BMI and platelet traits, linear regres-

sion models were used (lm() function from "stats" package). Two regression models were used: firstly, adjusting for age and sex and secondly, additionally adjusting for smoking and alcohol consumption as ordinal variables. The results of the regression reflect the change in platelet traits (in normalized SD units) per normalized SD higher BMI (4.7 kg/m^2). Only participants with all covariables were included in the linear regression models. The associations between potential covariables with both BMI and platelet traits were also explored using linear regression.

To understand properties of the GRS for BMI, the association between the GRS with both BMI and covariables were explored. The MR analysis was performed by using a two-stage least squares (2SLS) regression model (using systemfit() function from "systemfit" package).[54] The MR causal estimates reflect the change in platelet traits (in SD units) per SD increase in BMI. A Wu-Hausman test was performed to test for endogeneity between observational and MR estimates. Exact beta coefficients, confidence intervals and P values are provided throughout and guide strength of association.

1.2.6 Follow-up analysis to explore the association between immature platelets and aggregation

Motivated by findings from the primary analyses described above, additional observational analyses were conducted using data from the COagulation and Platelet laboratory Testing in Cardiac Surgery (COPTIC) study. The COPTIC study was an observational, single centre cohort study of adults undergoing cardiac surgery at the Bristol Heart Institute with primary objective of examining the relationship between coagulation laboratory parameters and bleeding outcomes after surgery in 2,541 participants.[55] This study was approved by the UK NHS Research Ethics Committee (09/H0104/53).

1.2.7 COPTIC study variables

Age and sex were reported at baseline. Height and weight were obtained from medical notes. BMI was derived from weight and height (kg/m^2). Smoking was reported as a categorical variable (0=never smoker, 1=ex-smoker for > 5 years, 2=ex-smoker for 1-5 years, 3=ex-smoker for 30 days-1 year, 4=current smoker). Platelet variables (PLT, MPV, IPF, IPC) were measured using the Sysmex XE-2100 Automated Haematology System (Oxford, UK). Blood samples were taken pre-operatively into 3.2% sodium citrate vacutainers (BD Biosciences, Milton Keynes, UK). Platelet aggregation was measured using Multiplate multiple electrode aggregometry (MEA) (Roche, Rotkreuz, Switzerland), which detects change in electrical impedance when platelets aggregate on metal electrodes. Aggregation was determined by the area under the curve (AUC) in response to platelet agonists including adrenaline (100 mg/mL), thrombin receptor activator peptide 6 test (TRAP-test), ADP-test and arachidonic acid (ASPI-test). Use of a combination of agonists therefore reflects activation of platelets via different pathways: adrenaline acts at the $\alpha 2$ adrenergic receptor, TRAP-6 acts at the protease-activated receptor-1 (PAR-1), ADP acts at the P2Y12 receptor and arachidonic acid at the Thromboxane A2 (TxA2) receptor.

1.2.8 COPTIC statistical analysis

The COPTIC dataset in the current analysis included 2,518 participants (23 out of 2,541 participants did not consent for future research). For the current study, only those who were not on antiplatelet therapies (prasugrel, clopidogrel or aspirin) were included (N=655). Extreme outliers that were ± 5 SDs from the mean were removed, and exposures and outcomes were rank normal transformed. Linear regression was used to explore the association between immature platelet count

Table 1.2: Characteristics of included INTERVAL participants

Variable	Mean (SD) or %	N
Age	45.3 (14.2)	33388
Sex		33388
Male	49.8%	
Female	50.2%	
Weight (kg)	78.2 (16.0)	33455
Height (cm)	171.8 (9.6)	33388
Body mass index (kg/m2)	26.4 (4.7)	33388
Smoking frequency		32867
Never	58.9%	
Previous	33.3%	
Current	7.8%	
Alcohol intake frequency		29538
Rarely	12.5%	
Less than once a week	17.1%	
One or two times a week	37.7%	
Three to five times a week / most days	32.7%	

and aggregation, adjusting for age, sex and smoking status.

1.3 Results

1.3.1 INTERVAL participant characteristics

Of INTERVAL participants included in the current study (N=33,388), 50.2 % were female. The mean age was 45.3 years (SD of 14.2 years, Table 1.2). The mean BMI was 26.4 kg/m2 (SD of 4.7 kg/m2). The majority of participants were never smokers (58.9 %), with 33.3% and 7.8 % reported as previous and current smokers, respectively. Nearly a third (32.7 %) of participants reported drinking alcohol at least three times a week.

1.3.2 Correlation between platelet traits in INTERVAL

The Sysmex XN-1000 haematology analyzer measures multiple platelet traits, however many of these traits are closely related measurements and therefore may not be completely independent. Indeed, platelet traits showed a high degree of correlation with each other (Figure 1.3), in particular among similar measures. For example, measures of PLT (PLT I/F) and PCT, the latter a measurement of platelet mass, were highly positively correlated with each other but were weakly inversely correlated with other platelet measures. Measures of platelet maturity (IPF, IPC, H-IPF) were highly correlated with each other. In addition, platelet size variables (MPV, P-LCR and PDW) showed strong positive correlations with each other as well as with measures of immature platelets.

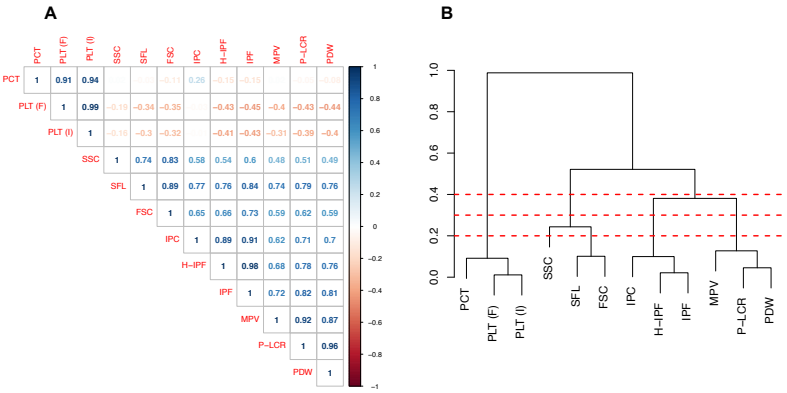

Figure 1.3: Correlation matrix and dendrogram of the relationship between platelet traits. A) Correlation coefficients are provided within the matrix. Dark blue indicates a correlation coefficient (r) of 1, with dark red indicating a correlation coefficient of -1. B) A dendrogram showing the hierarchical relationship between platelet traits, where the height at which two platelet traits join is (1 – Pearson correlation coefficient (r)). PCT = plateletcrit, PLT (F) = platelet count (PLT-F channel), PLT (I) = platelet count (impedance channel), SSC = side scatter, SFL = side fluorescence, FSC = side scatter, IPC = immature platelet count, H-IPF = high fluorescence immature platelet fraction, IPF = immature platelet fraction, MPV = mean platelet volume, P-LCR = platelet large cell ratio, PDW = platelet distribution width. Dashed red lines indicate a height of 0.2, 0.3 and 0.4.

1.3.3 Observational associations between BMI and platelet traits in INTERVAL

Linear regression was performed to determine whether a higher BMI was associated with a change in platelet characteristics. In the confounder adjusted linear regression (Table 1.3), BMI was positively associated with PCT (0.12 SD higher per SD higher BMI, 95% CI 0.11 to 0.13, P = 9.2 x 10^{-88}) and platelet count (PLT (I) 0.11 SD higher per SD higher BMI, 95% CI 0.09 to 0.12, P = 1.0 x 10^{-67}). The next strongest association with BMI was the association with SFL (0.06 SD higher per SD increase in BMI, 95% CI 0.05 to 0.07, P = 4.7x10^{-23}). BMI also showed

a positive association with IPC (0.06 SD higher per SD increase in BMI, 95% CI 0.05 to 0.08, P = 4.8×10^{-22}). These results demonstrate that BMI is positively associated with PCT, PLT (I), SFL and IPC in this population. These estimates were very similar to the age and sex only adjusted estimates (Table 1.4).

Table 1.3: Observational associations between BMI and platelet measures adjusted for age, sex, smoking status and alcohol consumption. βeta coefficient is the change in platelet measure in SDs per normalized SD increase in BMI. PCT = plateletcrit, PLT (I) = platelet count (impedance channel), PLT (F) = platelet count (PLT-F channel), SFL = side fluorescence, IPC = immature platelet count, PDW = platelet distribution width, FSC = forward scatter, H-IPF = high fluorescence immature platelet fraction, IPF = immature platelet fraction, P-LCR = platelet large cell ration, SSC = side scatter, MPV = mean platelet volume. Eta squared is the proportion of variance explained by the platelet trait in an ANOVA, whereas adjusted R squared is the variance explained by all the predictor variables in the regression model.

Platelet trait	N	βeta coefficient	SE	P value	R2	Eta squared	95% CI	Lower 95% CI	Upper 95% CI
PCT	24562	0.12	0.006	9.17e-88	0.133	0.107	0.012	0.11	0.13
PLT (I)	26044	0.11	0.006	1.03e-67	0.096	0.081	0.012	0.09	0.12
PLT (F)	23921	0.1	0.006	5.62e-57	0.094	0.082	0.012	0.09	0.11
SFL	26082	0.06	0.006	4.73e-23	0.02	0.000	0.012	0.05	0.07
IPC	23909	0.06	0.007	4.84e-22	0.008	0.001	0.013	0.05	0.08
PDW	24563	0.03	0.007	5.21e-05	0.006	0.004	0.013	0.01	0.04
FSC	26082	0.02	0.006	0.01	0.026	0.001	0.012	0.01	0.03
H-IPF	26068	-0.02	0.006	0.01	0.006	0.005	0.012	-0.03	-0.01
IPF	23909	0.01	0.007	0.04	0.008	0.007	0.013	0.00	0.03
P-LCR	27015	0.01	0.006	0.19	0.003	0.000	0.012	0.00	0.02
SSC	26082	0.01	0.006	0.28	0.04	0.000	0.012	-0.01	0.02
MPV	24559	0.01	0.007	0.37	0.007	0.000	0.013	-0.01	0.02

```
Warning in cbind(FullName, Name, N, Beta, SE, P, R, Eta, CI, Lower, Upper):
number of rows of result is not a multiple of vector length (arg 11)
```

Table 1.4: Age and sex adjusted estimates for the association between BMI and platelet traits. βeta coefficient is the change in platelet measure in SDs per SD higher BMI. Eta squared is the proportion of variance explained by the platelet trait in an ANOVA, whereas adjusted R squared is the variance explained by all the predictor variables in the regression model.

Full name	Short Name	N	βeta coefficient	SE	P value	R2	Eta2	95% CI	Lower 95% CI	Upper 95% CI
Plateletcrit	PCT	27997	0.126	0.006	1.46e-108	0.132	0.107	0.011	0.115	0.137
Platelet count (impedence channel)	PLT (I)	29682	0.11	0.006	4.5e-84	0.097	0.082	0.011	0.099	0.121
Platelet count (PLT-F channel)	PLT (F)	27268	0.104	0.006	6.5e-70	0.094	0.083	0.012	0.093	0.116
Side fluorescence	SFL	29721	0.064	0.006	4.08e-28	0.021	0.000	0.011	0.053	0.076
Immature platelet count	IPC	27255	0.067	0.006	1.27e-27	0.007	0.001	0.012	0.055	0.079
Platelet distribution width	PDW	27998	0.029	0.006	2.09e-06	0.006	0.004	0.012	0.017	0.041
Forward scatter	FSC	29721	0.022	0.006	0.00014	0.026	0.001	0.011	0.011	0.034
High fluorescence immature platelet fraction	H-IPF	29708	-0.017	0.006	0.00443	0.005	0.005	0.012	-0.028	-0.005
Immature platelet fraction	IPF	27255	0.015	0.006	0.0129	0.007	0.007	0.012	0.003	0.027
Side scatter	P-LCR	29721	0.01	0.006	0.0781	0.041	0.000158	0.011	-0.001	0.022
Platelet large cell ratio	SSC	30788	0.01	0.006	0.0909	0.002	4.37e-05	0.011	-0.002	0.021
Mean platelet volume	MPV	27994	0.009	0.006	0.155	0.006	4.57e-06	0.012	-0.003	0.137

1.3.4 Associations of covariables with BMI and platelet traits in INTERVAL

The association of covariables (age, sex, smoking and alcohol) with BMI were evaluated. Included covariables in the analysis showed associations with BMI (Table 1.5). Males had a higher BMI than females (0.18 SD, 95% CI 0.16 to 0.21, P = 9.64×10^{-64}). Age was positively associated with BMI (0.011 SD higher per year older, 95% CI 0.010 to 0.012, P = 1.11×10^{-182}). Alcohol showed an inverse association with BMI and smoking showed a weak positive association with BMI. Higher age was generally inversely associated with platelet measures (Table 1.6). Males had a lower platelet count and plateletcrit compared to females (PLT-F was 0.57 lower 95% CI -0.60 to -0.55, P = 9.9×10^{-324}, Table 1.7). Weak associations were detected between smoking and platelet traits such as a positive association between smoking status and immature platelets (Table 1.8). Higher alcohol consumption also showed inverse associations with measures of plateletcrit and platelet count (Table 1.9).

Table 1.5: Associations between covariables (exposure) and standardised BMI (outcome)

Variable	N	Beta coefficient per 1-unit increase in confounder (SDs)	Standard error	P value	Adjusted R2	F statistic	95% CI	Lower 95% CI	Upper 95% CI
Sex (1=female, 2=male)	33388	0.184	0.011	9.64e-64	0.008	285	0.021	0.162	0.205
Age (years)	33388	0.011	3.81E-04	1.11e-182	0.025	841	0.001	0.01	0.012
Alcohol intake frequency (1=rarely, 2=<1 a week, 3=1-2 times a week, 4=most days/every day)	29538	-0.096	0.006	1.57e-61	0.009	275	0.011	-0.108	-0.085
Smoking frequency (1=never, 2=occasinal, 3=most days/ every day)	32867	0.073	0.009	1.53e-17	0.002	72	0.017	0.057	0.09

Table 1.6: Association between age and platelet traits. βeta coefficient is the difference in platelet characteristics per 1 year increase in age.

Full name	Short Name	N	βeta coefficient	SE	P value	R2	95% CI	Lower 95% CI	Upper 95% CI
Side scatter	SSC	29721	-0.014	4.00E-04	4.59e-274	4.12E-02	7.84E-04	-0.015	-0.014
Forward scatter	FSC	29721	-0.011	4.03E-04	3.72e-160	2.41E-02	7.91E-04	-0.012	-0.01
Side fluorescence	SFL	29721	-0.009	4.05E-04	1.2e-111	1.68E-02	7.94E-04	-0.01	-0.008
Plateletcrit	PCT	27997	-0.009	4.18E-04	2.2e-92	1.47E-02	8.19E-04	-0.009	-0.008
Platelet count (impedence channel)	PLT (I)	29682	-0.006	4.07E-04	3.15e-51	7.58E-03	7.98E-04	-0.007	-0.005
Mean platelet volume	MPV	27994	-0.005	4.20E-04	8.11e-39	6.01E-03	8.22E-04	-0.006	-0.005
Platelet count (PLT-F channel)	PLT (F)	27268	-0.005	4.28E-04	1.09e-27	4.32E-03	8.38E-04	-0.006	-0.004
Platelet large cell ratio	P-LCR	30788	-0.003	4.01E-04	4.22e-17	2.26E-03	7.86E-04	-0.004	-0.003
Immature platelet count	IPC	27255	-0.003	4.28E-04	1.5e-13	1.96E-03	8.39E-04	-0.004	-0.002
Platelet distribution width	PDW	27998	-0.002	4.21E-04	1.32e-06	7.99E-04	8.24E-04	-0.003	-0.001
High fluorescence immature platelet fraction	H-IPF	29708	0.001	4.08E-04	0.0217	1.44E-04	8.01E-04	0	0.002
Immature platelet fraction	IPF	27255	-0.001	4.29E-04	3.79	1.21E-04	8.40E-04	-0.002	0

Table 1.7: Association between sex and platelet traits. Beta coefficient is the difference in platelet measure in SDs in men compared with women (for sex, 1=female and 2=male).

Full name	Short Name	N	βeta coefficient	SE	P value	R2	95% CI	Lower 95% CI	Upper 95% CI
Platelet count (PLT-F channel)	PLT (F)	27268	-0.574	0.012	9.88131298324e-02	1.68E-02	0.023	-0.597	-0.551
Platelet count (impedance channel)	PLT (I)	29682	-0.575	0.011	9.88131298324e-02	1.87E-02	0.022	-0.597	-0.554
Plateletcrit	PCT	27997	-0.663	0.011	9.88131291324e-01	1.80E-02	0.022	-0.685	-0.641
Immature platelet fraction	IPF	27255	0.164	0.012	4.41e-42	6.72E-03	0.024	0.141	0.188
High fluorescence immature platelet fraction	H-IPF	29708	0.142	0.012	2.24e-34	4.99E-03	0.023	0.119	0.164
Platelet distribution width	PDW	27998	0.118	0.012	3.76e-23	3.47E-03	0.023	0.095	0.142
Side scatter	SSC	29721	-0.076	0.012	5.76e-11	1.41E-03	0.023	-0.099	-0.053
Immature platelet count	IPC	27255	-0.075	0.012	4.75e-10	1.39E-03	0.024	-0.099	-0.052
Forward scatter	FSC	29721	0.039	0.12	0.000806	3.44E-04	0.023	0.016	0.062
Platelet large cell ratio	P-LCR	30788	-0.024	0.011	0.0329	1.15E-04	0.022	-0.047	-0.002
Side fluorescence	SFL	29721	-0.022	0.012	0.0592	8.61E-05	0.023	-0.045	0.001
Mean platelet volume	MPV	27994	-0.014	0.012	0.238	1.39E-05	0.023	-0.038	0.009

Table 1.8: Association between smoking and platelet traits. Beta coefficient is the change in platelet traits (SDs) per unit change in smoking category (1 = never, 2 = previous, 3 = current)

Full name	Short Name	N	βeta coefficient	SE	P value	R2	95% CI	Lower 95% CI	Upper 95% CI
High fluorescence immature platelet fraction	H-IPF	29241	0.047	0.009	2.81E-07	8.67E-04	0.018	0.029	0.065
Immature platelet count	IPC	26826	0.047	0.01	9.60E-07	8.57E-04	0.019	0.028	0.066
Immature platelet fraction	IPF	26826	0.038	0.01	6.74E-05	5.55E-04	0.019	0.019	0.057
Side fluorescence	SFL	29254	0.018	0.009	5.11E-02	9.59E-05	0.018	0	0.036
Forward scatter	FSC	29254	-0.014	0.009	1.34E-01	4.26E-05	0.018	-0.032	0.004
Mean platelet volume	MPV	30302	0.008	0.009	3.63E-01	-5.66E-06	0.018	-0.009	0.026
Platelet large cell ratio	P-LCR	29216	0.008	0.009	4.01E-01	-1.01E-05	0.018	-0.01	0.026
Platelet count (PLT-F channel)	PLT (F)	26839	0.008	0.010	4.07E-01	-1.17E-05	0.019	-0.011	0.027
Plateletcrit	PCT	27553	0.005	0.009	6.01E-01	-2.64E-05	0.019	-0.014	0.024
Platelet distribution width	PDW	27554	0.003	0.009	7.31E-01	-3.20E-05	0.019	-0.015	0.022
Side scatter	SSC	29254	0.003	0.009	7.72E-01	-3.13E-05	0.018	-0.015	0.021
Platelet count (impedance channel)	PLT (I)	29215	0.001	0.009	9.48E-01	-3.41E-05	0.018	-0.017	0.019

Table 1.9: Association between alcohol consumption and platelet traits. Beta coefficient is the change in platelet traits in SDs per unit increase in alcohol consumption (1=Rarely , 2= Less than weekly, 3=One or two weekly, 4= 3-5 weekly or every day)

Full name	Short Name	N	βeta coefficient	SE	P value	R2	95% CI	Lower 95% CI	Upper 95% CI
Plateletcrit	PCT	24757	-0.115	0.006	6.36E-73	1.31E-02	0.012	-0.128	-0.103
Platelet count (impedence channel)	PLT (I)	26251	-0.087	0.006	7.94E-45	7.45E-03	0.012	-0.099	-0.075
Platelet count (PLT-F channel)	PLT (F)	24110	-0.074	0.006	1.94E-30	5.40E-03	0.013	-0.087	-0.062
Side scatter	SSC	26289	-0.063	0.006	1.17E-24	3.95E-03	0.012	-0.076	-0.051
Forward scatter	FSC	26289	-0.06	0.006	3.81E-22	3.52E-03	0.012	-0.072	-0.048
Mean platelet volume	MPV	24754	-0.053	0.006	1.64E-16	2.70E-03	0.012	-0.065	-0.04
Side fluorescence	SFL	26289	-0.042	0.006	8.90E-12	1.73E-03	0.012	-0.054	-0.03
Platelet large cell ratio	P-LCR	27231	-0.036	0.006	4.39E-09	1.23E-03	0.012	-0.048	-0.024
Immature platelet count	IPC	24098	-0.033	0.006	2.55E-07	1.06E-03	0.013	-0.046	-0.021
Platelet distribution width	PDW	24758	-0.027	0.006	2.84E-05	6.67E-04	0.013	-0.039	-0.014
High fluorescence immature platelet fraction	H-IPF	26275	0.012	0.006	4.52E-02	1.15E-04	0.012	0	0.025
Immature platelet fraction	IPF	24098	0.002	0.006	8.10E-01	-3.91E-05	0.013	-0.011	0.014

1.3.5 GRS for BMI associations with BMI and covariables

To estimate the causal effect of BMI on platelet characteristics, MR was performed. The GRS for BMI (the average per SNP effect on BMI) showed a normal distribution (mean 0.08, SD 0.30, range -1.15 to 1.34) and was positively associated with BMI to the degree expected (R^2 = 0.042, P = $2.3x10^{-312}$, F=1458, Table 1.10). The GRS did not associate with sex, however weak associations were detected with alcohol consumption, age, and smoking status. The amount of variance explained by the GRS for BMI on any covariable included did not exceed R^2 = 0.002 (Table 1.10). As the GRS associates with BMI but does not strongly associate with the measured covariables, it is thus likely a valid instrument to perform MR.

Table 1.10: Association between genetic risk score for BMI with both BMI and covariables. βeta coefficient is the change in outcome variable per unit increase in the genetic risk score for BMI

Variable	N	Beta coefficient (per 1-unit increase in GRS)	Standard error	P value	Adjusted R2	F statistic
BMI (SDs)	33388	0.69	0.018	2.3E-312	0.042	1458.4
Age (years)	33388	-1.46	0.26	2.2E-08	0.001	31.3
Sex (1=female, 2=male)	33388	-0.01	0.009	4.7E-01	-1.4E-05	0.5
Smoking frequency (1=never, 2=previous, 3=current	32867	0.05	0.012	9.2E-05	4e-04	15.3
Alcohol intake frequency (1=rarely, 2=<1 a week, 3=1-2 times a week, 4=most days/every day)	29538	-0.15	0.02	2.9E-15	0.002	62.4

1.3.6 Mendelian randomization estimates for the association between BMI and platelet traits

In the MR analyses, BMI was associated with fewer traits than in the observational analysis (Table 1.11, 1.4). The causal estimate for the effect of BMI on SFL was 0.08 SDs per SD increase in BMI (95% CI 0.03 to 0.14, P = 0.003). This estimate was of larger magnitude than the observational estimate. The causal estimate for BMI and IPC was 0.06 SDs per SD increase in BMI (95% CI 0.006 to 0.12, P = 0.03), a similar magnitude of effect to the observational estimate. In the MR analysis, unlike in the observational analysis, the causal estimate did not suggest an effect of BMI on either PCT or PLT. MR estimates did not provide evidence for associations between BMI and other platelet variables. The point estimates were consistently positive, aligning with the positive observational associations seen across the platelet traits. Across all but one (H-IPF) of the platelet phenotypes we did not observe differences in directionality of effect comparing observational effects and causal effects predicted by Mendelian randomization (Figure 1.4). The Wu-Hausman test suggested that observational and MR estimates were similar for the majority of platelet traits ($P > 0.05$), except for measures of platelet count and plateletcrit ($P < 0.001$) (Table 1.11).

Table 1.11: Mendelian randomization estimates for the effect of BMI on platelet traits. βeta coefficient is the change in platelet measure in SDs per SD higher BMI

Full name	Short Name	N	βeta coefficient	SE	P value	R2	Eta2	95% CI	Lower 95% CI	Upper 95% CI	Wu-Hausman P value	F-statistic (GRS and BMI association)
Side fluorescence	SFL	29721	0.083	0.028	3.44E-03	0.132	0.107	0.056	0.027	0.139	0.973	676
Immature platelet count	IPC	27255	0.064	0.03	3.15E-02	0.097	0.082	0.058	0.006	0.122	0.672	641
Forward scatter	FSC	29721	0.052	0.028	6.62E-02	0.094	0.083	0.056	-0.003	0.108	0.774	652
Plateletcrit	PCT	27997	0.052	0.029	7.22E-02	0.021	0	0.057	-0.005	0.11	3e-04	706
Immature platelet fraction	IPF	27255	0.044	0.03	1.34E-01	0.007	0.001	0.058	-0.014	0.103	0.334	603
Platelet count (impedence channel)	PLT (I)	29682	0.035	0.028	2.19E-01	0.006	0.004	0.056	-0.021	0.091	0.001	729
Platelet large cell ratio	P-LCR	30788	0.034	0.028	2.25E-01	0.026	0.001	0.055	-0.021	0.088	0.527	678
Platelet distribution width	PDW	27998	0.035	0.029	2.29E-01	0.005	0.005	0.057	-0.022	0.093	0.936	625
Mean platelet volume	MPV	27994	0.029	0.029	3.23E-01	0.007	0.007	0.057	-0.028	0.086	0.746	614
Platelet count (PLT-F channel)	PLT (F)	27268	0.026	0.03	3.81E-01	0.041	0.000158	0.058	-0.032	0.084	0.001	671
Side scatter	SSC	29721	0.018	0.028	5.37E-01	0.002	4.37e-05	0.056	-0.038	0.073	0.433	661
High fluorescence immature platelet fraction	H-IPF	29708	0.015	0.028	6.10E-01	0.006	4.57e-06	0.056	-0.041	0.07	0.223	650

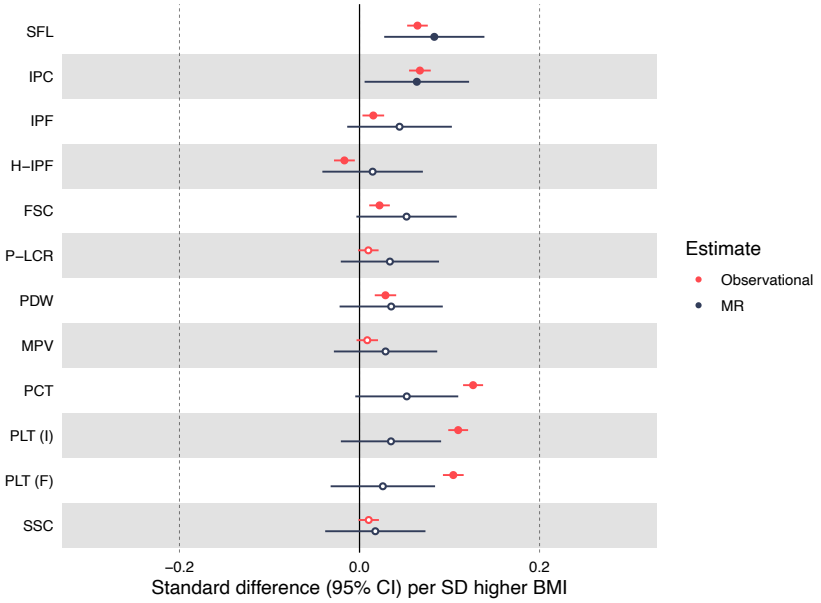

Figure 1.4: Forest plot of the Mendelian randomization (MR) and unadjusted observational estimates for the effect of BMI on platelet traits. Estimate points are filled where $P < 0.05$. SFL = side fluorescence, IPC = immature platelet count, H-IPF = high fluorescence immature platelet function, FSC = forward scatter, P-LCR = platelet large cell ratio, PDW = platelet distribution width, MPV = mean platelet volume, PCT = plateletcrit, PLT (I) = platelet count (impedance channel), PLT (F) = platelet count (PLT-F channel), SSC = side scatter.

1.3.7 Follow-up associations between IPC and whole blood aggregation in COPTIC

Given evidence for a causal effect of BMI on IPC in the MR analysis, we sought to evaluate the relationship between IPC and platelet activity as a biological parameter of clinical relevance. Whilst this analysis could not be conducted in INTERVAL due to a lack of suitable data, we were able to utilize data from the COPTIC study to address this question. The COPTIC study is a cohort of cardiac surgery patients, with samples taken pre-surgery. These participants have whole blood aggregation

measured, therefore making it possible to determine associations between IPC and aggregation in a clinical setting.

1.3.8 COPTIC participant characteristics

The total number of COPTIC participants was 2,541. Of these, 2,518 participants gave consent for future research (Table 1.12). Participants included in the analysis were those not on antiplatelet therapy (N=655). The majority of participants were male (61.8%), with a mean age of 63.9 years (SD of 16.1). Similar to the INTERVAL cohort, the mean BMI was in the overweight category (27.2 kg/m^2 with a SD of 4.9 kg/m^2). The majority of participants were either never smokers or ex-smokers for more than 5 years (89.2 %).

1.3.9 Association between IPC and aggregation in COPTIC cohort

To determine the potential functional effects of variation in IPC, the COPTIC cohort was used to assess the observational association between IPC and whole blood platelet aggregation in response to a range of platelet agonists (Figure 1.5A). In participants who were not on antiplatelet therapy (up to N=655), there was evidence for a positive association between IPC and aggregation induced by adrenaline (0.13 SD increase per SD increase in IPC, 95% CI 0.04 to 0.21, p=3.7x10-3), TRAP-6 (0.11 SD increase per SD increase in IPC, 95% 0.03 to 0.18, p=7.8x10-3) and ADP (0.08 SD per SD increase in IPC, 95% 0.01 to 0.0.15, p=0.04). As there was a high correlation among platelet traits (Figure 1.5A, Figure 1.5B), the effect of MPV, PLT and IPF were also assessed. PLT was associated with all four measures of aggregation, with a larger effect estimate than IPC (Figure 1.5A). The finding that IPC does not correlate with PLT however suggests that the effect of IPC on platelet

Table 1.12: Characteristics of COPTIC study participants

Variable	Mean (SD) or % (all)	N (all)	Mean (SD) or % (included)	N (included)
Age	66.7 (11.9)	2502	63.9 (16.1)	655
Sex		2518		655
Female	24.9 %		38.2 %	
Male	75.1 %		61.8 %	
Body mass index (kg/m2)	27.9 (4.6)	2422	27.2 (4.9)	649
Smoking		2469		654
Never	38.5 %		50.8 %	
Ex-smoker (>5 years)	47.4 %		38.4 %	
Ex-smoker (1-5 years)	8.3 %		6.3 %	
Ex-smoker (30 days to 1 year)	1.9 %		2.1 %	
Current smoker	3.9 %		2.4 %	
Antiplatelet medications		2467		655
0	26.6 %		100 %	
1	50.7 %			
2	22.5 %			
3	0.2 %			
ADP aggregation (AU)	138.2 (50.1)	2377	141 (46.5)	625
Adrenaline aggregation (AU)	53.0 (28.2)	2363	56.8 (32.3)	621
Arachidonic acid aggregation (AU)	86.9 (62.1)	2375	148.9 (52.4)	623
TRAP-6 aggregation (AU)	200.0 (50.1)	2373	198.7 (50.0)	623
Platelet count (x 109/L)	209.8 (60.5)	2409	200.9 (57.6)	637
Mean platelet volume (fL)	10.6 (1.0)	2400	10.6 (0.98)	634
Immature platelet fraction (%)	3.2 (1.9)	2379	3.3 (2.0)	628
Immature platelet count (x 109/L)	6.3 (3.4)	2383	6.2 (3.3)	629

aggregation is independent of the effect of PLT on platelet aggregation (Figure 1.5B). Fitting PLT alongside IPC as an additional predictor of aggregation had little effect on the IPC effect estimate providing further evidence of independent contributions from the two traits (Figure 1.5C). Adjustment for PLT in the regression model for the association between IPF and aggregation provided estimates of a similar magnitude to that of IPC and aggregation.

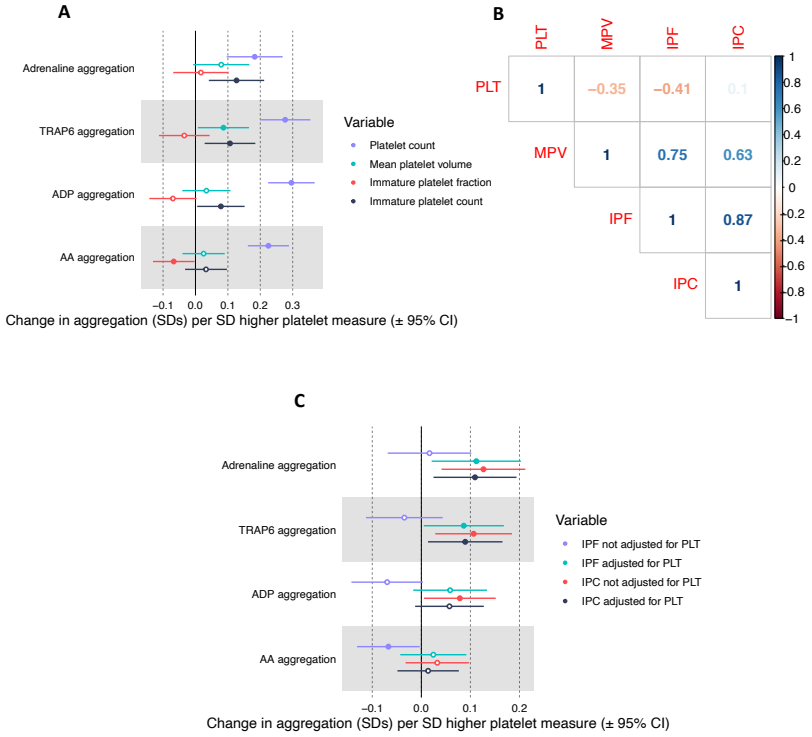

Figure 1.5: Association between platelet measures and whole blood platelet aggregation in COPTIC trial participants. A) Forest plot displaying the association between platelet measures (exposure) and aggregation (outcomes). B) Correlation matrix displaying the Pearson's correlation coefficient (r) between platelet traits. C) Forest plot displaying the effect estimate for the association between immature platelet count (IPC) or immature platelet fraction (IPF) and measures of aggregation, with or without adjustment for platelet count (PLT). ADP = adenosine diphosphate, AA = arachidonic acid, MPV = mean platelet volume.

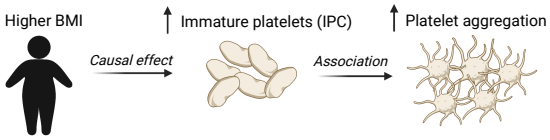

Figure 1.6: Graphical summary of findings

1.4 Discussion

In this study we used data from 33,388 healthy blood donors to study the effect of BMI on platelet phenotypes including measures of count, size and maturity. The combination of observational and MR estimates suggested a positive causal effect of BMI on SFL and IPC. The observational analysis revealed a strong association between BMI and both PLT and PCT, however the MR estimates did not provide evidence to support this association as causal. Observational analysis using data from a cardiac surgery cohort provided evidence for a positive association between IPC and aggregation induced by adrenaline, TRAP6 and ADP.

The observational and MR analyses in the current study provide evidence for a positive association between BMI and both IPC and SFL. As IPC is derived using SFL, the measure of mRNA content, these measures are highly correlated. Whilst few

studies have explored the direct association between BMI and immature platelet measures, there is some evidence for increased immature platelets in subjects with metabolic syndrome (MS) and subjects with type II diabetes compared with control subjects.[46,47] In these studies, participants with MS or type II diabetes also had a higher BMI, which could be a possible cause of this increase in immature platelets. Indeed, evidence from our analysis suggests that BMI, independent of metabolic syndrome or type II diabetes, may causally influence IPC and SFL.

As IPC is an indirect measure of platelet production,[39] these results suggest that an increased BMI may lead to an increased production of platelets. It has been reported that immature platelets have increased thrombotic potential[39,56] and may be predictive of cardiovascular events,[38,57] suggesting that an increase in the number of immature platelets may be of functional and clinical relevance. A follow-up analysis confirmed that higher IPC was associated with greater platelet aggregation independent of absolute platelet count within a cardiac surgery cohort. Positive associations were found between IPC and aggregation induced by adrenaline, TRAP-6 and ADP, suggesting general platelet hyperactivity with more immature platelets. A previous study in a coronary artery disease cohort also found a positive correlation between IPC and aggregation induced by arachidonic acid, collagen and ADP.[56] These findings suggest that IPC can be used as a proxy for platelet hyperactivity within a cohort with a history of cardiovascular disease. Measures of immature platelets are currently not considered in the clinic when prescribing antiplatelet therapy, however if a higher immature platelet count is demonstrative of newly produced, hyperactive platelets then this could be important in guiding dosage regimes to ensure sufficient platelet inhibition.[40] Observational studies have also found that patients with COVID-19 have elevated immature platelet counts and fraction, which could partly explain high rates COVID-19 induced thrombosis.[59]

Despite the association between BMI and IPC, there was less evidence for an effect of BMI on IPF. This lack of association may be because IPF is the proportion of immature platelets, therefore if someone had a higher number of immature platelets, but also a higher number of platelets overall (for example due to increased platelet lifespan), there would be no increase in the immature platelet fraction.

In general, previous studies have reported positive observational associations between BMI and both PLT and PCT.[42] More direct measures of body fat such as total fat mass, waist-hip ratio and waist circumference have also been reported to be positively associated with both PLT and PCT;[43] the observational evidence from the current study is in agreement with this. The MR analysis did not detect causal effects of BMI on measures of PLT/PCT, however the estimates were consistently positive. It is possible that IPC could be raised in absence of raised PLT as platelets may being produced at an increased rate, but could also be being destroyed at an increased rate. The small P values in the Wu-Hausman test for endogeneity for plateletcrit and platelet count variables support that the observational and MR estimates are different. Estimates could be biased due to confounding factors, reverse causation or other sources of bias. Possible routes of confounding could include stress, inflammation and nutrition, which could exert independent effects on both BMI and platelet count.

Both observational and MR estimates suggested that BMI is not associated with measures of platelet size, such as MPV and P-LCR. There are conflicting findings in the literature, with some studies suggesting that there is a positive association between BMI and MPV[44] and others showing no association.[42,43,45] The results of the current study suggest that associations seen previously between BMI and MPV are likely due to confounding of observational estimates. The lack of association between BMI and MPV may be surprising given a correlation coefficient of 0.62

between IPC and MPV. As BMI was positively associated with immature platelet count, and immature platelets tend to be larger, it may be surprising that BMI was not associated with measures of platelet size. However, immature platelets make up a small percentage of overall platelets (~3-6 %),[38,40] therefore the increase in immature platelets may not affect the overall median size of the whole platelet population.

Although this study suggests potential effects of BMI on platelet traits such as SFL and IPC, it does not provide mechanistic insight into how BMI exerts these effects. Previous studies have suggested that inflammation driven by adiposity can stimulate megakaryocyte proliferation, thereby increasing platelet numbers.[43] There is evidence that inflammatory mediators such as interleukin-6 (IL-6) could be one such factor.[60] Further study would be warranted to explore these mechanisms, as well as replicate the current findings, such as through independent population or clinical studies.

There are a few limitations to the study that should be recognised. Firstly, BMI was derived from self-reported height and weight. Although there is potential for this to bias observations, previous studies have found that self-reported BMI and BMI measured in the clinic are strongly associated.[61] The GRS also associates with BMI to the extent expected. Secondly, there may be other confounders which were not recorded within INTERVAL and therefore could not be accounted for in our models, such as, socio-economic position, which may affect both BMI and risk of thrombosis. Therefore, residual confounding of observational estimates cannot be ruled out. With respect to the observational analysis conducted in COPTIC, the sample size is modest, which may limit power to detect associations. Furthermore, this cohort required cardiac surgery and therefore it is possible that associations found may not be generalizable to the wider population, as participants also dis-

played greater aggregatory responses than the reference ranges.[62] However, these findings do indicate that immature platelets may be a biomarker of platelet hyperactivity in patients with a history of cardiovascular disease. It is important to note that the anticoagulant used (EDTA) has been reported to affect platelet parameters, such as increasing MPV and decreasing PLT.[63] Despite this, EDTA is the standard anticoagulant used for full blood counts in the NHS. The Sysmex analyzers used in INTERVAL and COPTIC were different models, however, it has been shown that there is strong association between platelet measures between the XE and newer XN analyzers.[64]

Altogether, we show observational and MR evidence that an increased BMI is associated with an increase in number of immature platelets. Observational evidence indicates that higher immature platelet count is associated with enhanced aggregation in a cardiac surgery cohort. Together, these results indicate that higher BMI may enhance platelet function and thrombosis by increasing platelet production and immature platelet count.

Chapter 2

The effect of obesity on platelet function: a clinical pilot study

The following results section is a clinical study. Recruitment into this study has been severely affected by COVID-19 as we had approvals to start recruitment in March 2020. All data in this chapter was collected by myself, apart from the platelet-neutrophil assay data, which was performed by Dr Chris Williams due to time constraints when receiving patient samples. The proteomic data was analysed by the Proteomics facility at the University of Bristol.

2.1 Background

2.1.1 Obesity is associated with cardiovascular events e.g. heart attacks and stroke

2.1.2 Platelets are a key cell involved in thrombosis

The aims of this clinical study were to determine whether an increased BMI is associated with an increase in platelet function and an alteration in platelet signalling compared to healthy controls. In light of the recent pandemic

2.2 Methods

2.2.1 Materials

Protease-activated receptor 1 (PAR-1)-activating peptide (SFLLRN-NH2) was from Bachem (Bubendorf, Switzerland), crosslinked collagen-related peptide (CRP-XL) from Prof. Richard Farndale (Department of Biochemistry, University of Cambridge, UK). Adenosine diphosphate (ADP) was from Sigma-Aldrich (Poole, UK). cOmplete mini protease inhibitor tablets and phosSTOP phosphatase inhibitors were from Roche Life Sciences (Welwyn Garden City, UK). The Pierce bicinchoninic acid (BCA) assay was from ThermoFisher Scientific (Altrincham, UK). Sodium Citrate Vacutainer® tubes, FixLyse, PE-Cy5-conjugated anti-CD42b, FITC-conjugated CD61 antibody, Fc block, PE-conjugated human platelet GPVI antibody, PAC1-FITC and anti-CD62P-PE antibodies were from BD (Wokingham, UK). FITC-conjugated anti-human CD41 and CD42b antibody were from BioLegend (London, UK).

2.2.2 Study population

This study was approved by the London Riverside Research Ethics Committee. Bariatric patients with a body mass index of >40kg/m^2 were recruited from the Department of Bariatric and Metabolic Surgery at Southmead Hospital, Bristol. Participants with a BMI between 18 and 25 kg/m^2 were recruited from the University of Bristol. Inclusion criteria for participants were: aged 18 years and over and able to give written informed consent. Participants were excluded if they were pregnant or lactating, had any known clotting or bleeding disorders (e.g. von Willebrand disease or drug-induced thrombocytopenia), or if they had taken any antiplatelet medication within the previous 14 days such as clopidogrel, ticagrelor or aspirin. Recruitment occurred between May and June 2021. As a pilot study, a total of four bariatric patients with obesity and four control participants were recruited. Bariatric patients donated blood before bariatric surgery.

2.2.3 Isolation of platelet rich plasma (PRP) and platelet lysates

Blood was taken by venipuncture into vacutainers containing sodium citrate (3.2%). Blood was centrifuged (1000 RPM, 17 mins). PRP was isolated and diluted 1:40 in HEPES Tyrode's (145 mM NaCl, 3 mM KCl, 0.5 mM Na_2HPO_4, 1 mM $MgS0_4.7H_2O$, 10 mM HEPES, pH 7.4, 0.1% [w/v] D-glucose). To create platelet lysates, acid citrate dextrose (1:7) and apyrase (0.02 U/mL) were added to isolated PRP. PRP was centrifuged (1700 RPM, 10 mins). Platelets were double washed in CGS buffer (120 mM NaCl, 25.8 mM sodium citrate dihydrate, 0.1% [w/v] D-glucose, 0.02 U/mL apyrase, pH 6.5). Platelets were resuspended in radioimmunoprecipitation assay buffer (RIPA: 25 mM HEPES, 200 mM NaCL, 1mM EDTA, 1 % (v/v) NP40, 0.5 % (w/v) sodium deoxychelate, 0.1 % (w/v) SDS) supplemented with a protease and phosphatase inhibitor for 10 minutes on

ice. Lysates were centrifuged at 10000 RPM, at 4°C for 5 minutes. Supernatant was removed and a bicinchoninic acid (BCA) assay was performed to calculate sample protein concentrations. Samples were stored in a -80°C freezer until used for proteomics analysis.

2.2.4 Platelet parameters measured by Sysmex

Citrated whole blood (200 uL) was analysed using the Sysmex XE-2100 haematology analyser. Platelet readouts were platelet count (PLT), immature platelet fraction (IPF), immature platelet count (IPC), side fluorescence (SFL, a measure of mRNA content), forward scatter (FSC, a measure of platelet size), side scatter (SSC, a measure of granularity).

2.2.5 Integrin $\alpha_{\mathbf{IIb}}\beta_{\mathbf{3}}$ activation and P-selectin expression measured by flow cytometry

Isolated PRP was diluted 1:40 in HEPES Tyrode's, with the antibodies FITC-conjugated PAC1 and PE-conjugated CD62P in a 2:1 ratio. Master mix was added to pre-prepared vacuum packed 96 well plates to a final folume of 50 uL. Agonists used were PAR1-AP, ADP and CRP, which were freeze-dried onto the plates,[65] as well as HEPES Tyrode's which was used for a measurement of unstimulated integrin $\alpha_{\mathrm{IIb}}\beta_3$ activation and P-selectin expression. Platelets were stimulated with agonists for 10 minutes and were fixed with 2 % paraformaldehyde (PFA) for 20 minutes. Plates were analysed using the BD Accuri C6 Plus flow cytometer (BD Biosciences, Wokingham, UK), where the median fluorescence intensity was derived from 10,000 gated platelets.

2.2.6 Surface receptor levels measured by flow cytometry

Diluted PRP (1:40) was used for unstimulated surface receptor detection. Antibody was added (1:10 v/v) to a final volume of 50 uL. Antibodies used include anti-human CD41, CD42b, CD61 (FITC-conjugated) and GPVI and CD110 (PE-conjugated). Platelets were gated and the median fluorescence was reported from a total of 10,000 events.

2.2.7 Platelet-neutrophil assay measured by flow cytometry

Citrated whole blood was diluted 1 in 10 in HEPES-Tyrode's buffer and incubated with FITC-conjugated anti-CD41 and PE-conjugated anti-C45 antibodies at room temperature for 15 mins. Samples were either stimulated (+ 5 ug/mL CRP) or unstimulated (+ vehicle). Samples were fixed with FixLyse for 10 minutes, followed by 4 % PFA for 10 minutes. Fluorescence was quantified using the flow cytometer, where 1000 neutrophil events were analysed based on a gating of CD45 and SSC parameters. Aggregates of platelets and neutrophils were defined by being both CD41+ and CD45+.

2.2.8 Tandem Mass Tag Mass Spectrometry (TMT-MS) quantification of platelet proteins

Platelet lysate samples (50 ug total protein) were digested with trypsin and labelled with TMT 15 plex reagents (Thermo Fisher Scientific, Loughborough, UK). Labelled samples were pooled. The pooled sample was evaporated to dryness then resuspended in formic acid (5%). The pooled sample was then desalted using a SepPak cartridge (Waters, Milford, Massachussetts, USA). Eluate from the SepPak cartridge was evaporated to dryness and resuspended prior to fractionation by high pH reversed-phase chromatography using an Ultimate 3000 liquid chromatog-

raphy system (Thermo Fisher Scientific). Further fractionation occurred using high pH RP fractions using the Ulitamte 3000 nano-LC system in line with an Orbitrap Fusion Tribrid Mass Spectrometer (Thermo Fisher Scientific).

2.2.9 Statistical analysis

Data was analysed using GraphPad Prism 8. If data displayed a normal distribution (based on Shapiro-Wilk p-value and W statistic), a parametric test was used (e.g. unpaired t-test or one way ANOVA), otherwise a nonparametric test was used. Concentration-response curve were plotted using a four parameter variable slope. To compare concentration-response curve parameters, logEC50s and curve maxes were compared. A Fisher's exact test was used to compare proportions of categorical variables. Due to limited power, caution should be taken when considering p-values.

2.3 Results

2.3.1 Participant characteristics

The characteristics of participants included in the study are shown in Table 2.1. The mean BMI in the control group was 21.8 kg/m^2 (SD 2.1 kg/m^2) and 49.3 kg/m^2 (SD 8.8 kg/m^2). Participants in the control group and bariatric patient group were age and sex matched. Bariatric patients exhibited similar smoking habits to control participants, however they displayed increased rates of T2D and previous symptomatic COVID-19 infection.

Table 2.1: Characteristics of included participants

Variable	Control mean (SD) or %	Bariatric patient mean (SD) or %	P value for difference
Age	42.2 (14.1)	46 (18.0)	0.77
Sex	75 %	75 %	1
Body mass index (kg/m2)	21.8 (2.1)	49.3 (8.8)	9e-04
Smoker	0 %	0 %	1
Hypertensive (140/90 mmHg)	0 %	25 %	1
Type 2 Diabetes	0 %	50 %	0.43
Type 2 Diabetes medication	NA		NA
Metformin		25 %	
Dapagliflozin		25 %	
Dulaglutide		25 %	
Previous COVID-19 infection	0 %	75 %	0.14

2.3.2 Platelet parameters

Platelet traits measured by Sysmex were compared across bariatric and control groups (Table 2.2). There was no statistical evidence for an effect of obesity on platelet parameters. However, IPC and IPF did display larger mean values compared to controls, but they also had relatively large SDs in the bariatric patient group. A higher N would be required to detect whether there is any difference between groups.

2.3.3 Basal receptors

Unstimulated platelet receptor levels were compared across bariatric patient and control participant groups. Basal integrin $\alpha_{\mathrm{IIb}}\beta_3$ activation and P-selectin expression were similar across groups. Similarly, other platelet receptors were not different across groups.

Table 2.2: Platelet traits measured by Sysmex XE-2100

Variable	Control mean (SD)	Bariatric patient mean (SD)	P value for difference
Platelet count	257.8 (29.1)	235.8 (61.5)	0.54
Immature platelet fraction	3.8 (2.3)	7.3 (8.3)	0.44
Immature platelet count	9.3 (4.9)	16.3 (16.3)	0.45
Side fluorescence	82.0 (5.8)	85.1 (9.8)	0.61
Forward scatter	53.8 (5.5)	60.5 (8.8)	0.24
Side scatter	40.3 (1.1)	41.8 (3.0)	0.37

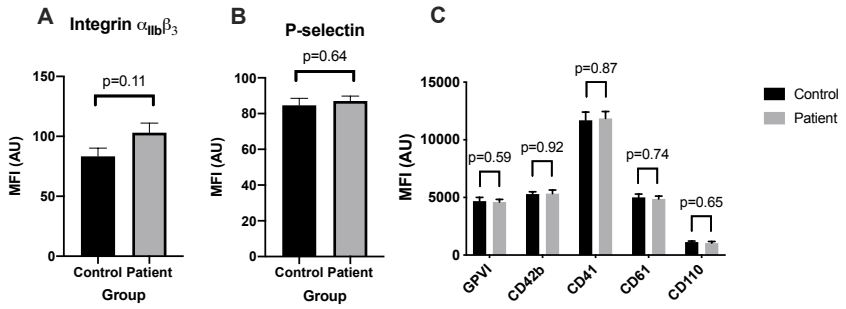

Figure 2.1: Diluted PRP was used for experiments, where platelets were incubated with antibodies for 10 minutes before fixing. Bar graphs comparing A) Basal integrin $\alpha_{IIb}\beta_3$ activation B) Basal P-selectin expression and C) Basal surface receptor expression across bariatric patient and control groups. N=4. P-values indicated for unpaired t-test results.

2.3.4 Agonist-induced integrin $\alpha_{IIb}\beta_3$ and P-selectin expression

Markers of platelet activation were compared in response to different platelet agonists.

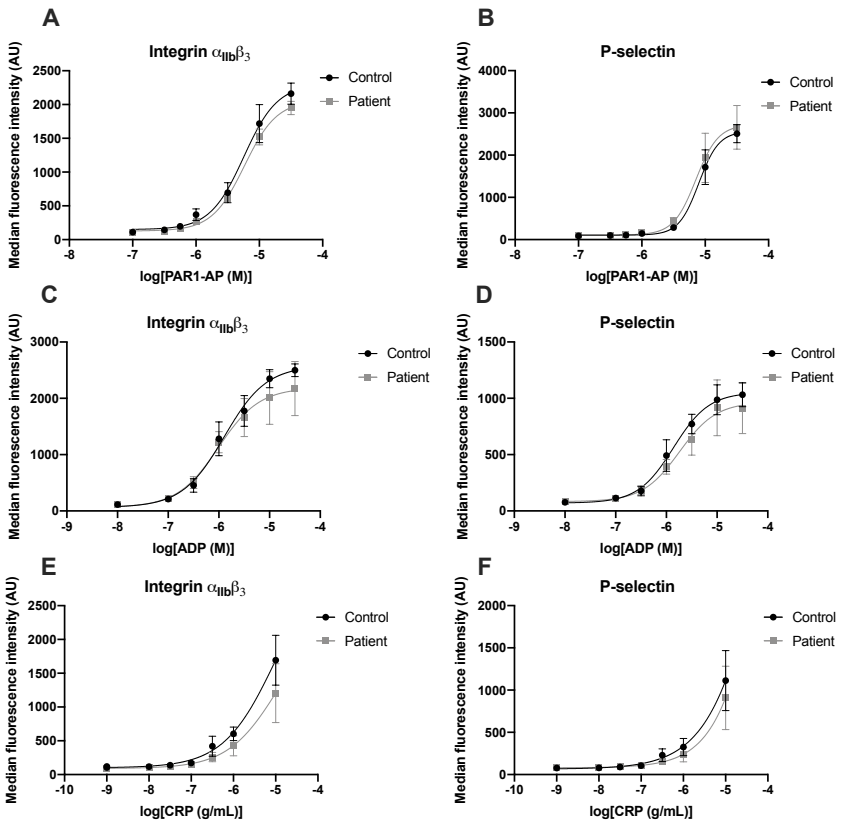

Figure 2.2: Concentration-response curves of integrin $\alpha_{\mathrm{IIb}}\beta_3$ activation in platelet rich plasma diluted in HEPES-Tyrode's in response to A) PAR1-AP, B) ADP and C) CRP. Concentration-response curves of P-selectin expression in response to D) PAR1-AP, E) ADP, F) CRP. N=4.

2.3.5 Platelet-neutrophil aggregates

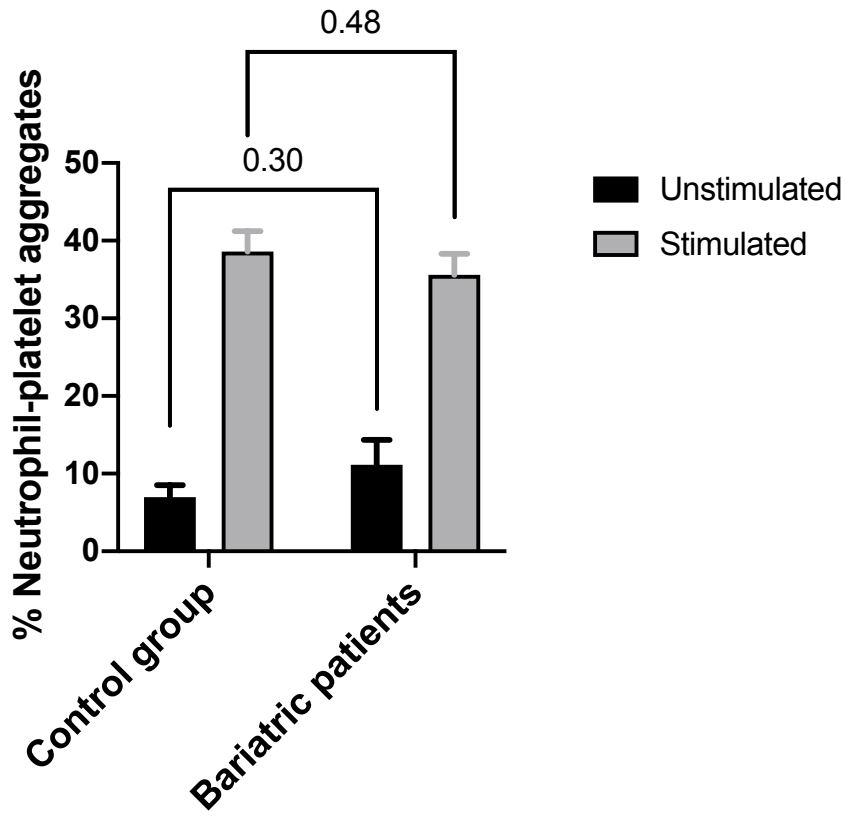

Figure 2.3: Comparison of platelet-neutrohpil aggregates across bariatric patient and control participant groups in unstimulated and stimulated (5 ug/mL CRP) conditions. N=4. P-values provided for the results of a two-way ANOVA with Sidak's multiple comparisons tests.

2.3.6 Proteomics

A PCA was performed. . . .

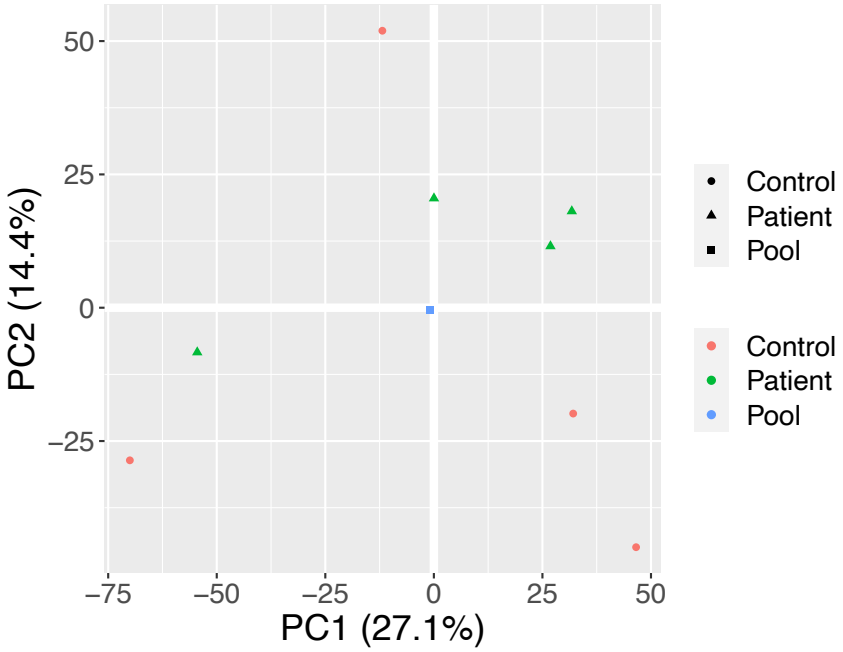

Figure 2.4: A principal component analysis to determine whether PC1 and PC2 separate out control and bariatric groups

A volcano plot was . . .

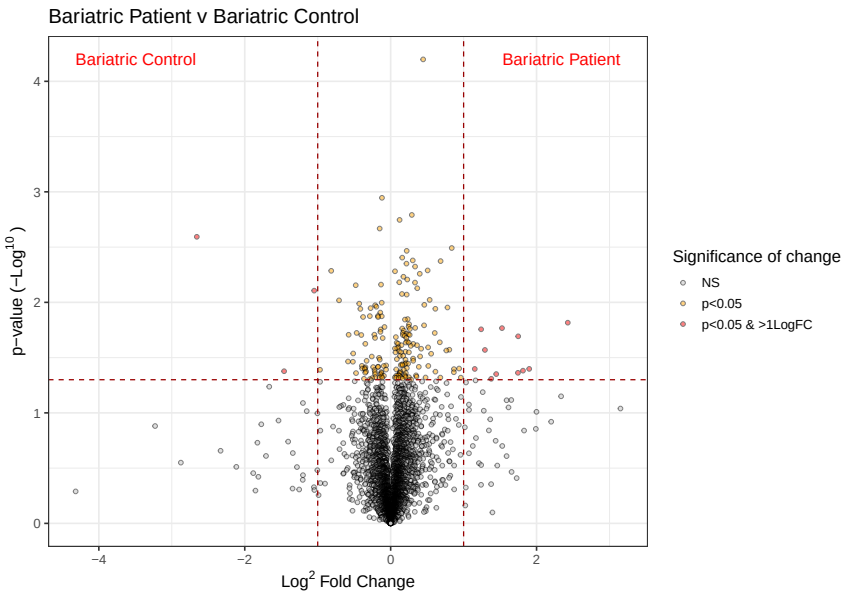

Figure 2.5: A principal component analysis to determine whether PC1 and PC2 separate out control and bariatric groups

The full proteomic results are here . . .

2.4 Discussion

This study has explored the effect of obesity on platelet parameters and platelet function. Initial analyses are inconclusive as the experiment lacks power. Despite this, optimisation of experimental techniques used within the study provides a means of exploring effects in the future.

Overall effects We have been able to optimise common platelet experiments to be more high throughput for patient studies. Limitations - what we initially aimed to do wasnt possible so power is limited.

Chapter 3

Pathways linking BMI and platelet function: do the chemokines MDC and TARC play a role?

3.1 Background

Platelets play an important role in haemostasis.[66] In healthy people, prostacyclin (PGI2) and nitric oxide (NO) are released by the endothelium to suppress platelet activity.[67] When the endothelium is damaged, platelets adhere to the injured vessel wall through the glycoprotein Ib-IXV receptor and the GPVI collagen receptor. As a result, platelets secrete alpha and dense granules and undergo shape change.[68] Alpha granules release growth factors (e.g. IGF-1), clotting factors and chemokines,[69] whereas dense granules release molecules such as ADP, which fur-

ther activate platelets by interacting with the $P2Y^1$ and $P2Y^{12}$ platelet receptors. Platelet activation also results in activation of integrins such as $\alpha_{IIb\beta}3$~ and subsequent fibrinogen binding results in platelet aggregation, thrombus formation and cessation of bleeding.[66] Despite these processes being essential for haemostasis, when platelets become hyperactive, the balance is tipped in favour of thrombosis.

Platelets likely contribute to the increased CVD risk related to obesity and type II diabetes as platelets from patients with obesity are reported to be hyperactive and have a lower sensitivity to antiplatelet therapies such as clopidogrel.[32] These therapies are critical in the prevention of ischaemic events. The underlying mechanism that causes platelet hyperactivity is not completely understood. It is likely to involve a combination of factors. One such factor is a decrease in the production of endothelial PGI2 and nitrous oxide.[70] Another contributing factor could be an alteration in the circulating levels of soluble ligands, growth factors and cytokines which can enhance platelet function (known as platelet primers). Platelet primers are signaling molecules which enhance agonist-induced platelet function but in isolation are unable to stimulate platelet aggregation or integrin activation. Molecules that have been reported to be able to prime platelets include insulin-like growth factor-1 (IGF-1)[32] and thrombopoietin (TPO).[71]

Previous studies have shown that multiple platelet primers increase platelet function through activation of the lipid kinase phosphoinositide-3-kinase (PI3K).[72–74] Another pathway that has been implicated in platelet priming effects is the MAPK pathway. TPO can enhance TxA2 synthesis through increasing the activation of extracellular signal regulated kinase (ERK2).[75,76] The common involvement of both PI3K and MAPK in platelet priming mediated by IGF-1, TPO and matrix metalloproteinase-2 (MMP-2) suggests that platelets may also be activated by

other platelet primers. There are other biomarkers which are also suggested to be raised with obesity as a result of low-grade chronic inflammation,[19] including the chemokines macrophage-derived chemokine (MDC/CCL22) and thymus and activation regulated chemokine (TARC/CCL17).[77] The inflammatory chemokines MDC and TARC act at the receptor CCR4, which is expressed on platelets.[78] These chemokines have been shown to potentiate platelet aggregation.[79] For example, the presence of TARC and MDC have been demonstrated to greatly enhance platelet aggregation in the presence of low levels of ADP or thrombin.[79] Furthermore, gene polymorphisms of MDC and CCR4 are associated with myocardial infarction (MI).[80] As blood/plasma levels of MDC have been reported to be elevated in obese people[77] and polymorphisms of its receptor are associated with MI, understanding the mechanisms by which MDC can enhance platelet function may give us an important insight in to how platelets are hyperactive in obesity and potentially provide a novel target for new antiplatelet/thrombotic or preventative therapies for those with an increased risk of cardiovascular events.

The aim of this chapter is to determine the effects and mechanisms of MDC and TARC on platelet function and signalling.

3.2 Methods

3.2.1 Materials

PAR1 activating peptide (SFLLRN) was purchased from Bachem. MDC and TARC were from PeproTech (London, UK). Indomethacin, PGE1, Triton X-100, apyrase, HEPES Tyrode's (10 mM HEPES, 1 mM magnesium, 0.5 mM monosodium phosphate, 145 mM sodium chloride, 3 mM potassium chloride, pH 7.2), 2-butanol, ethanol, methanol and bovine serum albumin (BSA) were

purchased from Sigma-Aldridge (Poole, UK). Phosphate buffered saline (PBS), trisodium citrate, acid citrate dextrose (ACD), D-glucose, ammonium persulfate (APS), Tris buffered saline (TBS), polysorbate 20 (Tween 20), Donkey anti-Rabbit IgG (H+L) Highly Cross Adsorbed Secondary Antibody Alexa Fluor 647 and Fura-2 were purchased from Thermo Fisher Scientific (Loughborough, UK). FITC PAC1 and FITC-CD62P from BD bioscience (Berkshire, UK). 30 % acrylamide mix (37.5:1), 1.0 M Tris pH 6.8 solution and 1.5 M Tris pH 8.8 solution purchased from National Diagnostics (Nottingham, UK). Immobilon-FL PVDF membranes (0.45-micron filter) were purchased from Merck Millipore (Hertfordshire, UK). Sodium dodecyl sulphate (SDS) was purchased from GE Healthcare worldwide (Hatfield, UK). SDS solution (10 %), tetramethylethylenediamine (TEMED), filter paper and Precision Plus protein pre-stained standards were purchased from Bio-Rad laboratories (Bredbury, UK). Blocking buffer was purchased from LI-COR Biosciences (Cambridge, UK). pSer473 PKB, pSer239 VASP, pSer157 VASP, pThr202/Tyr204 ERK and pSer425 Talin were purchased from Cell Signaling Technology (Hertfordshire, UK). Alexa Fluor 680 AffiniPure Donkey Anti-Goat IgG and Alexa Fluor 680 AffiniPure Donkey Anti-Rabbit IgG were obtained from Jackson Immunoresearch (Cambridgeshire, UK). Paraformaldehyde was purchased from Acros Organics (Geel, Belgium).

3.2.2 Isolation of human platelets

Fresh human venous blood was obtained from healthy volunteers by venipuncture. Blood was taken into the syringe with a 1 in 10 volume of 4 % trisodium citrate, then mixed with a 1 in 7 volume of acid citrate dextrose (ACD). Blood was placed into 5 mL LP4 tubes then centrifuged at 1000 revolutions per minute (RPM) for 17 minutes. Platelet rich plasma (PRP) was extracted. PRP was supplemented

with inhibitors such as 0.02 U/mL apyrase, 140 nM PGE1 or 10 μM indomethacin (exact inhibitors for each experiment used will be shown in results). Supplemented PRP was either used for the experiment of centrifuged for a further 10 minutes at 1700 RPM. Platelet poor plasma (PPP) was removed leaving a platelet pellet. This pellet was resuspended in HEPES Tyrode's (supplemented with 0.1 % glucose and 0.02 U/mL apyrase, and 10 μM indomethacin if already applied to PRP). Platelets were counted using a Z1 coulter particle counter by diluting 1 in 2000 in 10 mL MQ water. Platelets were diluted further using the supplemented HEPES Tyrode's to a platelet concentration of 4 x 10^8/mL and left to rest for 30 mins in a 30°C water bath.

3.2.3 Platelet aggregation

Platelets were diluted to a concentration of 2 x 10^8/mL using supplemented HEPES Tyrode's buffer. A volume of 250 μL platelets was added to aggregometer cuvettes containing a Teflon stir bar. HEPES Tyrode's (500 μL) was added to the reference PPP (blank) channel. Platelets were left to rest in the cuvettes for 2 mins. For priming experiments, platelets were either preincubated with vehicle (MQ) or 1 μg/ml MDC or TARC for 5 mins. A Chrono-log model 490 aggregometer was used. For experiments which were not exploring platelet priming, platelets were added to the cuvettes for 2 minutes and agonist was added for 5 minutes. Aggregation traces were derived from the amount of light transmitted through the sample. Recordings were taken for 5 minutes with a stirring speed of 1200 rpm and at 37 °C. Platelets were used within 3 hours from resting.

3.2.4 Plate platelet aggregation

PRP supplemented with apyrase was used. For each donor, the platelet count was recorded. PPP was used as a reference. PRP was added to the wells of a 96 well plate and left to rest for 2 minutes. 200 ng/ml MDC, 200 ng/mL TARC or vehicle were preincubated with platelets for 5 minutes. After 5 minutes, PAR1-AP was added to the wells and the plate was shaken with an Eppendorf ThermoMixer® for 5 minutes at 1200 rpm and at 37 °C. The plate was immediately read using a Labtech LT-4500 automatic microplate absorbance reader. Responses were normalized so that basal PRP = 0 % aggregation and PPP = 100 % aggregation.

3.2.5 Flow cytometry: integrin $\alpha_{\mathbf{IIb}}\beta_{\mathbf{3}}$ activation and P-selectin expression

Platelets were diluted to a concentration of 2 x 10^7/mL using supplemented HEPES Tyrode's buffer. Platelets were incubated with 200 ng/ml MDC, TARC or vehicle for 5 mins, then added to wells in a 96 well plate already containing a mixture of PAC1 and CD62P antibodies (in a 2:1 ratio), as well as increasing concentrations of PAR1-AP. After a further 5 minutes, platelets were fixed with 1 % paraformaldehyde (PFA). Plates were read using a BD Accuri C6 Plus flow cytometer. The relative fluorescence of the antibody in each well was measured over 10,000 platelets and median statistics were used.

3.2.6 Flow cytometry: phospho-VASP

Washed platelets were diluted to a concentration of 1 x 10^8/mL. Platelets were left to rest in a 96 well plate for 2 minutes. 200 ng/ml MDC, 200 ng/mL TARC, 10 µM ADP or vehicle were preincubated with platelets for 5 mins at 37 °C. PGE1 (concentration-response curve) was added for a further 5 minutes. Platelets were

fixed with 1 % PFA for 10 minutes. Platelets were centrifuged at 1000g for 10 minutes. Supernatant was then discarded, and platelets were permeabilized with 0.1% TritonTM X-100 for 10 minutes. The plate was centrifuged again at 1000g for 10 minutes. Supernatant was discarded and platelets were washed and resuspended with 1 x PBS. Platelets were centrifuged at 1000g for 10 minutes. Supernatant was discarded and platelets were resuspended with rabbit pSer157 VASP primary antibody (1 in 1000) for 30 minutes on ice. Platelets were centrifuged at 1000g for 10 minutes. Supernatant was discarded and platelets were washed and resuspended with 1 x PBS. Platelets were centrifuged at 1000g for 10 minutes. Supernatant was discarded and platelets were resuspended with secondary antibody (Alexa Fluor™ 647 anti-rabbit IgG, 1 in 1000) for 45 minutes on ice. Platelets were centrifuged at 1000g for 10 minutes. Supernatant was discarded and platelets were washed with 1 x PBS. Platelets were centrifuged once more at 1000g for 10 minutes and the plate was read using a BD Accuri C6 Plus flow cytometer.

3.2.7 Western blotting

Platelets were used at a concentration of 4 x 10^8/mL. Platelets were preincubated with 200ng/mL MDC, 200ng/mL TARC or vehicle for 5 minutes followed by 5 minutes application of PAR1-AP (all at 30 °C). For phospho-VASP assay, MDC, TARC or ADP were preincubated with platelets for 5 minutes, with PGE1 added for a further 5 minutes. After 10 minutes, platelets were lysed with 4 X LDS NuPAGE sample buffer supplemented with 50 mM dithiothreitol (DTT). Samples were stored at -20 °C. When required, samples were heated to 70 °C for 10 minutes. A 10 % Resolving gel was cast for Tris-glycine SDS-polyacrylamide gel electrophoresis, and a 5 % Stacking gel was cast for loading. As a standard, 4 µl Precision Plus protein pre-stained standard was loaded, along with 20 µl of the

lysate sample. The samples were placed in a tank containing sample buffer and protein electrophoresis was carried out for 90 minutes at 100 V. Following this, proteins were transferred to a PDVF membrane in the presence of transfer buffer for 60 minutes and at 100 V. The membrane was then put in blocking buffer and 0.1 % TBS-Tween (1:1 ratio) for 1 hour. Membranes were incubated with the primary antibody (1 in 1000 dilution) at room temperature for 90 minutes. Talin was used as a loading control. The membrane was washed with 0.1 % TBS-Tween 3 times for 5 minutes. The membrane was incubated with secondary antibody (1 in 5000 dilution) for 1 hour. The membrane was washed 6 times for 5 minutes, then membranes were scanned using the LI-COR Odyssey® CLx Imaging System. Bands were quantified using Image Studio Lite.

3.2.8 Calcium mobilisation

Fura-2 (4 μM) was added to PRP after the addition of indomethacin and apyrase and left for 1 hour wrapped in foil at 37 °C. PRP was centrifuged and platelets were resuspended in supplemented HEPES Tyrode's as described above, with the washed platelets (at 4 x 10^8/mL) protected from the light with foil. Platelets were recalcified with 1 mM calcium chloride. Fluorescence of fura-2 was measured as the ratio at wavelengths 340:380 nm and provided a measurement of free calcium. 10 basal recordings were taken over 1 minute before addition of MDC, TARC or ADP. Upon agonist stimulation, 20 readings were taken over 2 minutes. A final concentration of 1 % Triton X-100 was added to permeabilise platelets and further readings were taken.

3.2.9 Statistical analysis

GraphPad Prism 7 was used to analyse data. A Shapiro-Wilks test was used to test for normality of the data. If normally distributed, a parametric test was used. Where data was not normally distributed a non-parametric test was used. Statistical tests used for each experiment are in figure legends. Error bars plotted on graphs are standard error of the mean (SEM). Concentration-response curves were generated using a four parameter variable slope nonlinear regression. Parameters from curves including the logEC50 and the curve max were used as summaries of the curve fit.

3.2.10 Two sample MR lookup using EpigrahDB

Two sample MR uses summary statistics and allows for causal associations to be estimated where the exposure and outcome are measured in different populations. EpigraphDB is a tool which performs two sample MR to determine how the human proteome contributes to complex diseases.[81] Using protein quantitative trait loci (pQTLs) for MDC and TARC[82] (N=3301) and GWAS summary statistics for disease outcomes (mainly from Neale lab: `https://github.com/Nealelab/UK_Biobank_GWAS`), it is possible to explore whether plasma levels of these chemokines contribute to disease. For MDC, a Wald ratio was used to estimate the causal effect and for TARC the inverse variance weighted (IVW) method was used.

3.3 Results

3.3.1 Priming effects of MDC and TARC on platelet aggregation in washed platelets using light transmittance aggregometry

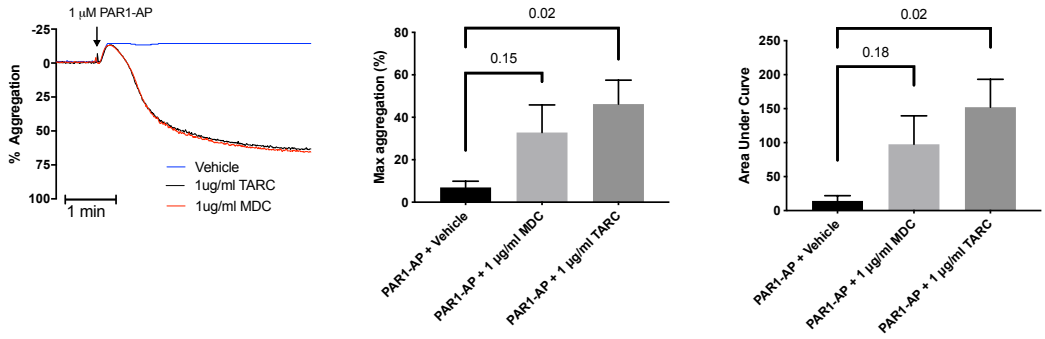

Figure 3.1: Washed platelets at 2x10^8/mL were preincubated with MDC or TARC (5 mins) before application of PAR1-AP-induced aggregation (5 mins). A) Raw aggregation trace comparing the effect of vehicle, 200 ng/mL MDC or 200 ng/mL TARC on PAR1-AP induced aggregation. B) Bar chart comparing the maximum aggregation reached when adding vehicle, MDC or TARC. C) Bar chart comparing the area under the curve of the aggregation trace in response to vehicle, MDC or TARC. Results analysed with a repeated measures one-way ANOVA with Dunnett's multiple comparisons. N=7. Mean+SEM displayed.

3.3.2 Priming effects of MDC and TARC on PAR1-AP induced integrin $\alpha_{IIb}\beta_3$ activation and P-selectin expression

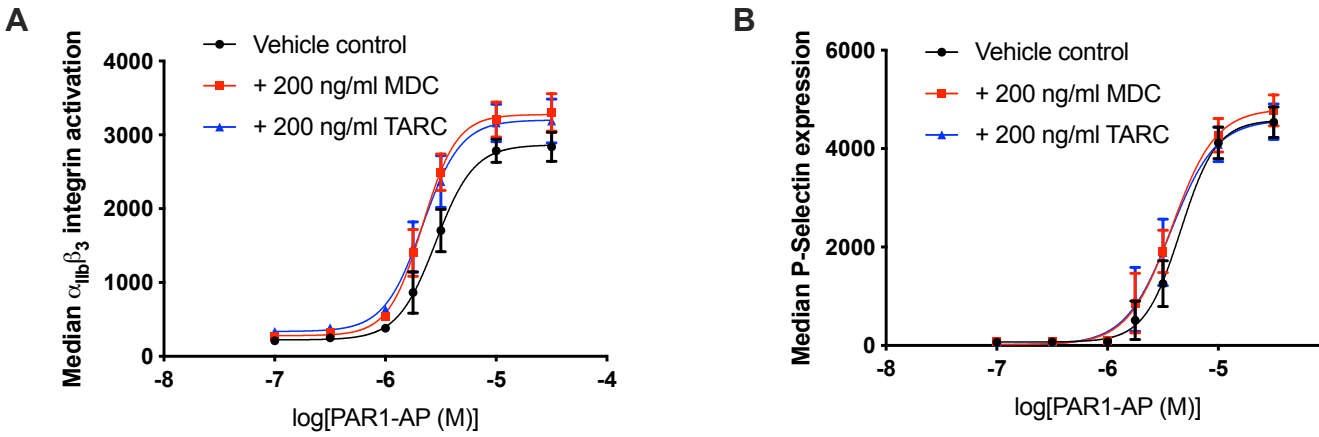

Figure 3.2: Washed platelets at 2x10^7/mL were preincubated with MDC or TARC (5 mins) before application of PAR1-AP (5 mins). Integrin $\alpha_{IIb}\beta_3$ activation and P-selectin expression were measured using flow cytometry. A) The effect of vehicle, 200 ng/mL MDC or 200 ng/mL TARC on PAR1-AP induced $\alpha_{IIb}\beta_3$ activation. B) The effect of vehicle, MDC or TARC on PAR1-AP induced P-selectin expression. LogEC50s compared using a one-way ANOVA with Dunnett's multiple comparisons in Table 3.1

Table 3.1: Comparison of the effect of MDC and TARC on the logEC50 for PAR1-AP induced integrin activation and P-selectin expression compared with a one-way ANOVA (N=4)

Condition	Integrin activation logEC50 (M)	Integrin SEM	Integrin P value	P-selectin logEC50 (M)	P-selectin SEM	P-selectin p value
Vehicle	-5.54	0.1		-5.36	0.07	
MDC	-5.66	0.08	0.019	-5.44	0.07	0.006
TARC	-5.64	0.11	0.014	-5.46	0.11	0.21

3.4 Effects of MDC and TARC on PAR1-AP induced PS exposure

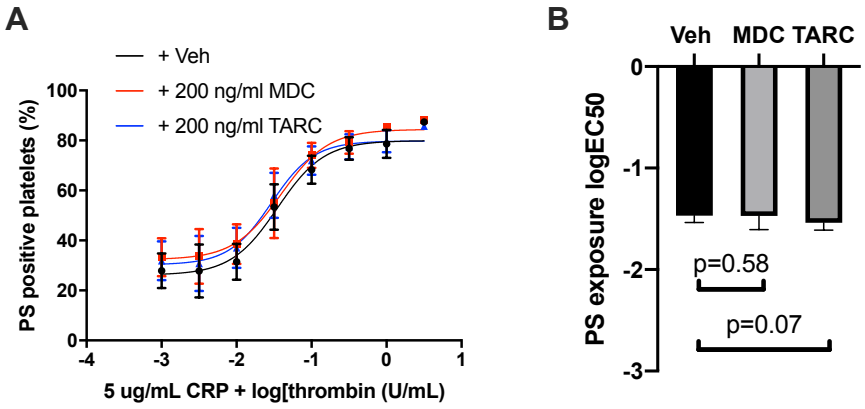

Figure 3.3: A) Washed platelets at $2x10^7$/mL were preincubated with vehicle, 200ng/mL MDC or 200 ng/mL TARC (5 mins) before application of increasing concentrations of PAR1-AP (10 mins). PS exposure was measured using flow cytometry (N=4). B) Bar chart comparing the effect of vehicle, MDC or TARC on the logEC50s of PAR1-AP induced PS exposure. LogEC50s compared by a Friedman's test.

3.4.1 Mechanism of MDC and TARC priming in washed platelets

3.4.2 Effects of MDC and TARC on phospho-VASP levels

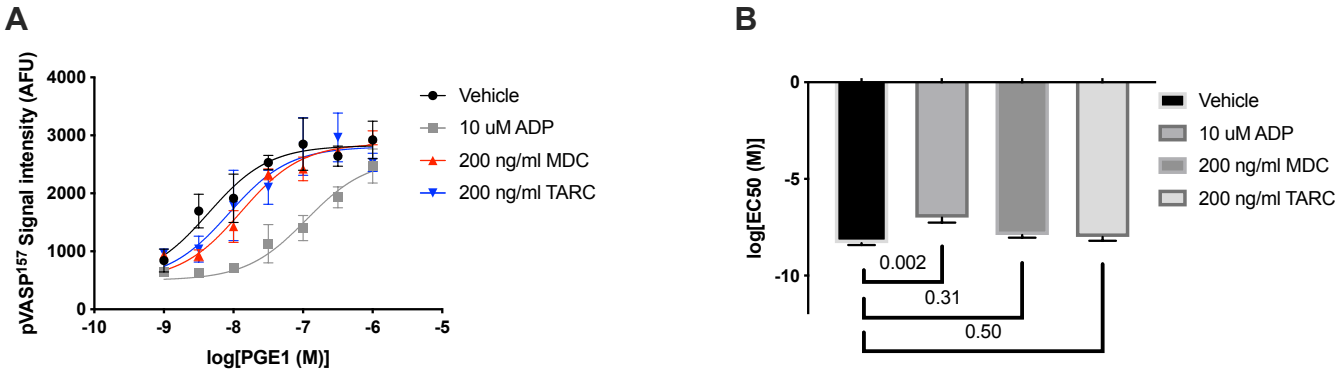

Figure 3.4: A) Washed platelets at $1x10^8$/mL were preincubated with vehicle, 200ng/mL MDC, 200 ng/mL TARC or 10 uM ADP (5 mins) before application of increasing concentrations of PGE_1 (5 mins). PS exposure was measured using flow cytometry (N=3). B) Bar chart of the logEC50s of each concentration response curve from A. LogEC50s were compared using a one-way ANOVA with Dunnett's multiple comparisons.

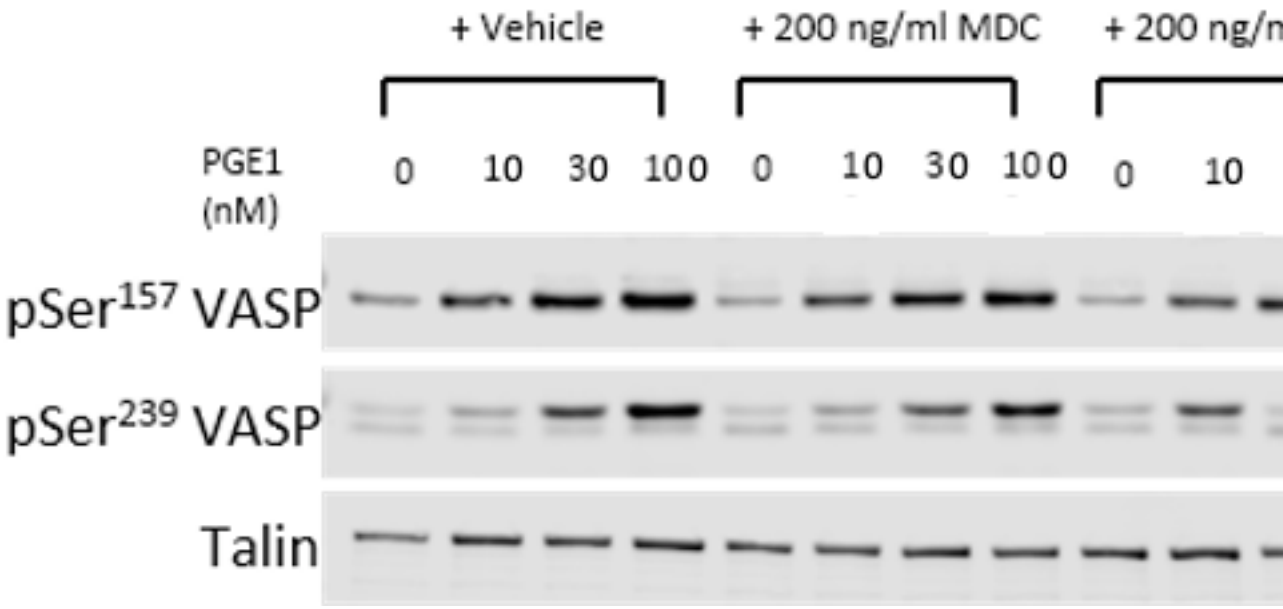

Figure 3.5: A representative blot of the effect of 200 ng/mL MDC, 200 ng/mL TARC and 10 uM ADP on phospho-VASP at Serine 157 and 239 in the presence of increasing concentrations of PGE_1

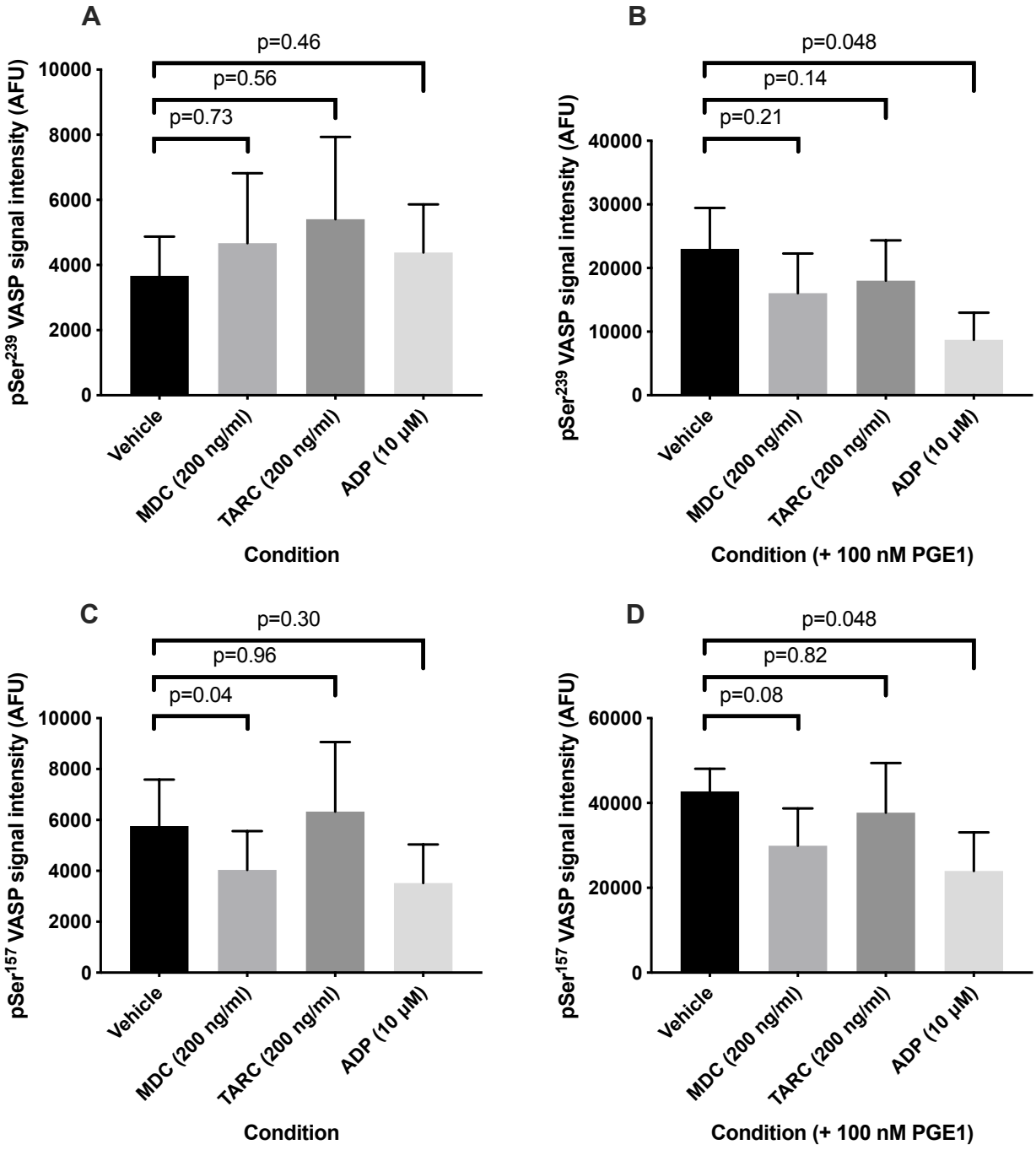

Figure 3.6: Bar charts of the effect of 200 ng/mL MDC, 200 ng/mL TARC and 10 uM ADP on phospho-VASP at A) Serine 239 in the absence of PGE_1, B) Serine 239 in the presence of 100 nM PGE_1, C) Serine 157 in the absence of PGE_1, D) Serine 157 in the presence of 100 nM PGE_1. N=4. Results analysed with a repeated measures one-way ANOVA with Dunnett's Multiple comparisons.

3.4.3 Effects of MDC and TARC alone on aggregation in washed platelets

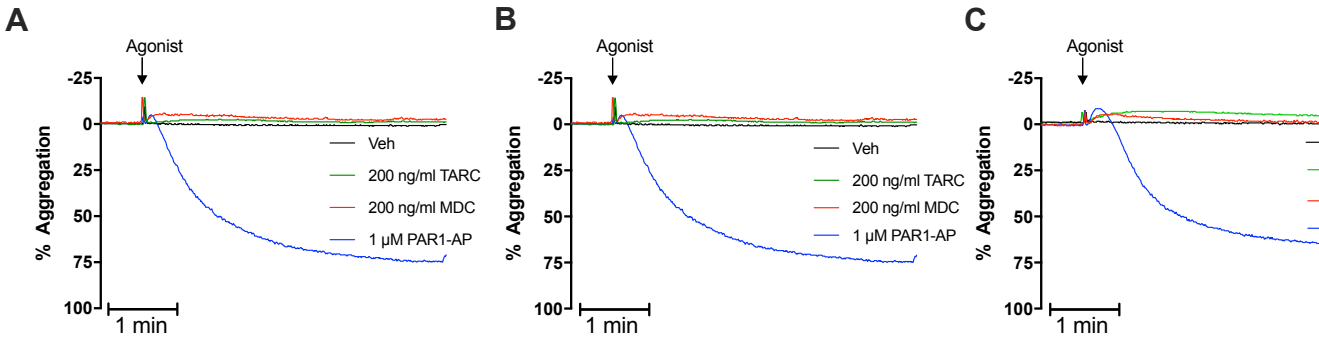

Figure 3.7: A-C) Washed platelets at $2x10^8$/mL were stimulated with vehicle, 200ng/mL MDC, 200 ng/mL TARC or 1 uM PAR1-AP (5 mins). Aggregation was measured using light transmittance aggregometry. Each graph is a representative trace from a separate donor

3.4.4 Effects of MDC and TARC alone on calcium mobilisation

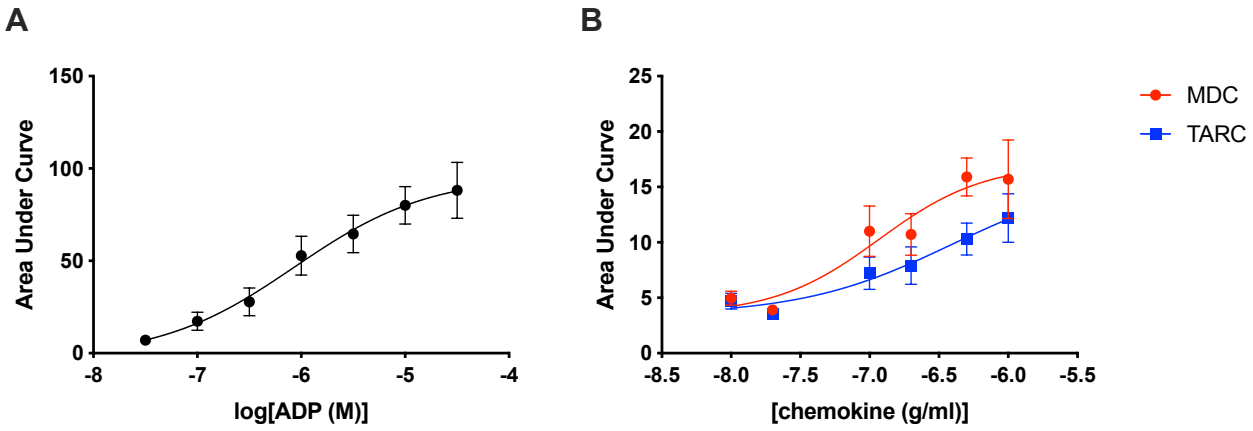

Figure 3.8: Washed platelets at $2x10^8$/mL loaded with Fura-2 were stimulated with increasing concentrations of A) ADP or B) 200ng/mL MDC or 200 ng/mL TARC. Area under curve represents the free calcium and is ratio of fluorescence at wavelengths 340:380 nm.

3.4.5 Priming effects of MDC and TARC on platelet aggregation in PRP

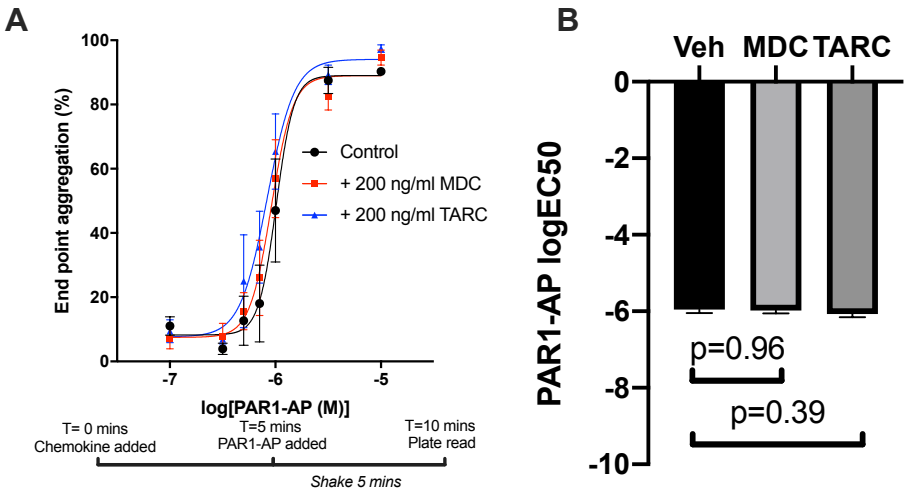

Figure 3.9: PRP was preincubated with MDC or TARC (5 mins) before application of PAR1-AP-and aggregation induced by a plate shaker (5 mins). A) A concentration response curve of the effect of vehicle, 200 ng/mL MDC or 200 ng/mL TARC on PAR1-AP induced aggregation. N=7. B) A bar chart of the logEC50s of the PAR1-AP concentration response curve in the presence of vehicle, MDC or TARC. Results analysed using a repeated measures one-way ANOVA with Dunnett's multiple comparisons. N=7.

3.4.6 Mechanisms of the effect of MDC on platelet aggregation in PRP

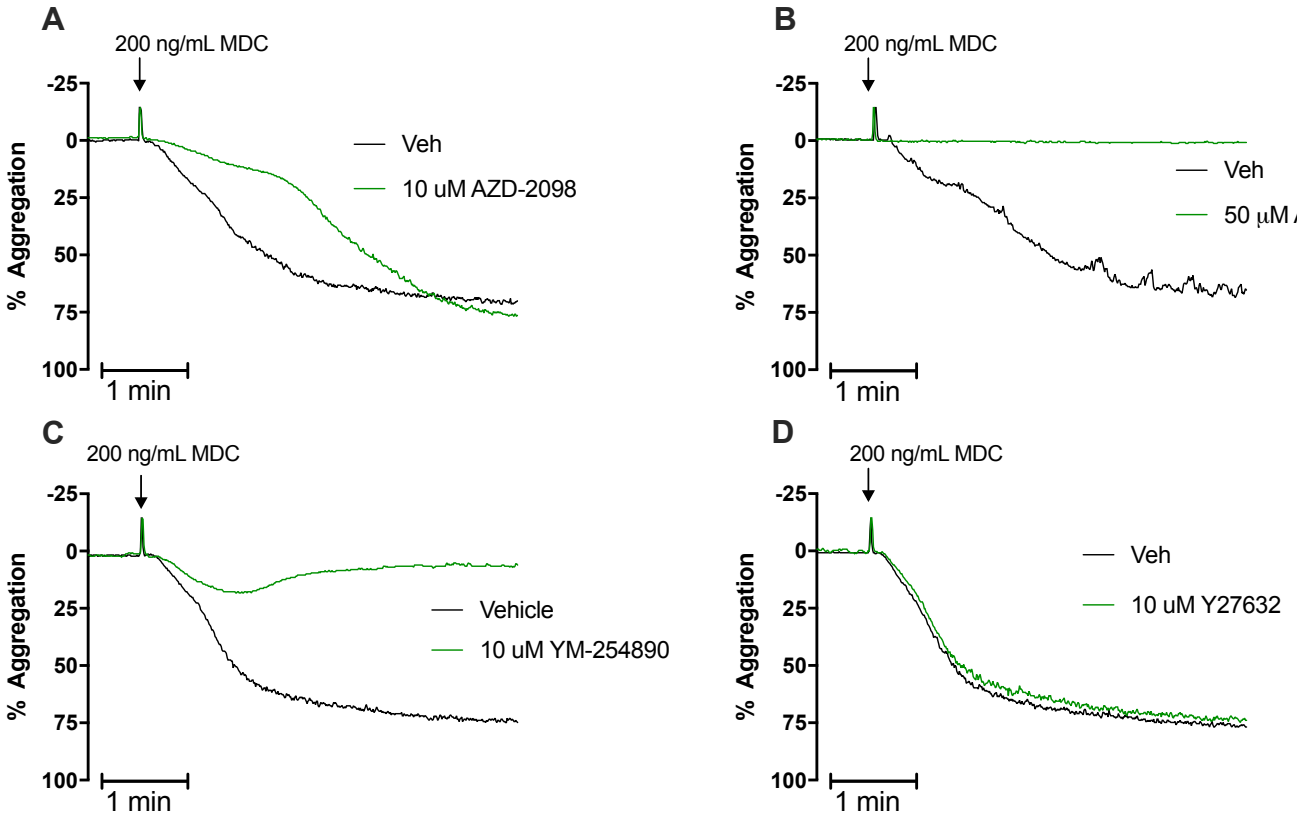

Figure 3.10: (ref:caption11)

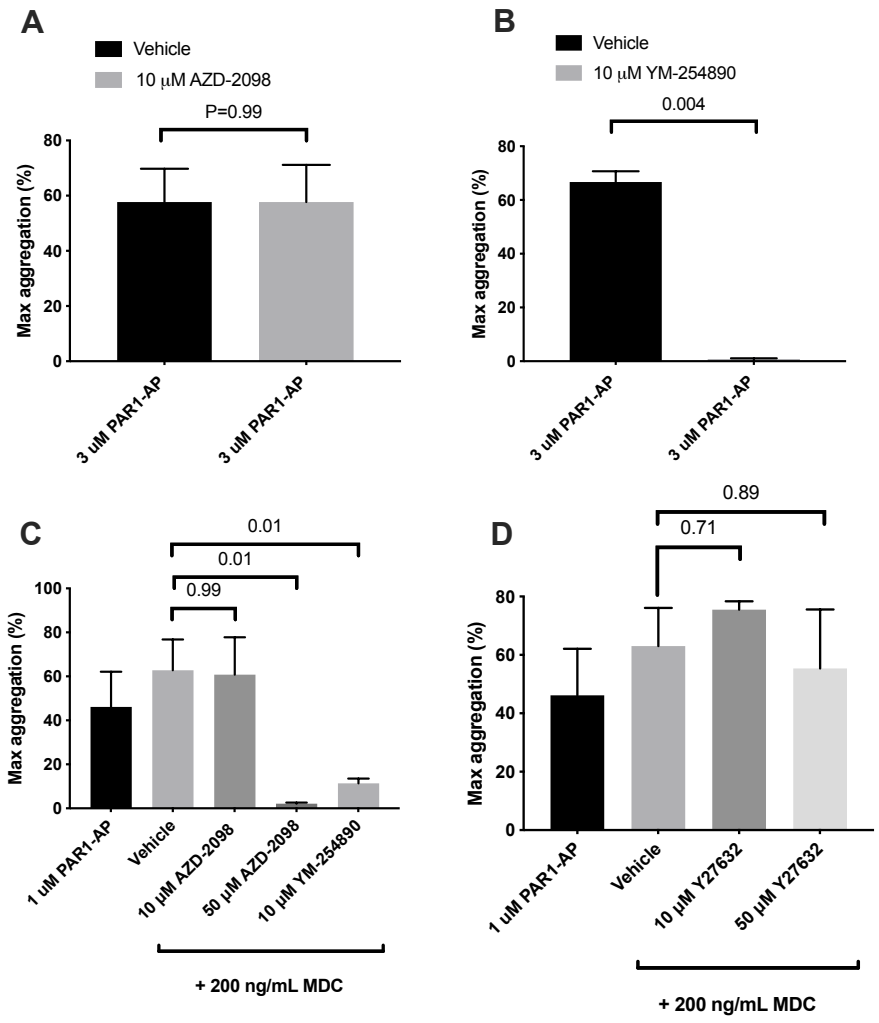

Figure 3.11: Results analysed with a repeated measures one-way ANOVA with Dunnett's Multiple comparisons.

3.5 Two sample Mendelian randomiation to assess the role of MDC (CCL22) and TARC (CCL17) in disease

3.5.1 MDC implicated in DVT

Table 3.2: Estimates for the effect of MDC on disease outcomes using two sample Mendelian randomization

Protein	Outcome trait	N SNP (rsID)	Method	Effect sie	Standard error	P value
MDC/CCL22	Non-cancer illness code self-reported: high cholesterol	1 (rs77542162)	Wald ratio	0.2224	0.0209	2.2e-26
MDC/CCL22	Non-cancer illness code self-reported: deep venous thrombosis (dvt)	1 (rs77542162)	Wald ratio	0.2936	0.0512	9.81e-09
MDC/CCL22	Non-cancer illness code self-reported: pulmonary embolism (with or without) dvt	1 (rs77542162)	Wald ratio	0.3308	0.0775	1.969e-05
MDC/CCL22	Forced expiratory volume in 1-second (FEV1)	1 (rs77542162)	Wald ratio	0.0312	0.0084	0.0001961
MDC/CCL22	Forced vital capacity (FVC)	1 (rs77542162)	Wald ratio	0.0285	0.0079	0.0003201
MDC/CCL22	Eye problems or disorders: Glaucoma	1 (rs77542162)	Wald ratio	0.2292	0.0654	0.000457
MDC/CCL22	Diagnoses - main ICD10: M16 Coxarthrosis [arthrosis of hip]	1 (rs77542162)	Wald ratio	0.2217	0.0648	0.0006186
MDC/CCL22	Non-cancer illness code self-reported: gout	1 (rs77542162)	Wald ratio	-0.4045	0.1237	0.001075
MDC/CCL22	Non-cancer illness code self-reported: hypothyroidism or myxoedema	1 (rs77542162)	Wald ratio	0.1203	0.0381	0.001607
MDC/CCL22	Schizophrenia	1 (rs77542162)	Wald ratio	-0.1477	0.048	0.002086
MDC/CCL22	Systolic blood pressure automated reading	1 (rs77542162)	Wald ratio	-0.0283	0.0099	0.004246
MDC/CCL22	Alcohol intake frequency	1 (rs77542162)	Wald ratio	-0.04	0.0143	0.005124
MDC/CCL22	Non-cancer illness code self-reported: osteoporosis	1 (rs77542162)	Wald ratio	0.1593	0.0665	0.01655
MDC/CCL22	Diagnoses - main ICD10: K80 Cholelithiasis	1 (rs77542162)	Wald ratio	-0.1965	0.0826	0.01731
MDC/CCL22	Fracture resulting from simple fall	1 (rs77542162)	Wald ratio	0.0528	0.0236	0.02527
MDC/CCL22	Cancer code self-reported: small intestine or small bowel cancer	1 (rs77542162)	Wald ratio	0.5966	0.2745	0.02976

3.5.2 TARC

Table 3.3: Estimates for the effect of TARC on disease outcomes using two sample Mendelian randomization

Protein	Outcome trait	N SNP (rsID)	Method	Effect sie	Standard error	P value
TARC/CCL17	Rheumatoid arthritis	2 (rs983545, rs222846)	Inverse variance weighted	0.1015	0.0386	0.00851
TARC/CCL17	Anorexia nervosa	2 (rs983545, rs222846)	Inverse variance weighted	0.2182	0.0845	0.009826
TARC/CCL17	Non-cancer illness code self-reported: gastro-oesophageal reflux (gord) or gastric reflux	2 (rs983545, rs222846)	Inverse variance weighted	0.003	0.0012	0.01172
TARC/CCL17	Systemic lupus erythematosus	2 (rs983545, rs222846)	Inverse variance weighted	0.2716	0.114	0.01717
TARC/CCL17	Ischemic stroke	2 (rs983545, rs222846)	Inverse variance weighted	-0.081	0.0373	0.02999
TARC/CCL17	Red blood cell count	2 (rs983545, rs222846)	Inverse variance weighted	-0.0116	0.0058	0.04445

3.6 Discussion

Chapter 4

Effects of adiposity on the human plasma proteome: Observational and Mendelian randomization estimates

The work within this chapter has been published in the International Journal of Obesity.[33] I am the first author of the publication; I performed all analyses, created all figures and tables and wrote the manuscript.

Supplementary Tables 1-7 can be found here `https://github.com/lucygoudswaard/mythesis/blob/master/index/figure/BMI_protein_INTERVAL/BMI_protein_supplementary_Tables.xlsx`

4.1 Introduction

Obesity has tripled worldwide since 1975, now affecting around 40% of adults in the United States and 26% of adults in the UK.[1] The average body mass index (BMI) of the UK adult population is now in the conventional 'overweight' category (BMI between 25 and 30 kg/m^2)[3] and 'overweight' is now more common than 'normal-weight' in middle age in many high-income countries.[4] BMI is often used as a proxy for adiposity given high correlations between BMI and more objectively measured fat mass indices.[20] Higher adiposity is a major risk factor for various noncommunicable diseases including type II diabetes, cardiovascular diseases, musculoskeletal diseases, and cancer,[7–9] which collectively put a strain on health services.[11] These BMI-disease associations are supported by prospective observational studies and, more recently, by Mendelian randomization (MR) studies,[13–15] which use genetic variation reliably associated with BMI to re-estimate effects of BMI on disease outcomes. Given the properties of genetic variation, this method helps to overcome issues such as confounding and reverse causation which commonly occur with observational studies.[83] Despite MR studies supporting a causal role of adiposity for cardiometabolic diseases, and randomised trials supporting the effectiveness of weight loss in reducing disease risk,[16] the molecular footprint of adiposity is not well understood. Previous studies have largely focused on the impact of higher BMI on the lipidome including traits such as cholesterol and triglycerides in lipoprotein subtypes (e.g. low-density lipoprotein (LDL) and high-density lipoprotein (HDL) particles)[20,21] and on inflammatory molecules such as C-reactive protein (CRP).[21,23] A benefit of studying the systematic effects of BMI on the circulating proteome is that proteins are often more suitable pharmacological targets than metabolites. Efforts to study the effect of BMI on the proteome have generally been in an observational framework.[24] It is estimated that 25% of

proteins in the human proteome circulate in blood,[25] which is important as the majority of druggable targets are such proteins.[22] Studying the effect of BMI on a large set of proteins has only recently become possible with newly developed proteomic technologies such as the SomaLogic platform, with the ability to quantify enzymes, protein kinases and transport proteins with unprecedented sensitivity.[84] Utilisation of SomaLogic within a trial or cohort setting has recently become more widespread, such as within the INTERVAL study, a UK cohort of blood donors.[82] There is evidence that proteins which change as a result of a higher BMI may contribute to cardiometabolic disease: identification of such proteins is important in understanding how higher BMI causes disease and to identify targets which may benefit from pharmacological intervention. In this study, we aimed to measure associations between adiposity and the human proteome and to also estimate the underlying effects in a causal framework. Using data on 2737 participants from INTERVAL, we estimated effects of BMI on 4034 (3622 unique) plasma protein traits, in both observational and MR frameworks. We examined the agreement between effect estimates from different methods and performed enrichment analyses of the most strongly altered proteins to map their potential relevance to disease.

4.2 Methods

4.2.1 Study population

INTERVAL is a prospective cohort study which was initially a randomised trial that aimed to test the efficiency and safety of reducing the time between whole blood donation in approximately 50000 participants.[48] Upon informed consent, eligible participants who were: aged 18 years and over, willing to complete online

questionnaires and without a self-reported history of major disease were recruited between June 11th, 2012 and June 15th, 2014 from 25 National Health Service Blood and Transplant (NHSBT) centres across England. Participants filled out questionnaires including self-reported height and weight, smoking frequency and alcohol consumption. Blood samples were taken at baseline which were analysed for full blood counts and blood biomarkers. This study was approved by Cambridge (East) Research Ethics Committee. Access to the data was granted by the Data Access Committee. The present study was conducted on a random subset of participants from INTERVAL who had basic phenotype data and plasma proteins measured by SomaLogic. This included up to 2737 participants mostly of European descent across analyses described below.

4.2.2 Assessment of BMI and covariables

Participants completed online questionnaires wherein they reported their height and weight. BMI was calculated as weight in kilograms divided by the square of their height in metres (kg/m^2). Available covariables were age, sex, previous or current smoking frequency (in three categories of: never, occasional, most days or every day) and alcohol intake frequency (in four categories of: rarely, less than once a week, 1-2 times a week, 3-5 times a week or most days). These covariables were chosen as they were measured in the INTERVAL collection and are measures which are thought to influence adiposity and cardiometabolic health.[20]

4.2.3 Measurement of circulating proteins

Plasma proteins were measured in INTERVAL participants at baseline (before randomisation of assignment to the time interval between blood donation) using the SomaScan® by SomaLogic.[82] This platform uses 4034 modified nucleotides

known as Slow Off-rate Modified Aptamers (SOMAmers) which make direct contact with proteins, enabling detection of 3622 unique proteins or protein complexes and quantifies them using a DNA microarray.[84] Separate SOMAmers can bind to isoforms of the same protein, but can also bind to the same protein at different sites (which can be impacted by post-translational modifications or complexes formed with other proteins). We therefore have included all 4034 SOMAmers. The extensive number of proteins measured, with no missingness and in a cohort of 2737 participants, provides a rich proteomic dataset. The proteins were measured in relative fluorescence units (RFUs) and quality control (QC) was performed as described by Sun et al..[82] There was no missingness across protein variables. The proteomic data used had been pre-adjusted (using linear regression) for age, sex, duration between blood draw and sample processing (1 day or less vs >1 day), and the first three genetic principal components, with the residuals inverse normal rank transformed. All following analyses use this 'pre-adjusted' data as input.

4.2.4 Genetic data and instrument for BMI

INTERVAL participant genotyping was performed on the Affymetrix GeneTitan Multi-Channel (MC) Instrument using the UK Biobank Axiom Array (ThermoFisher Scientific, Loughborough, UK) and the QC of genotype data was implemented as described by Astle et al..[49] The imputation panel used was the 1000 genomes phase-3-UK-10K.[49] A genetic instrument for BMI was constructed using 654 genetic variants that were associated with BMI at $P<5x10^{-8}$ in the inverse variance weighted fixed-effect meta-analysis of GWAS of 700000 individuals of European ancestry.[51] This meta-analysis consisted of around 250000 adults from the Genetic Investigation of ANthromopetric Traits (GIANT) consortium[85] and 450000 adults from the UK Biobank study. Only 0.05 % of UK Biobank

participants were included in the current INTERVAL study (of N=2737). These participants were not excluded to increase power. The weighted GRS was made using PLINK 2.0 software[52] using the effect alleles and beta coefficients from the source GWAS. The score was calculated by multiplying the number of effect alleles at each SNP by its effect estimate (beta), summing these, and dividing by the total number of SNPs included. The GRS therefore can be interpreted as the average per-SNP effect on BMI for each individual.

4.2.5 Statistical analyses

The population characteristics of INTERVAL participants with SomaLogic data who were included in this study (N range: 2422 to 2737 due to missing data for co-variables) were compared to those INTERVAL participants who were not included (N range: 27174 to 30721) to assess generalisability of any BMI-protein associations to the wider INTERVAL sample. Population characteristics evaluated were age, sex, weight, height, BMI, smoking frequency, and alcohol intake. Differences in population characteristics among the two INTERVAL sub-sets were tested by a two-sided t-test for continuous traits and a two-sided Chi-square test for categorical variables. Observational analyses were conducted using linear regression to examine associations between BMI (in normalized standard deviation (SD) units based on a rank normal transformation (rntransform() from "moosefun" package `https://github.com/hughesevoanth/moosefun`) and each standardised protein trait as a dependent (outcome) variable. Two linear models were used (using lm() function from R "stats" package): 1) adjusted for age and sex and 2) additionally adjusted for smoking and alcohol consumption (each as an ordered categorical variable). Given that the procedure which generates the "pre-adjusted" data (adjustment for covariables before rank normal transformation of proteins)

can reintroduce correlations,[86] age and sex are again used as covariables here. The estimates derived from models 1) and 2) therefore reflect the normalized SD-unit difference in each protein trait per normalized SD-unit (4.8 kg/m^2) higher BMI. Associations of covariables with BMI and protein traits were also examined using linear regression.

A Shapiro-Wilk test was used to confirm whether the GRS showed a normal distribution. MR analyses were conducted using two-stage least squares (2SLS) regression models with robust standard errors, using the systemfit function as part of the systemfit package,[54] with measured BMI in SD units and the GRS for BMI as the instrumental variable. These MR estimates reflect the normalized SD-unit difference in each protein trait per normalized SD-unit (4.8 kg/m^2) higher BMI. We report estimates from the direct linear associations between BMI and proteins as "observational" estimates and those from the 2SLS causal effect estimates as "MR estimates". Agreement between observational and MR estimates was examined using a separate linear regression. This was performed: 1) for all proteins, and 2) excluding the proteins that fell below our P-value reference point for strong evidence (defined below) to examine whether agreement is limited to "top hits" or applies throughout the effect distribution. Agreement between observational estimates and MR estimates would suggest that there are causal effects of BMI across the general proteome, with differences in estimates suggesting confounding of observational estimates.

To account for multiple testing, a Bonferroni correction was used to adjust results. This was informed by the correlation between proteins, adjusting only for the estimated number of independent traits (Figure 4.1). Correlation was assessed by a Spearman's correlation matrix. From a starting number of 4034, the number of independent proteins was 3655 (using a correlation cut-off of r = 0.8 or tree cut

height = 0.2 between proteins, Figure 4.2), dendrogram made using "iPVs" package `https://github.com/hughesevoanth/iPVs`). We utilised a Bonferroni adjusted P-value of 0.05/3655 = $1.4x10^{-5}$ to indicate strong evidence in this sample. Full results are presented in the supplementary material.

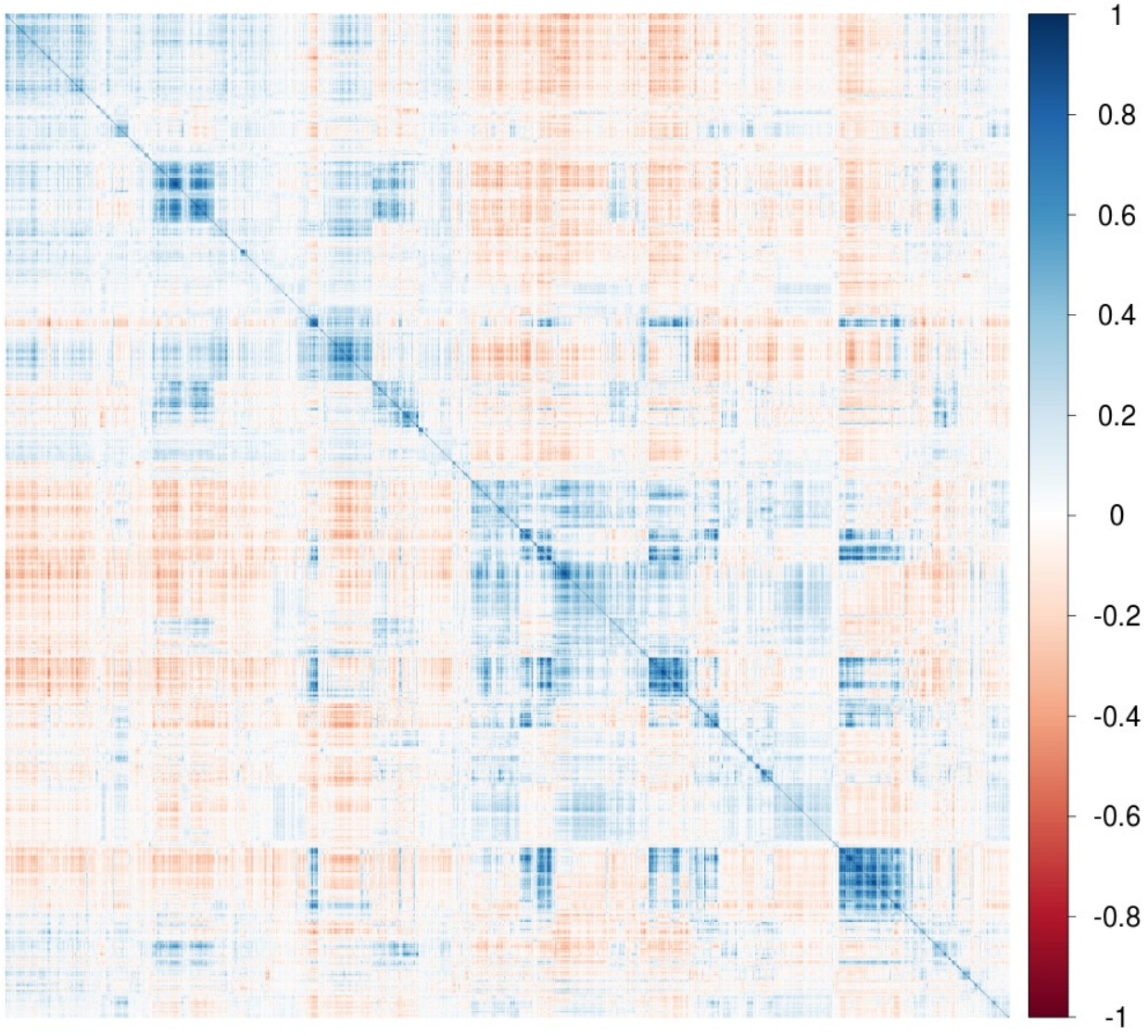

Figure 4.1: Correlation matrix of all protein traits analysed by SomaLogic (4,034) where the colour corresponds to the correlation coefficient ranging from 1 (dark blue) to -1 (dark red)

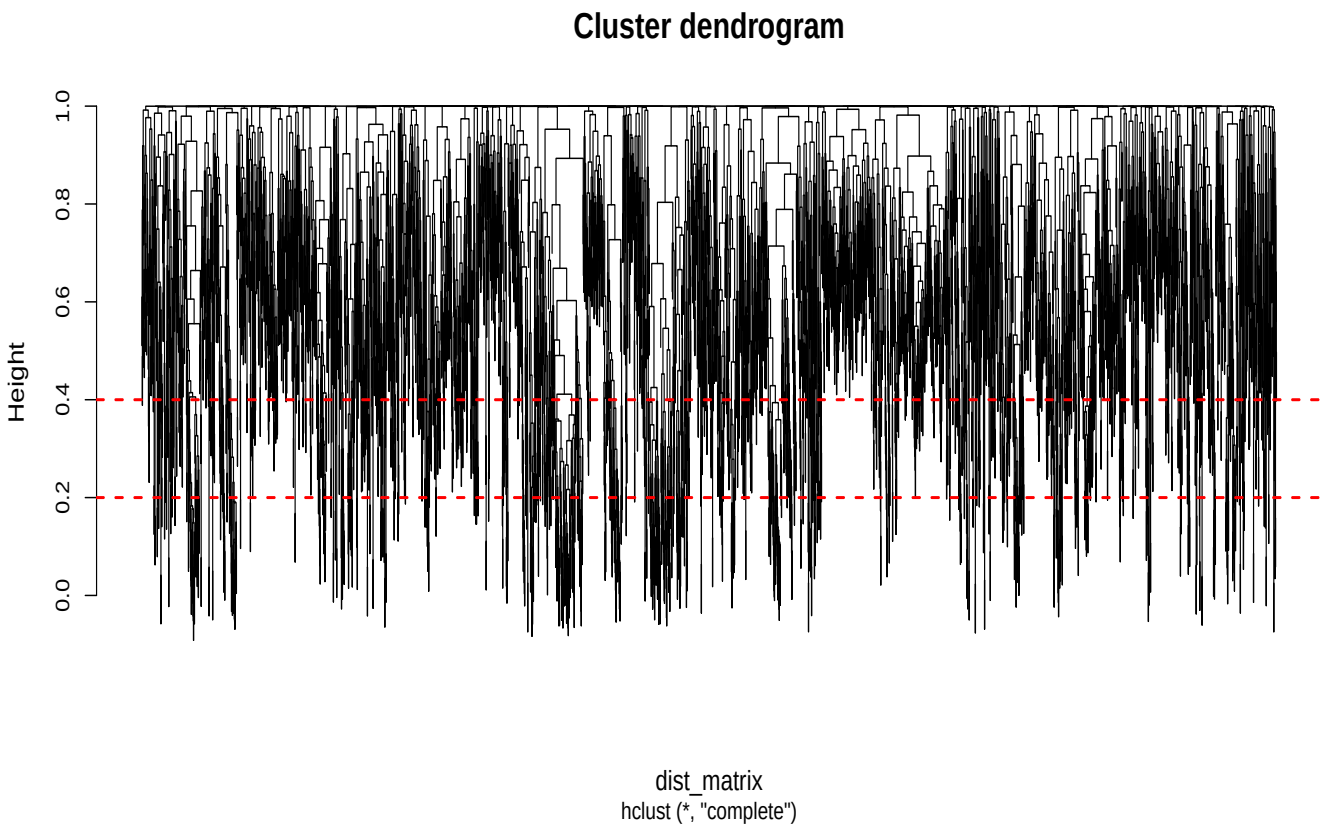

Figure 4.2: Cluster dendrogram showing the hierarchical relationship between protein levels. Height is calculated as (1-correlation coefficient). At a height of 0.2 and 0.4 (red dashed lines), the number of independent proteins was 3,655 and 3,016 respectively.

4.2.6 Enrichment analysis

To investigate whether any protein groups showed a particularly strong relationship with BMI and disease (signal detection), protein features were clustered for further analysis. First, a principal component analysis (PCA; prcomp() function from the R "stats" package) on the proteins, not the individuals, was performed on the "pre-adjusted" (see above) dataset (Figure 4.3A-D). The top "n" PC eigenvectors, as identified by a scree plot of the PCA eigenvalues (Figure 4.3E), were carried forward into an unsupervised k-means analysis (kmeans() function from the R "stats" package). Nineteen k-means analyses were run altering the value of k (number of clusters) from two to 20. To identify an appropriate number of protein clusters (k) we generated a scree like plot (Figure 4.3F). Here we plotted the variance explained by clusters, for each k, as estimated as the sum of squares explained by clusters (betweenness) over the total sums of squares, and looked for the smallest k with the maximum variance explained (a plateau). In summary, we used a data reduction method (PCA) to identify major axes (PCs) of the protein data that were then utilised in a machine learning clustering algorithm (k-means) to identify clusters of proteins that share abundance similarities across individuals.

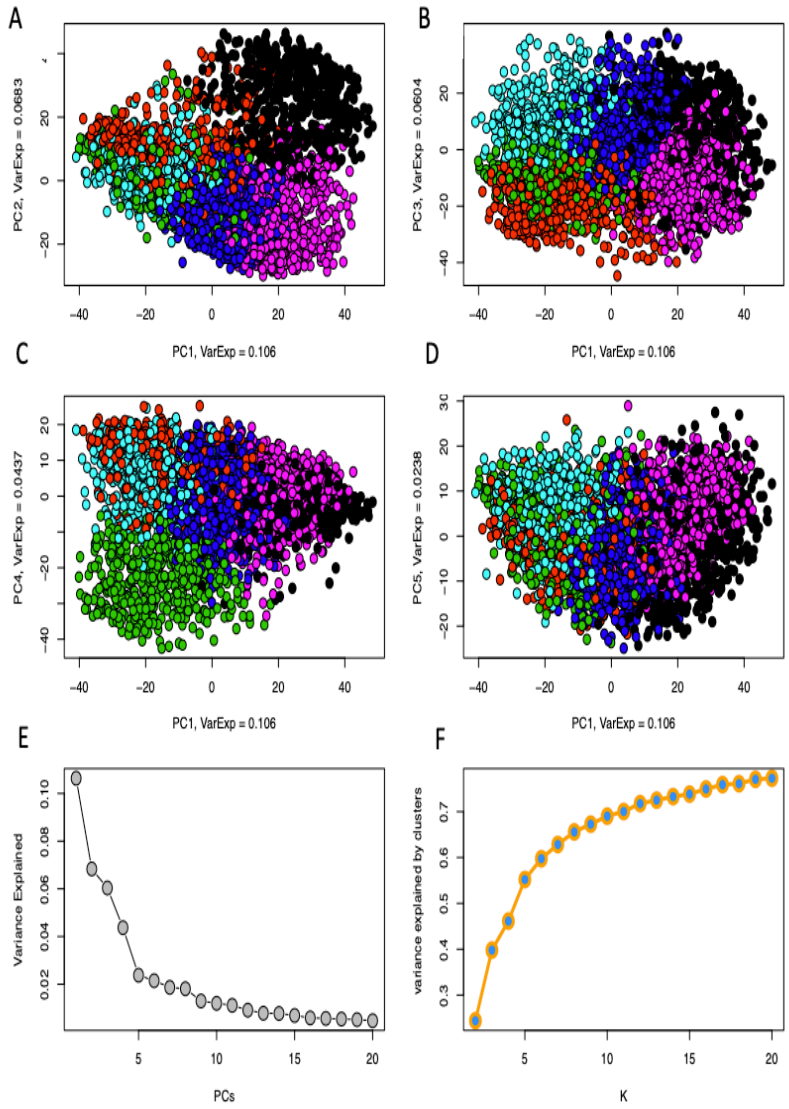

Figure 4.3: Principal component analysis (PCA) and k-means clustering of proteins provide evidence for five clusters. A-D) Principal component (PC) 1 vs PC2-PC5 for the study protein data. Each dot represents a protein, and the colors represent the clusters identified by the k-means analysis (1=black, 2=red, 3=green, 4=dark blue, 5=light blue, 6=pink). E) PC scree plot displaying the proportion of variance explained by each of the first 20 PCs. F) K-means scree like plot displaying the variance explained by clusters for each of the 19 k-means analysis, where in each analysis k (the number of clusters) was set from two to 20.

To explore whether there was a systematic difference in the association of proteins within these clusters and BMI, the beta coefficients from the observational linear regressions or MR models were transformed into their absolute values and divided by their standard error (SE). The absolute betas divided by their SEs in each cluster was compared using a one-tailed pairwise Wilcox test to identify which clusters showed a stronger association with BMI. For the cluster(s) showing evidence for larger absolute beta coefficients, an enrichment analysis was performed using DAVID bioinformatics resources 6.8.[87] Enrichment was assessed by using the uniprot IDs for the proteins in the cluster and comparing these proteins with the uniprot IDs of the full SomaLogic protein list. Enrichment for protein involvement in disease (using the genetic association database (GAD) disease classes)[88] of the protein cluster was assessed by fold enrichment and a Bonferroni-corrected P-value to account for multiple testing. Proteins that were associated with BMI in confounder-adjusted observational analyses at $P<1.4x10^{-5}$ were also entered into the disease enrichment tool and compared with the total proteins (as described for cluster enrichment). Analyses were performed using R version 3.4.2.[53] R code used for analyses available upon request.

4.3 Results

4.3.1 Participant characteristics

INTERVAL participants included in this study (those with proteomic data), had a mean age of 45.0 years (SD of 14.1 years) and 48.3% were female (Table **??**). Mean BMI was 25.9 kg/m^2 (SD of 4.8 kg/m^2) and the majority of participants were non-smokers (59.1%). Nearly a quarter (23.5%) reported currently or previously smoking daily and 71.5% reported drinking alcohol at least once a week. Partici-

pants with proteomic data were representative of the full INTERVAL cohort (Table **??**).

Table 4.1: Comparison of included vs excluded INTERVAL participants

Variable	Included Mean (SD) or %	Included N	Excluded Mean (SD) or %	Excluded N	P value for difference (Two tailed t-test or Chi2 test)
Age	45.0 (14.1)	2737	44.9 (14.2)	30721	0.76
Sex		2737		30721	0.04
Male	51.70%		49.70%		
Female	48.30%		50.30%		
Weight (kg)	78.5 (16.1)	2736	78.6 (16.0)	30719	0.26
Height (cm)	172.7 (9.7)	2730	172.2 (9.6)	30661	0.005
Body mass index (kg/m2)	25.9 (4.8)	2729	26.4 (4.7)	30659	0.66
Smoking frequency		2675		29958	0.88
Never	59.10%		59.40%		
Occasional	11.90%		12.00%		
Most days or every day	29.00%		28.50%		
Alcohol intake frequency		2422		27124	0.34
Rarely	11.70%		12.60%		
Less than once a week	16.90%		17.10%		
One or two times a week	37.40%		37.70%		
Three to five times a week / most days	34.10%		32.50%		

4.3.2 Observational estimates of associations of BMI with protein traits

In a linear regression model adjusted for age and sex among 2729 adults, BMI (per SD higher) was associated with 1576 proteins (39%) at the level $P < 1.4 \times 10^{-5}$ (multiple testing reference point, Supplementary Table 1). In a second model additionally adjusting for frequencies of smoking and alcohol intake among 2 380 adults, there were 1447 associations at the same reference point (Supplementary Table 2). The strongest positive associations were with leptin (0.74 SD, 95% CI 0.71-0.77, $P=9.9 \times 10^{-324}$) and adipocyte fatty acid binding protein (FABP4) (0.58 SD, 95% CI 0.55-0.62, $P=6.4 \times 10^{-211}$). BMI (per SD) was also strongly positively associated with inflammatory proteins such as Complement Factor I (0.46 SD, 95% CI 0.43-0.50, $P=5.6 \times 10^{-122}$) and CRP (0.44 SD, 95% CI 0.41-0.48, $P=8.2 \times 10^{-112}$). BMI (per SD) also showed strong negative associations with proteins such as insulin-like growth factor-binding protein (IGFBP) 2 (-0.48 SD, 95% CI -0.51 to -0.44, $P=2.7 \times 10^{-133}$) and sex hormone-binding globulin (SHBG) (-0.43 SD, 95% CI -0.47 to -0.39, $P=2.4 \times 10^{-106}$).

4.3.3 Observational associations of covariables with BMI and protein traits

Age, sex, and frequencies of smoking and alcohol intake were each associated with BMI (Table **??**). Males had a higher BMI than females (0.17 SD, 95% CI 0.10-0.25, $P=5.8 \times 10^{-6}$). Age was positively associated with BMI (0.01 SD higher per year older, 95% CI 0.009-0.015, $P=1.2 \times 10^{-18}$). Smoking frequency was positively associated with BMI, but alcohol intake frequency was negatively associated with BMI. Covariables (age, sex, smoking and alcohol) showed associations with protein traits (Supplementary Tables 3-6 and Figures 4.4 A-D). There was evidence

for 18 associations between age and proteins, 26 associations between sex and proteins, 38 proteins associated with smoking and 137 proteins associated with alcohol at the Bonferroni-adjusted level of $p<1.4x10^{-5}$.

Table 4.2: Associations between covariables (exposure) and standardised BMI (outcome)

Variable	N	Beta coefficient per 1-unit increase in confounder (SDs)	Standard error	P value	Adjusted R2	F statistic	95% CI	Lower 95% CI	Upper 95% CI
Age (years)	2729	0.012	0.001	1.23e-18	0.028	78.8	0.003	0.009	0.015
Sex (1=female, 2=male)	2729	0.17	0.038	5.8e-06	0.007	20.6	0.075	0.1	0.25
Smoking frequency (1=never, 2=occasinal, 3=most days/ every day)	2667	0.022	0.022	3.16e-05	0.006	17.4	0.043	0.05	0.13
Alcohol intake frequency (1=rarely, 2=<1 a week, 3=1-2 times a week, 4=most days/every day)	2415	0.021	0.021	7.08e-05	0.006	15.8	0.041	-0.12	-0.04

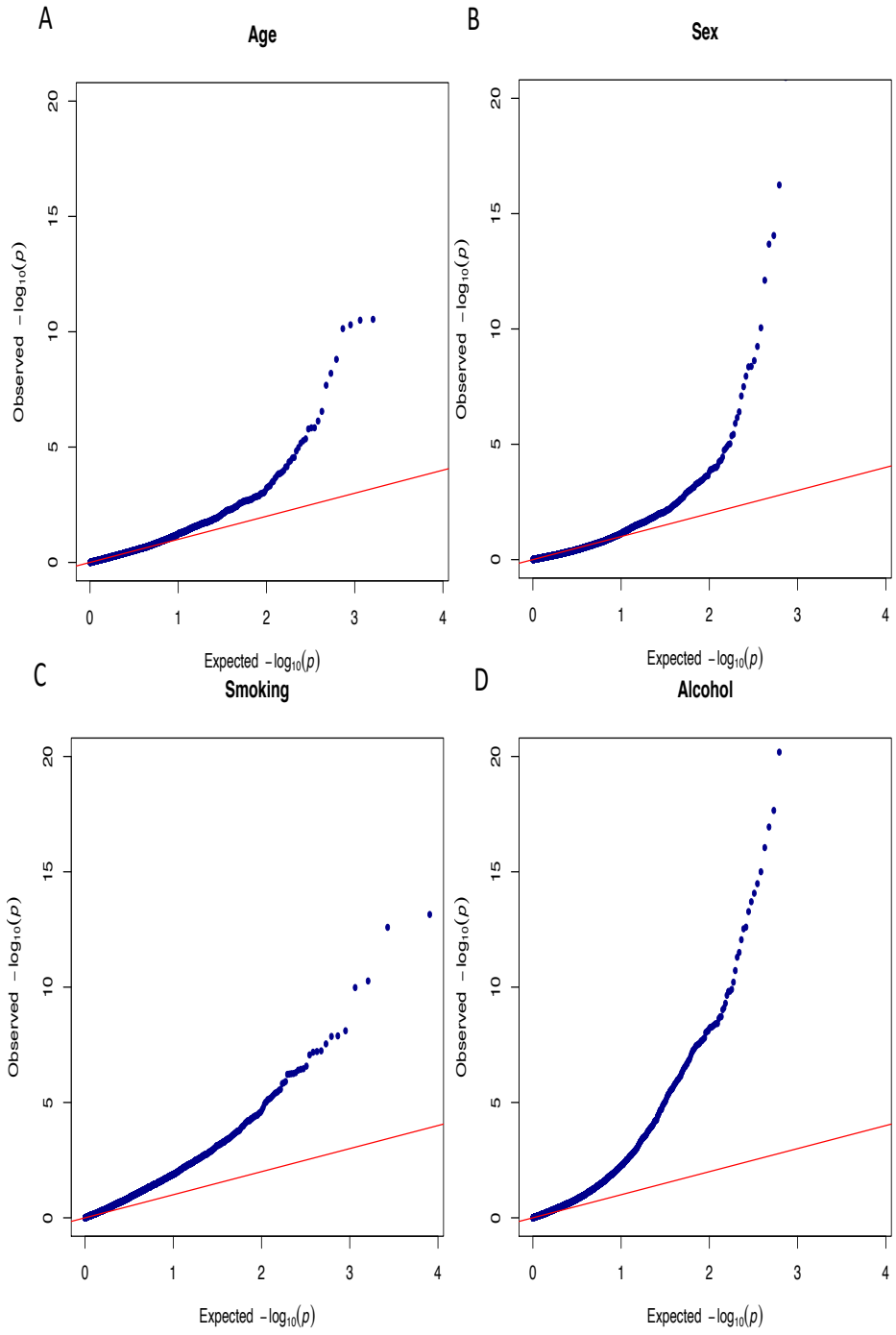

Figure 4.4: Q-Q plots for the expected against observed -log10(pvalues) of the association of age (A), sex (B), smoking (C) and alcohol (D) with protein traits

4.3.4 Associations of the GRS for BMI with measured BMI and covariables

The distribution of the GRS among participants was normal (mean=0.08, SD=0.29, W=0.99, P=0.73, N=2729). The GRS was associated with BMI, explaining 2.8% of its variance (R^2=0.028, P=1.6x10-18, Table **??**). There was no strong evidence of association between GRS and age (R^2=0.001, P=0.11), sex (R^2=6x10-5, P=0.28), smoking frequency (R^2=< 0.0001, P=0.91), or alcohol intake (R^2<0.0001, P=0.44).

Table 4.3: Associations of the genetic risk score for BMI with reported BMI and covariables

Variable	N	Beta coefficient (per 1-unit increase in GRS)	Standard error	P value	Adjusted R2	F statistic
BMI (SDs)	2729	0.57	0.06	1.64e-18	0.028	78.2
BMI (kg/m2)	2729	2.54	0.31	4.82e-16	0.024	66.68
Age	2737	-1.47	0.92	0.11	0.001	2.55
Sex (1=female, 2=male)	2737	-0.04	0.03	0.28	5.96e-05	1.16
Smoking frequency (1=never, 2=occasinal, 3=most days/ every day)	2675	-0.02	0.14	0.91	-4e-04	0.01
Alcohol intake frequency (1=rarely, 2=<1 a week, 3=1-2 times a week, 4=most days/every day)	2422	-0.06	0.07	0.44	-2e-04	0.6

4.3.5 MR estimates of associations between BMI and protein traits

In MR analyses, eight unique BMI-protein associations were detected at the level $P<1.4x10^{-5}$ (multiple testing reference point, Figure **??**). MR estimates provide an estimate of the causal association between protein (in SDs) per SD higher BMI. The strongest association of BMI was again with leptin (0.63 SD, 95% CI = 0.48-0.79; $P=1.6x10^{-15}$); this was followed by the association with FABP4 (0.65 SD, 95% CI = 0.46-0.83; $P=6.7x10^{-12}$). A strong negative association was also seen between BMI (per SD) and SHBG (-0.45 SD, 95% CI -0.65 to -0.25, $P=1.4x10^{-5}$). Other BMI-protein associations ($P<1.4x10^{-5}$) included positive associations with fumarylacetoacetase (FAAA), inhibin beta C chain and complement C5, and negative associations with receptor-type tyrosine-protein phosphatase delta and PILR alpha-associated neural protein. Supplementary Table 7 provides the full MR results.

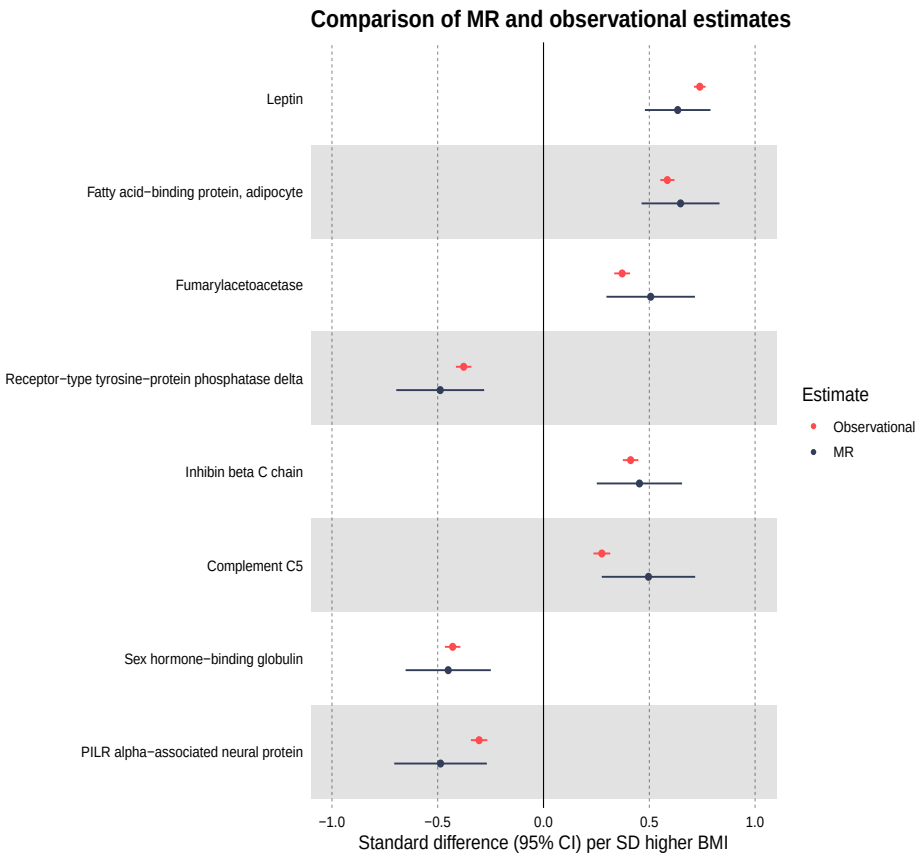

Figure 4.5: Forest plot of the strongest BMI-protein Mendelian randomisation estimates and the corresponding observational estimates

4.3.6 Comparison of observational and MR estimates

The distribution of P-values for associations between BMI and protein traits suggested an overrepresentation of signal for the observational estimates of BMI and protein traits; far more than expected from chance alone (Figure 4.6A). In contrast to this, the extent of this overrepresentation was reduced considerably in the MR (Figure 4.6B).

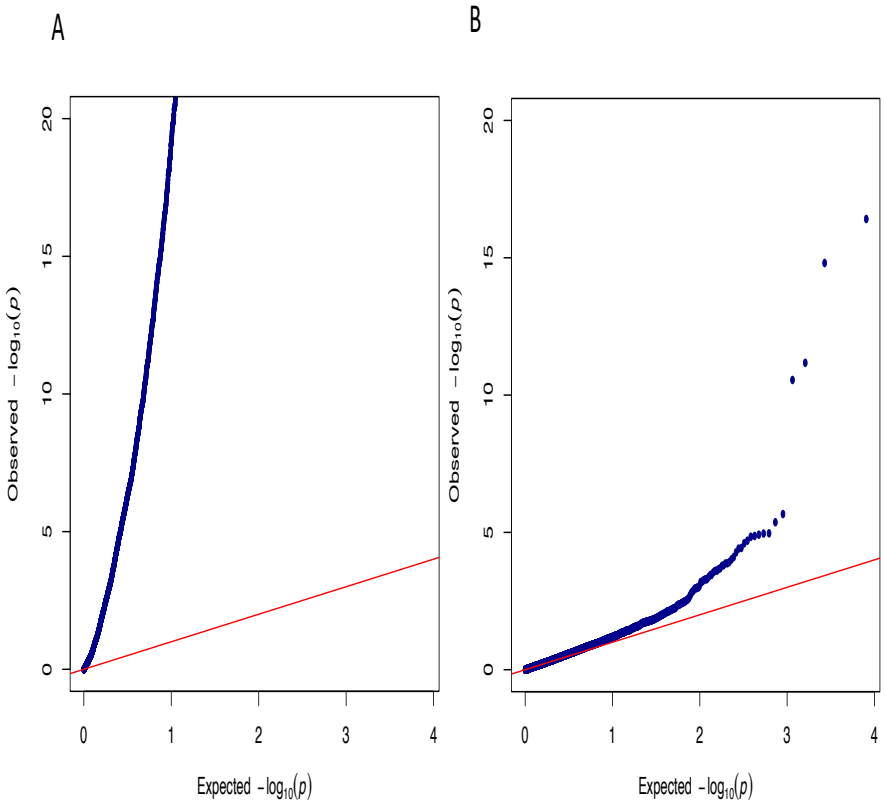

Figure 4.6: A) Q-Q plot of expected against observed -log10(pvalues) for the unadjusted observational BMI-protein trait estimates B) Q-Q plot of expected against observed -log10(p) values for the Mendelian Randomization BMI-protein trait estimates.

The unadjusted and confounder-adjusted regression coefficients for BMI and protein traits were strongly associated (Beta=0.99 SDs, R^2=0.99, P=9.9x10^{-324}, Figure 4.7). Compared with the observational estimates, the MR estimates were less precise, but there was a strong positive association between the beta coefficients from observational and MR estimates (Beta=0.68 SDs, R^2=0.33, P=9.9x10^{-324}) (Figure 4.8). After removing the proteins where P<1.4x10^{-5}, the strength of association be-

tween unadjusted and adjusted observational estimates remained, but the association between observational and MR estimates attenuated slightly (Figures 4.9 A/B). These results suggest causal effects of BMI across the general proteome.

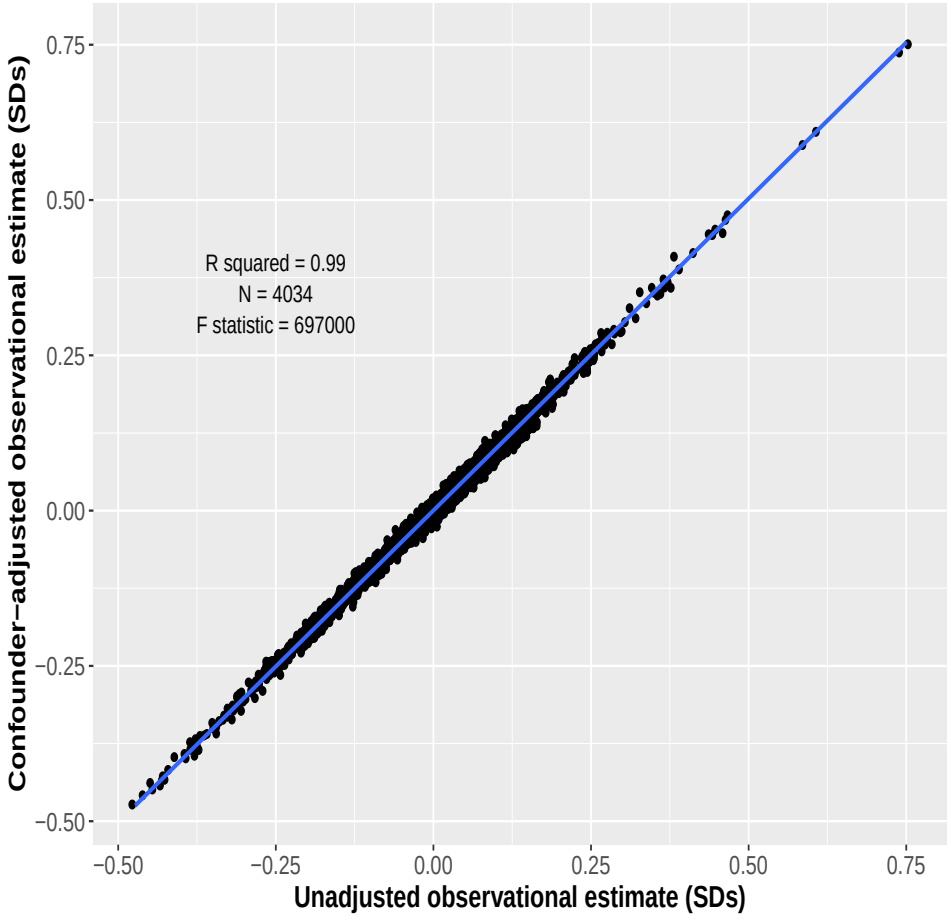

Figure 4.7: Scatter plot of the unadjusted (age and sex adjusted) observational estimates and the confounder-adjusted observational estimates for BMI and protein traits with a regression line.

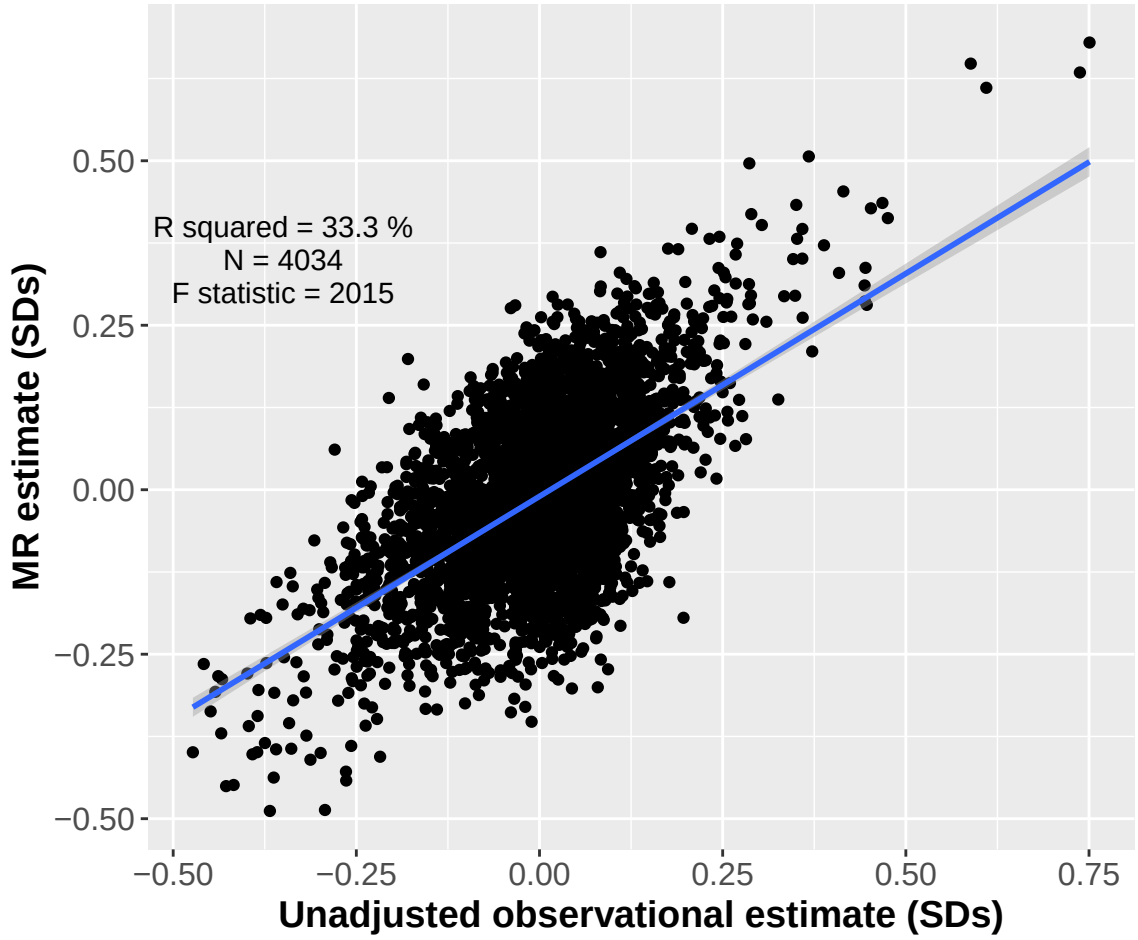

Figure 4.8: Scatter plot of the unadjusted (age and sex adjusted) observational estimates and the MR estimates for BMI and protein traits with a regression line

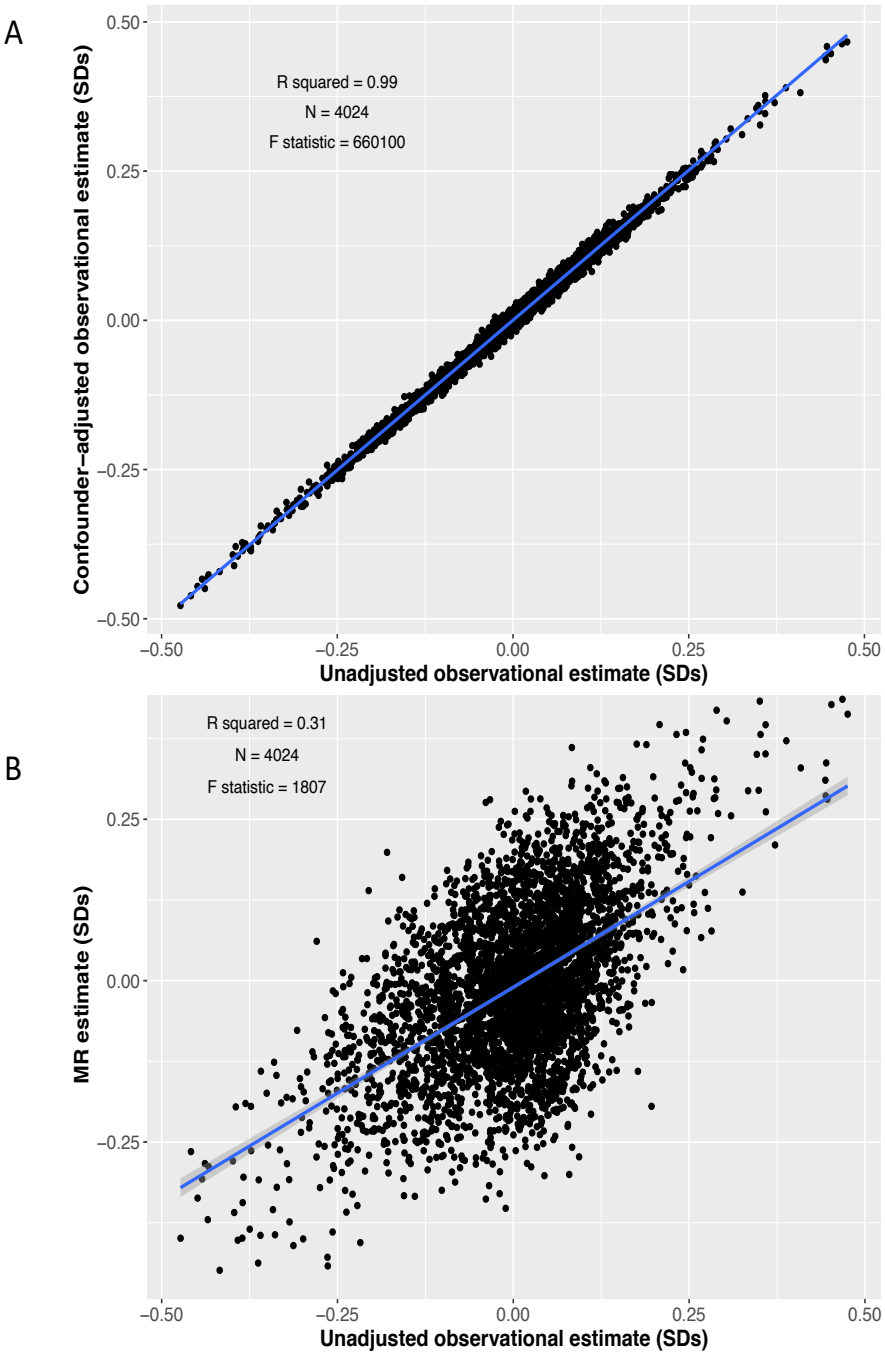

Figure 4.9: A) Scatter plot of the unadjusted (age and sex adjusted) observational estimates and the confounder-adjusted observational estimates for BMI and protein traits with a regression line (blue), with the top eight MR BMI-associated proteins excluded. 6B) Scatter plot of the unadjusted (age and sex adjusted) observational estimates and the MR estimates for BMI and protein traits with a regression line (blue), with the top eight MR BMI-associated proteins excluded.

4.3.7 Enrichment analysis of strongest BMI-protein associations

In examining the clustering of proteins, visual representation using a scree plot suggested there were five PCs that explained 30.3% of the variance (Figure 4.3). After PC5 there was clear drop in variance explained, therefore other PCs were excluded. These five PCs were entered into a k-means analysis, which provided evidence for five clusters (grouping of individual proteins is included in Supplementary Table 7). To identify which cluster was most strongly affected by BMI, the median absolute beta coefficient divided by the SE for each cluster was compared with the overall estimate. Six of the proteins out of the eight strongest BMI-protein MR estimates were in cluster 2 (Supplementary Table 7). There was consistent evidence that cluster 2 showed a stronger association with BMI than the overall average BMI-protein effect both observationally (3.79 (IQR 1.62-7.06) vs 3.35 (IQR 1.57-5.83) respectively, $P=3.7x10^{-4}$) and in MR (0.85 (IQR 0.41-1.46) vs 0.74 (IQR 0.32-1.14), $P=5.3x10^{-6}$, Table **??**). Cluster 2 showed consistent evidence of a having the largest BMI effect. Compared with the full protein list in SomaLogic, the proteins in cluster 2 were enriched for disease (Table **??**), including cardiovascular disease (1.14 fold enrichment, $P=1.3x10^{-4}$), renal disease (1.22 fold enrichment, $P=1.0x10^{-3}$) cancer (1.1 fold enrichment, $P=9.5x10^{-3}$) and metabolic disease (1.08 fold enrichment, $P=4.2x10^{-2}$). No other individual cluster showed enrichment for disease. Enrichment for disease was also explored by comparing the proteins which had an association with BMI ($P<1.4x10^{-5}$) in the confounder-adjusted regression model with the total protein list. Compared with the full protein list, the proteins which showed a stronger observational association with BMI were enriched for renal disease (1.21 fold enrichment, P=0.001) and metabolic disease (1.9 fold enrichment, P=0.015, Table **??**).

Table 4.4: Comparison of cluster median(absolute beta coefficient / SE) with the overall median for observational and MR analyses, *One-tailed pairwise Wilcox test

Cluster	N	Median absolute observational beta coefficient divided by SE (IQR)	P value* for cluster v all proteins (Observational)	Median MR beta coefficient divided by SE (IQR)	P value* for cluster v all proteins (MR)
1	1061	3.28 (1.56-5.26)	1	0.69 (0.31-1.22)	1
2	1406	3.79 (1.62-7.06)	0.00037	0.85 (0.41-1.46)	5.3e-06
3	581	2.77 (1.39-4.30)	1	0.74 (0.37-1.25)	1
4	502	3.01 (1.25-5.6)	1	0.68 (0.34-1.19)	1
5	484	4.44 (2.2-7.05)	1.8e-05	0.63 (0.32-1.14)	1
All	4034	3.35 (1.57-5.83)		0.74 (0.35-1.28)	

Table 4.5: Cluster 2 vs full SomaLogic protein enrichment results for disease class using DAVID bioinformatics 6.8

Term (Genetic Association Database disease class)	Count	%	List Total	Population Hits	Population Total	Fold Enrich-ment	P Value	Bonferroni-adjusted P value
Cardiovascular	439	34.1	1024	1023	2723	1.14	7.3e-06	0.00013
Renal	216	16.8	1024	472	2723	1.22	5.8e-05	0.001
Cancer	401	31.2	1024	958	2723	1.11	0.00053	0.0095
Pharmacogenomic	357	27.8	1024	856	2723	1.11	0.0019	0.034
Metabolic	504	39.2	1024	1243	2723	1.08	0.0024	0.042
Vision	100	7.8	1024	216	2723	1.23	0.0059	0.1
Haematological	168	13.1	1024	388	2723	1.15	0.0098	0.16
Reproduction	152	11.8	1024	351	2723	1.15	0.014	0.23
Immune	360	28	1024	886	2723	1.08	0.015	0.24
Neurological	311	24.2	1024	759	2723	1.09	0.016	0.25
Aging	117	9.1	1024	278	2723	1.12	0.075	0.75

Table 4.6: Disease enrichment of proteins associated with BMI in confounder adjusted model compared with all proteins in SomaLogic (DAVID Bioinformatics 6.8)

Term (Genetic Association Database disease class)	Count	%	List Total	Population Hits	Population Total	Fold Enrichment	P Value	Bonferroni-adjusted P value
Renal	218	16.4	1039	472	2723	1.21	7.7e-05	0.001
Pharmacogenomic	372	28	1039	856	2723	1.14	8.8e-05	0.002
Metabolic	515	38.8	1039	1243	2723	1.09	0.00083	0.015
Aging	129	9.7	1039	278	2723	1.22	0.0027	0.047
Cardiovascular	424	32	1039	1023	2723	1.09	0.0041	0.072
Unknown	233	17.6	1039	541	2723	1.13	0.0064	0.109
Psych	228	17.2	1039	534	2723	1.12	0.012	0.189
Reproduction	154	11.6	1039	351	2723	1.15	0.014	0.228
Normal variation	86	6.5	1039	185	2723	1.22	0.015	0.236
Other	211	15.9	1039	497	2723	1.11	0.021	0.313
Cancer	391	29.5	1039	958	2723	1.07	0.022	0.332
Haematological	164	12.4	1039	388	2723	1.11	0.05	0.605
Neurological	309	23.3	1039	759	2723	1.07	0.055	0.638
Developmental	160	12.1	1039	381	2723	1.1	0.066	0.708

4.4 Discussion

This study sought to estimate the effects of adiposity on a comprehensive set of protein traits only recently measurable by untargeted proteomics using observational and MR methods. Observational results provided evidence for associations between BMI and 1576 proteins and MR was performed to reduce confounding. MR results suggest that BMI alters protein traits involved in regulating appetite, sex hormones, inflammation, and other systems; specific proteins most altered by BMI include leptin, FABP4, and SHBG. Results of follow-up analyses suggest that the cluster of proteins most altered by BMI is enriched for genes associated with cardiovascular and metabolic disease.

This study explored the effect of BMI on a large set of circulating proteins in an MR framework. Previous studies have used observational epidemiology to explore the effect of obesity on the plasma proteome: one study used mass spectrometry and found an increase in Complement Factors I, B and H and an increase in CRP.[24] These findings were replicated in our current observational analysis using the SomaLogic platform, indicating that associations are detectable across different proteomic platforms. The only association that did not replicate in the current study was the positive association with protein S100-A9. Although the MR analysis did not support some of these BMI-protein associations as being causal based on a P-value reference point, the strong association between the observational and MR estimates throughout the entire effect distribution suggests that disagreements between methods are likely an issue of power given current sample sizes.

Previous work implementing MR to examine the relationship between BMI and ~1000 proteins (measured using the same SomaLogic array) provided corroborative evidence to that shown here.[89] Both studies suggested a positive association

between BMI and leptin, as well as a negative association with SHBG. Other proteins, such as IGFBP1/2 and growth hormone receptor, did not pass our multiple testing threshold, but the direction and magnitude of estimates were in agreement, suggesting a possible causal effect that was not detectable in the current study. Building on previous work, the current study provides MR estimates for >3600 proteins, offering a wider proteomic profile and detecting additional associations such as that between BMI and fumarylacetoacetase and inhibin beta C chain. Furthermore, the inclusion of over threefold more proteins allowed a more comprehensive enrichment analysis to be performed.

For proteins with stronger MR derived association evidence, it is important to explore whether they have a potential role in disease. Identification of individual proteins could help to guide future intervention if changes in proteins can be mapped to disease outcomes. Our results suggest a strong positive effect of BMI on levels of leptin, a hormone released by white adipose tissue which suppresses appetite.[90] The direction of effect agrees with estimates from previous cross-sectional and MR studies,[21,91] indicating leptin receptor resistance.[92] There is observational evidence in humans that higher leptin can induce greater aggregation of platelets (cells involved in haemostasis).[93] In a larger observational study, leptin was found to be associated with higher risk of coronary events independent of BMI.[94]

Our results help to provide contextualisation for proteins which have already been implicated in disease. For example, results suggest a strong positive effect of BMI on FABP4, an adipokine found primarily in adipocytes and macrophages.[95] This MR estimate supports the association which has been suggested in previous observational studies.[96] FABP4 has been implicated in cardiometabolic disease: a SNP which increases FABP4 was found to raise the odds of type II diabetes among adults,[97] potentially through its contribution to higher insulin resistance.[98] FABP4

has also been associated with higher risk of atherosclerosis among adults.[99] A strong SHBG-lowering effect of higher BMI was also suggested here. The SHBG molecule is a glycoprotein which binds androgens and oestrogens and suppresses their activity;[100] a reduction in SHBG is therefore expected to lead to higher levels of circulating sex hormones. The negative effect of BMI on SHBG seen here supports observational findings.[101–103] When evaluating the role of SHBG in disease, MR analysis suggests that an increase in SHBG contributes to a decrease in risk of cardioembolic stroke.[81] Other studies have also implicated lower SHBG levels in increasing type II diabetes risk.[104] The exact mechanisms leading from decreased SBHG to ill-health is unclear, but may arise as a result of the increased bioavailability of testosterone and oestrogen.[100]

Despite these possible protein involvements in cardiometabolic disease, it remains difficult to assess the contribution of individual proteins as they are not entirely independent and any pathological effects would likely be due to a global change in protein composition. There are not distinct groupings in the SomaLogic data as there often are with, for example, metabolomics data. We therefore examined proteins grouped into clusters of similar features, compared BMI-protein estimates of each cluster with overall estimates and explored enrichment for genes related to disease. The cluster most altered by BMI (cluster 2) included most of the eight proteins with the strongest BMI effects from MR analyses, as well as various complement factors, chemokines and coagulation factors, and was found to be enriched for genes related to cardiovascular disease, renal and metabolic diseases, and cancer. Enrichment was similar when comparing the proteins that had an observational association with BMI with all proteins included, with enrichment appearing greatest for renal and metabolic disease. Together, this suggests that changes in proteins may mediate effects of obesity on cardiometabolic diseases; more focused investi-

gations of these proteins are now needed.

This study has some limitations. Firstly, although INTERVAL is one of the largest existing cohorts to have untargeted proteomic data based on the SomaLogic platform, the sample size is still relatively modest and may have low power to detect some associations when using MR: based on the detectable (P<1.4x10-5) median absolute observational effect size (0.13 SDs), our analyses had 80% power to detect MR effect sizes >= 0.33 SDs (alpha=0.05) for our sample size (N=2737).[105] With greater statistical power, there would likely be more proteins detected with MR. This was reinforced by the high agreement in the magnitude of effect estimates seen in observational and MR analyses which applied throughout the effect distribution. Secondly, height and weight were self-reported which could bias results towards the null due to systematic error in BMI measurement. However, strong correlations are often reported between self-reported and measured BMI[61] and the validity of self-reported BMI is supported by the association between the GRS for BMI and self-reported BMI in INTERVAL to the degree expected. Thirdly, the small degree of overlap between INTERVAL and UK Biobank (participants used for the source GWAS for BMI who were also in INTERVAL) may have a biasing effect on estimates, though this is likely to be towards the null. We anticipate that overall, this bias would make estimates more conservative.[106] Fourth, we recognise a lack of availability of possible confounders such as socio-economic position (which likely affect both BMI and protein traits related to cardiovascular disease processes.[107,108] Residual confounding may help account for the divergence between observed and expected P-values seen in observational versus MR models. Fifth, the proteins examined are highly correlated and we therefore may not fully be describing changes in individual proteins. Evidence from case-control cohorts as well as functional and animal studies would help isolate individual proteins that

are altered and contribute to disease. Finally, although analyses provide insight into the proteomic effects of BMI, it does not distinguish between the type of adiposity. It would be useful to distinguish between the effects of subcutaneous and visceral fat using dual-energy X-ray absorptiometry (DXA) derived measurements, but these were not available in the INTERVAL dataset.

This study utilised SomaLogic to explore the relationship between BMI and plasma proteins in unprecendented scope and detail, in both an observational and MR framework. We provide evidence for a broad impact of higher adiposity on the human proteome. Causal evidence was strongest for BMI in relation to proteins involved in regulating appetite, sex hormones, and inflammation. Identification of BMI-driven protein changes could provide therapeutic targets for prevention of obesity-related disease. Protein alterations were also found to be enriched for genes related to cardiovascular and metabolic disease. Altogether, these results help to focus attention onto new potential proteomic signatures of obesity-related disease. Further characterisation of the role of such proteomic profiles in disease using MR is warranted.

Chapter 5

Using a randomised control trial to explore the effects of caloric restriction-induced weight loss on the plasma proteome

5.1 Background

Obesity is associated with an increased risk of type II diabetes (T2D), cardiovascular diseases, musculoskeletal diseases, and types of cancer.[7–9] These associations are well established, however mechanisms of disease are less clear. It is likely that a change in the composition of proteins circulating in the blood plays a role in obesity-related disease.[33] The previous results chapter (Chapter 4) examined the effect of BMI on the plasma proteome and provided estimates for the average change in proteins per average difference in BMI in a general population, using

Mendelian randomization (MR). There was evidence of a BMI effect on proteins such as leptin, sex hormone binding globulin (SHBG) and fatty acid binding protein 4 (FABP4). A change in circulating levels of these proteins have been implicated in cardiovascular diseases.

Although MR helps to overcome issues inherent to observational studies such as confounding and reverse causation, MR has its own limitations. For example, it is possible that genetic variants used for the genetic risk score could affect protein levels through routes other than changes in BMI. This is known as horizonal pleiotropy.[109] Triangulation can be used in efforts to further determine the BMI effects on the plasma proteome. Triangulation is defined as strengthening causal inference through the combination of study designs which each have separate sources of bias.[110] Therefore, to complement the Mendelian randomization in the previous chapter, the Diabetes REmission Clinical Trial (DiRECT) can be utilised. The DiRECT trial consisted of a group of patients who had overweight/obesity and T2D who were either given guideline T2D care or an intervention to help with weight loss and T2D remission. This chapter therefore aimed to:

1) Examine the effect of BMI change on the plasma proteome.
2) Compare BMI-protein effects across MR and RCT study designs (Results from Chapter 4).

5.2 Methods

5.2.1 Study design and participants

Samples analysed within this study were collected from participants enrolled in the Diabetes Remission clinical trial (DiRECT). DiRECT was a cluster-randomised trial which took place at 49 primary care practices in Scotland and Tyneside.[111]

Ethics approval was granted from West 3 Ethics Committee (January, 2014) and the National Health Service (NHS). Participants were recruited between 25th July 2014 and 5th August 2016. Details of the protocol have previously been published.[112] Participants enrolled were between 20-65 years, diagnosed with T2D within previous 6 years, had a BMI of between 27-45kg/m^2. Participants were excluded if they were: using insulin, had a HbA1c concentration of $\geq$12 % ($\geq$108 mmol/mol), weight loss of >5 kg in the preceding 6 months, an estimated glomerular filtration rate of <30 mL/min per 1.732 m^2. More exclusion criteria are included in the main trial paper.[111] Participants in the control group received best-practice care by guidelines. The intervention group were asked to follow the Counterweright-Plus weight management programme.[113] This programme consisted of a total diet replacement (TDR) phase using a low energy diet (825-853 kcal/day) for 3-5 months, followed by structured food reintroduction of 2-8 weeks, with ongoing monthly long-term weight loss maintenance visits. Those in the intervention group had their antidiabetic and antihypertensive drugs discontinued. In total, there were 146 patients in the control group and 119 in the intervention group who remained enrolled in the study at 1 year, as detailed in Figure 5.1.

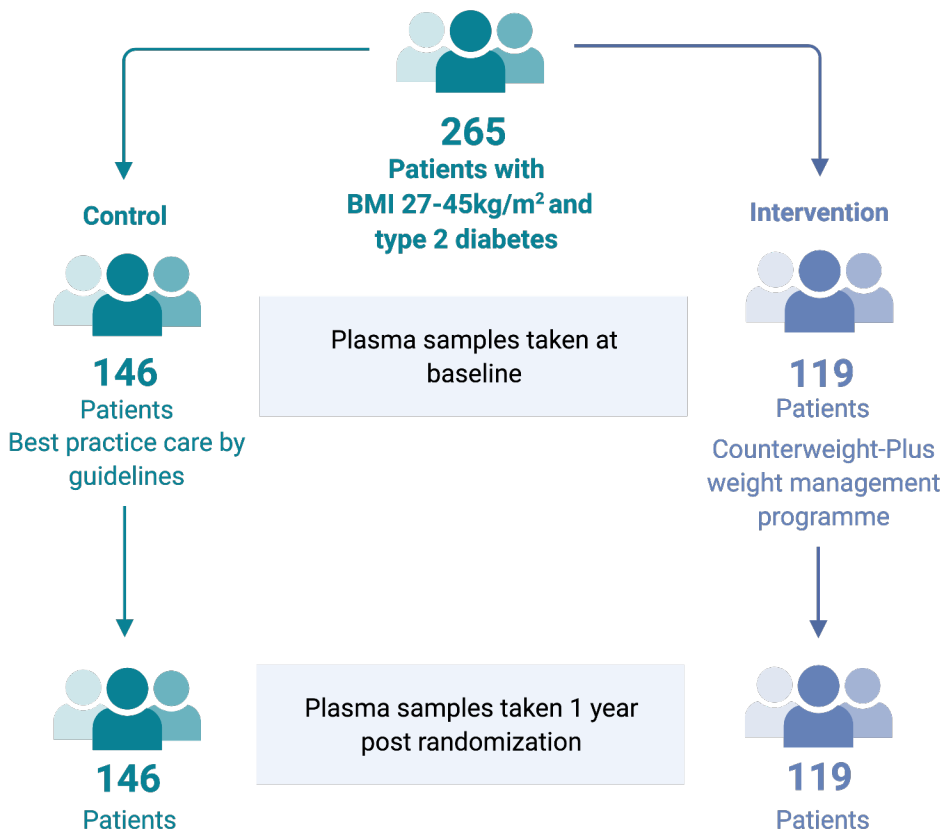

Figure 5.1: Overview of participants in the DiRECT trial

5.2.2 Included variables

Age and sex were self-reported. Height was measured with the Frankfort plane horizonal with a protable stadiometer (Chasmors Ltd, London). Weight was measured using Class 111 approved calibrated scales (Marsden Group UK). Blood was donated at various timepoints including at baseline (week 0) and at 1 year (week 52), where HDL cholesterol, tryglycerides, HbA1c and plasma glucose were measured. Systolic blood pressure was measured with the patient seated, rested and with legs uncrossed for $\geq$5 mins. BMI was calculated by dividing

the weight (kg) by the square of the height (m). BMI change was calculated by subtracting BMI at baseline from BMI at 1 year. This BMI change was rank normal transformed (function rntransform() from "moosefun" R package `https://github.com/hughesevoanth/moosefun/`). The centre attended and list size of the practice (>5700 or $\leq$5700) determined whether participants were allocated to control or intervention and therefore were recorded.

5.2.3 Proteomics

Plasma samples were taken from participants at baseline at and 1-year post-randomisation. Protein detection was performed by the SomaScan assay by SomaLogic. As mentioned before, this technique uses Slow Off-rate Modified Aptamers (SOMAmers) which make direct contact with proteins and quantifies protein levels by using a DNA microarray.[84] There were 5284 proteins included in the array, of which 4601 proteins passed internal quality control checks. The proteomic data was further cleaned using the "metaboprep" R package,[114] with protein levels from both timepoints all QC'd together. This R package identified no extreme missingness, removed outliers based on principal component analyses and calculated the number of independent proteins by using pairwise correlation coefficients between proteins (869 representative proteins). The package also implemented the Shapiro-Wilk test and took the W statistic to identify proteins which have a normal distribution (W$\geq$0.95). Only 399 proteins had W statistics $\geq$0.95. This data was then log2 transformed to match protein transformation by Sun et al.[82] then adjusted for age and sex (data had already undergone technical adjustment). Protein change was calculated for each individual by subtracting the baseline protein levels from protein levels at one-year. These protein change values were rank normal transformed.

5.2.4 Statistical analysis

Analyses were performed using R version 3.6.1.

A total of 265 participants were included in the current analysis. Participant baseline characteristics were summarised as the mean and SD. Baseline characteristics were calculated for both control group (N=146) and intervention (N=119) and were compared across groups using either a two-tailed t-test or a Chi^2 test.

The association between BMI change and protein change was explored using multiple linear regression adjusting for centre, list size, age and sex (function lm() from base R "stats" package). Due to the nature of linear regression, outputs of regression provide the mean individual change in protein (in SDs) per SD increase in BMI. Therefore if more weight loss reduces the levels of protein, the beta coefficient is positive.[115] The association between potential confounders (age and sex) and both exposure (BMI change) and outcome (protein change) were explored using linear regression. The difference in BMI change in intervention group compared to control was performed using linear regression, adjusting for centre, list size, age and sex. Covariables (age, sex, centre and list size) were compared across treatment groups. Treatment group was therefore used as an instrument for BMI change in a two stage least squares analysis :ref(fig:direct-tsls) to estimate the effects of BMI change on protein change using ivreg() function from the "AER" R package (`https://github.com/cran/AER/blob/master/R/ivreg.R`). As a sensitivity analysis, for proteins which associate with BMI change, baseline protein levels were compared across control and intervention groups. P values were adjusted for multiple testing based on the number of representative proteins ($0.05/869 = 5.8 \times 10^{-5}$).

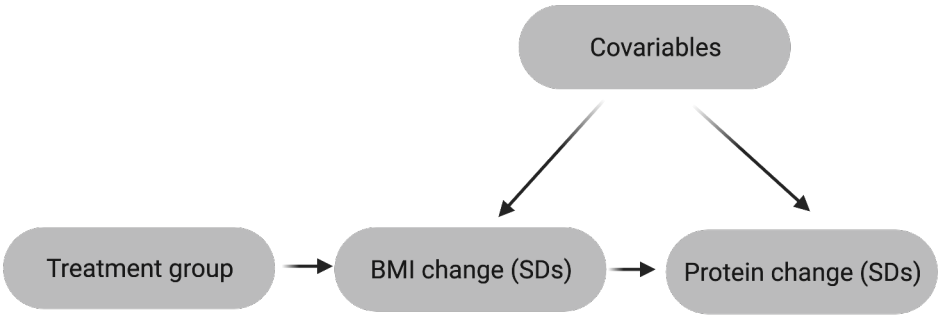

Figure 5.2: Schematic of two stage least squares analysis

Enrichment analysis was performed by xxxxxxxxxx.

BMI change-protein change estimates from the multiple linear regression analyses and two stage least squares analyses were compared using a Deming regression (deming() function from "Deming" R package. Deming regression is a special type of total least squares (TLS) and differs from linear regression (OLS) as it accounts for error in both the x- and y- axes. In efforts to further elucidate which proteins are altered by adiposity, a comparison between the BMI-protein associations in INTERVAL (Chapter 4) and the BMI change-protein change associations in the current analyses was also performed using a Deming regression. There were 2803 unique proteins overlapping across the two cohorts. Estimates across the two cohorts were compared using a Deming regression.

5.3 Results

5.3.1 Participant characteristics

Participants displayed similar characteristics across control and intervention groups (Table 5.1). Baseline BMI was similar across treatment groups (34.2 kg/m^2 with

SD of 4.3 in the control group vs 35.0 kg/m^2 with SD of 4.6 in the intervention group, p=0.19). Other variables relating to cariometabolic health were also similar across groups, including systolic blood pressure, HbA1c, glucose and insulin levels. The participants in the intervention group were younger and included fewer participants from Scotland but more from Tyneside.

Table 5.1: DiRECT participant characteristics

Variable	Control mean (SD) or %	Control N	Intervention mean (SD) or %	Intervention N	P value for difference (Two tailed t-test or Chi2 test)
Age	56.2 (6.9)	146	53.9 (7.1)	119	0.01
Sex		146		119	0.54
Male	62 %		57 %		
Female	38 %		43 %		
Body mass index (kg/m2)	34.2 (4.3)	146	35.0 (4.6)	119	0.19
Systolic Blood Pressue (mmHg)	137 (15)	146	135 (17)	119	0.19
HbA1c (mmol/mol	58 (11)	146	60 (13)	119	0.15
Glucose (mmol/l)	8.8 (2.5)	146	9.2 (3.2)	118	0.24
Insulin (uu/ml)	22 (14)	146	24 (15)	119	0.27
Cholesterol (mmol/l)	4.3(1.1)	146	4.3 (1.1)	118	0.9
HDL (mmol/l)	1.2 (0.3)	146	1.1 (0.3)	118	0.04
Triglycerides (mmol/l)	1.9 (0.9)	146	2.0 (1.4)	118	0.48
Diabetes duration (years)	3.0 (1.7)	146	3.2 (1.6)	119	0.37
Number of antidiabetic medications		146		119	0.5
0	23 %		24 %		
1	53 %		44 %		
2	18 %		23 %		
3	5 %		7 %		
4	1 %		3 %		
Centre		146		119	0.01
Scotland	77 %		62 %		
Tyneside	23 %		38 %		
List size		146		119	0.2
>5700	49 %		58 %		
=<5700	51 %		42 %		

5.3.2 Observational association between BMI change and protein change

Linear regression was used to estimate the effect of BMI change on protein change. After adjusting for multiple testing ($p<5.8x10^{-5}$), 254 proteins out of 4601 proteins were associated with BMI change (5.5 % of all protein tested). Although overall there was weight loss and a reduction in BMI, beta coefficients can be interpreted as the change in protein level (SDs) per SD increase in BMI. Negative associations were observed with BMI change and scavenger receptor class A member 5 (SCAR5, -0.62 SDs per SD increase in BMI, 95% CI -0.72 to -0.53 $p=3.7x10^{-29}$) and apolipoprotein F (-0.58 SDs per SD increase in BMI, 95% -0.68 to -0.48, $p=1.8x10^{-24}$). BMI change was positively associated in changes in levels of Proto-oncogene tyrosine-protein kinase receptor Ret (RET) (0.59 SDs increase per SD increase in BMI, 95% CI 0.49 to 0.69 $p=1.6x10^{-25}$), growth hormone receptor (GHR) (0.59 SDs increase per SD increase in BMI, 95% CI 0.49 to 0.69 $p=5.8x10^{-25}$).

5.3.3 Association between confounders and exposure/outcomes

There was weak evidence for an association between age and BMI change (0.02 SDs increase in BMI change per 1 year, 95 % CIs -0.0003 to 0.034, p=0.054), where an increase in BMI change means that the participant is less likely to reduce their BMI. There was no evidence that sex was associated with BMI change (BMI change 0.073 SDs increase in BMI change in females compared with males, 95 % CI -0.17 to 0.32, p=0.56). Age and sex displayed weak associations with protein changes, but these did not pass multiple testing adjustment.

5.3.4 Association between treatment group and BMI change

Mean BMI change in all participants was -1.8kg/m^2 (SD 2.5 kg/m^2), with a mean change in the control group of -0.4 kg/m^2 (SD 1.3) vs -3.6kg/m^2 (SD 2.5) in the intervention. The difference in BMI change comparing intervention to control was -3.2 kg/m^2 (SE 0.25, p=1.52x10^{-29}). These analyses suggest that treatment group is a valid instrument for BMI change.

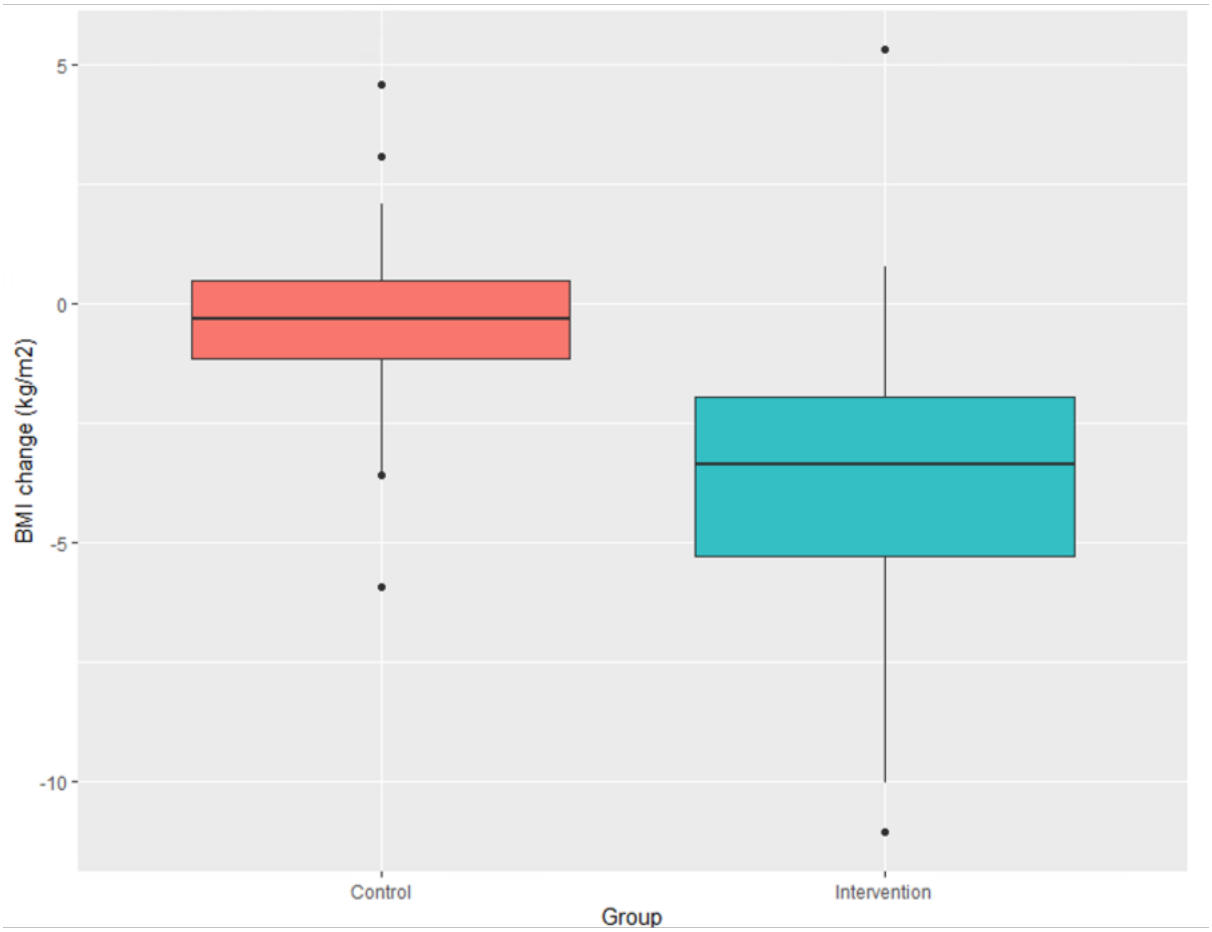

Figure 5.3: Box plot of the distribution of BMI change (kg/m^2) by treatment group with the median and interquartile ranges displayed by the box

5.3.5 Two-stage least squares analysis

Treatment group was used as an instrument for BMI change in a two-stage least squares analysis. After adjustment for multiple testing, BMI change was associated with changes in levels of 171 proteins (3.7 %). The protein which showed the strongest association with BMI change was SCAR5 (-0.91 SDs per SD increase in BMI, 95% CI -1.07 to -0.75, $p=1.54x10^{-23}$). Other proteins displayed a positive association with BMI change, including Fatty acid-binding protein, heart (FABP1, 0.86 SDs increase per SD increase in BMI, 95% CI 0.69-1.04, $p=6.3x10^{-19}$). Aminoacyclase-1 was also positively associated with BMI change (0.80 SDs increase in protein per SD increase in BMI, 95% CI 0.64 to 0.97, $p=4.9x10^{-18}$).

5.3.6 Comparison of observational linear regression and TSLS results for the effect of BMI change on protein change

A total of 110 proteins were only associated within the observational linear regression model, whereas 27 proteins were only associated in the TSLS analysis using treatment group as an instrument for BMI change. 144 proteins were associated with BMI in both analyses (5.4), with largest effect estimates across methods for SCAR5, Adiponectin, Apolipoprotein F, growth hormone receptor (GHR), receptor-type tyrosine-protein phosphatase U (PTPRU) and sex hormone binding globulin. These can be seen in Figure 5.5.

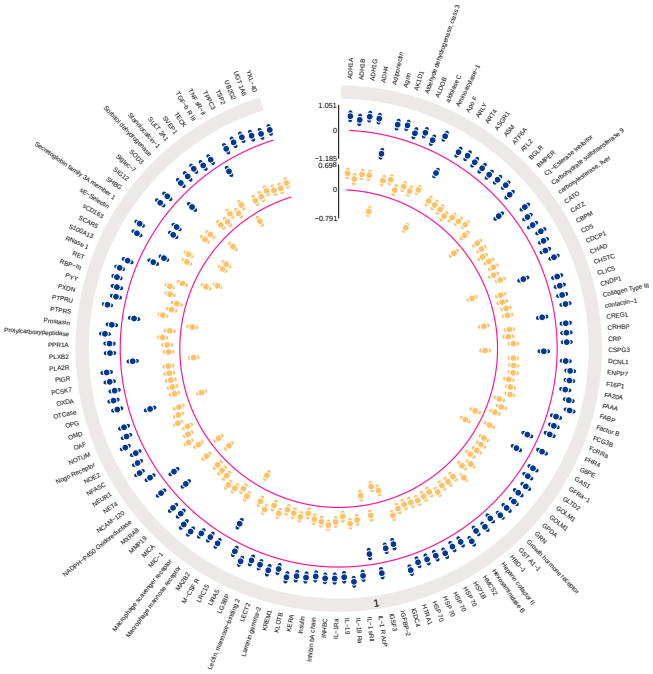

Figure 5.4: Circos plot displaying the protein change (SDs) per SD increase in BMI in DiRECT, where the blue estimates are the two stage least squares estimates and the yellow estimates are from the multiple linear regression. Proteins presented were associated in both models at p=5.8x10^{-5}

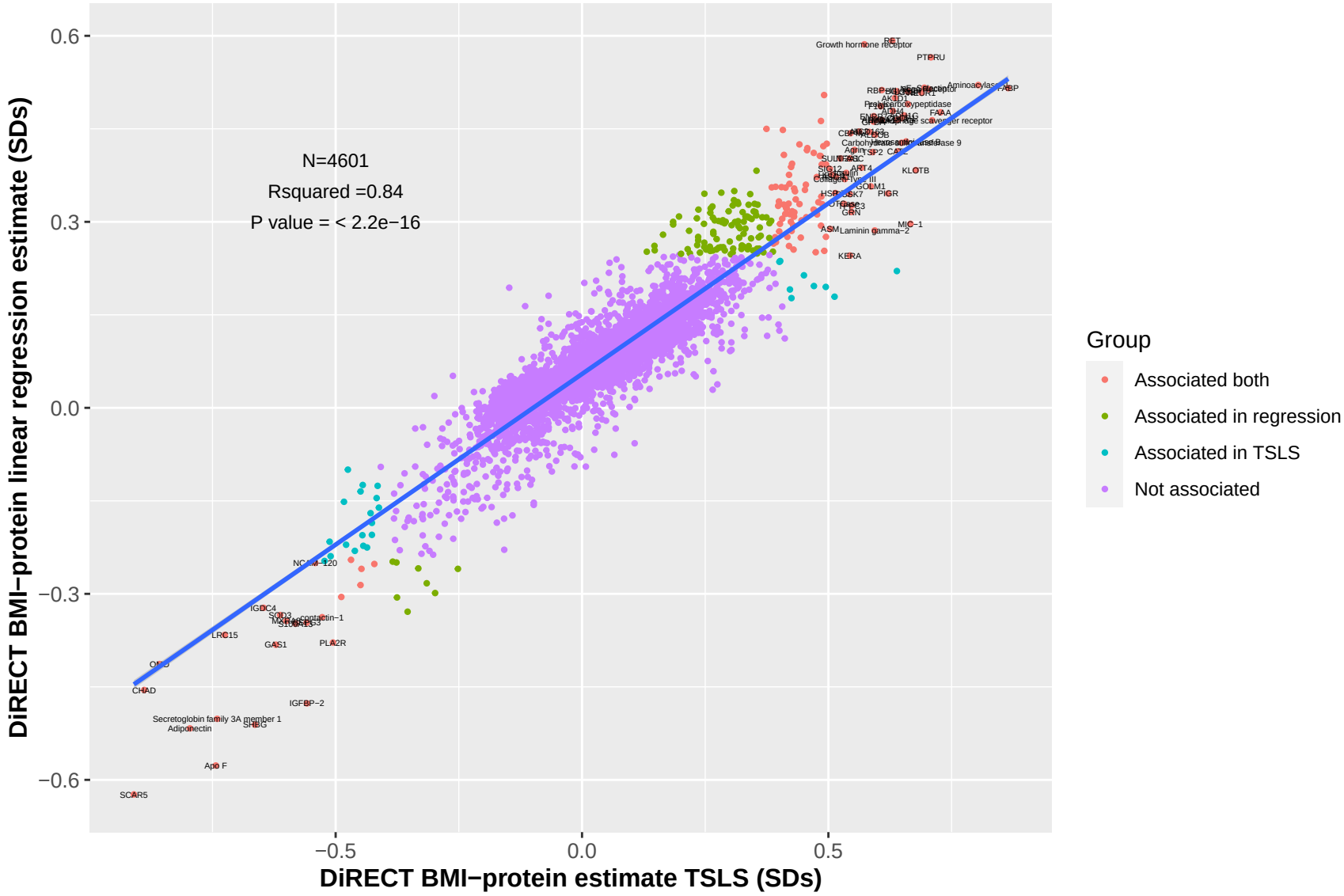

Figure 5.5: Scatter plot displaying the estimate for the change in protein (SDs) per SD increase in BMI in the two stage least squares and linear regression models. Points are coloured by which models the proteins were associated with BMI change ($p<5.8x10^{-5}$).

5.3.7 Sensitivity analysis

Proteins which were associated with BMI change were compared across control and intervention groups at baseline. There was not strong evidence for any differences in baseline levels of proteins ($p<5.8x10^{-5}$).

5.3.8 Enrichment analysis

5.3.9 Comparison of BMI effects across DiRECT and INTERVAL

BMI-protein estimates from this study were compared with effect estimates from INTERVAL (Chapter 4). Both cohorts provided evidence for an effect of BMI on a range of proteins. Figure 5.6 shows the BMI-protein estimates for the proteins which were included in both DiRECT and INTERVAL. There was consistent evidence across both studies for a BMI-increasing effect on proteins including fumarylacetoacetase (FAAA), alcohol dehydrogenase 4 (ADH4), growth hormone receptor (GHR) and carboxypeptidase M (CBPM). Furthermore, higher BMI had a lowering effect on levels of sex hormone binding globulin (SHBG), as well as insulin-like growth factor binding protein 1/2 (IGFBP1/2).

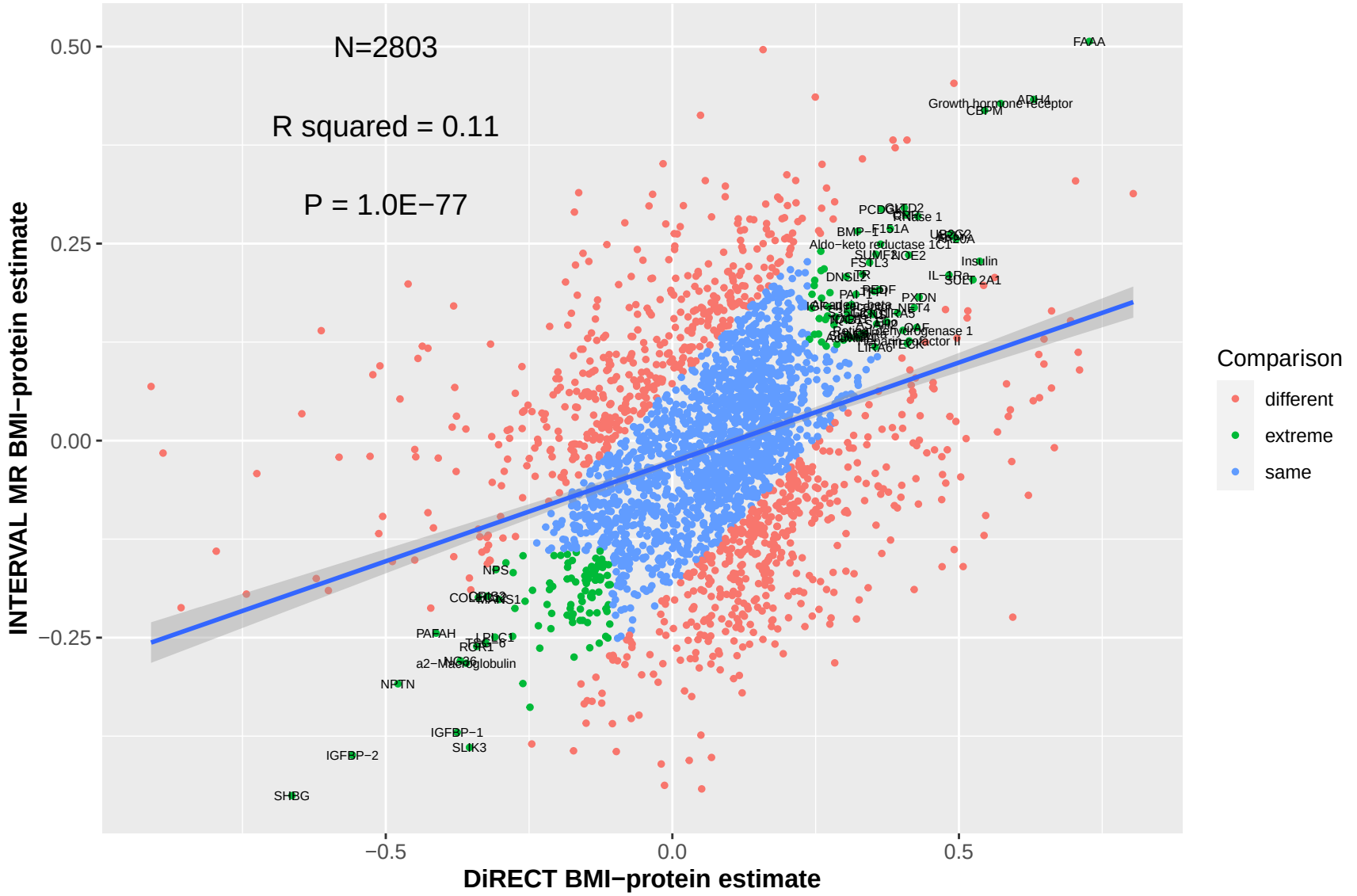

Figure 5.6: Scatter plot comparing the two stage least squares estimate for the effect of BMI change on protein change in DiRECT with the Mendelian randomisation estimate for the effect of BMI on protein levels in INTERVAL. Proteins that are similar across analyses with a large effect size in both studies are coloured in green, whereas those which display similar estimates across studies with small effect size are shown in blue. Proteins do not show similar effects across studies are shown in pink. A linear regression line is indicated in blue.

5.4 Discussion

In this chapter, a clinical trial implementing caloric restriction-induced weight loss was utilised to assess the effect of BMI change on the plasma proteome.

This analysis provided evidence for a broad effect of weight loss on the plasma proteome. This trial

Table 5.2 provides a summary of findings from the literature from previous studies which have explored the effect of caloric restriction on plasma proteins. As detailed in the table, previous studies have mainly used mass spectrometry. One study used SomaLogic, however this was the panel that

Table 5.2: Summary of current literature on the effect of weight loss on the plasma proteome

Paper	Author, year	Proteomic platform	Study summary
Proteomics reveals the effects of sustained weight loss on the human plasma proteome	Geyer, 2016	Mass spectrometry	Participants: 43 individuals with obesity before and after 8-week weight loss. Total proteins associated: 93/1294, Overlapping findings: Weight loss leads to increase in ApoF, ITIH3, SHBG, corticosteroid-binding globulin.
Proteomic profiles before and during weight loss: Results from randomized trial of dietary intervention	Figarska, 2020	Olink CVD II, CVD III and inflammation	Participants: 609 from the diet intervention examining the factors interacting with treatment success (DIETFITS) study. Total proteins associated: 130/26 associated with change in BMI. Overlapping findings: Changes in IGFBP1/IGFBP2, E-selectin, CD163.
The differential plasma proteome of obese and overweight individuals undergoing a nutritional weight loss and maintenance intervention	Oller-Moreno, 2018	Mass spectrometry	Participants: 473 in the nutritional programme DiOGenes. Total proteins associated: 39/183. Overlapping findings: Changes in SHBG, Adiponectin, IGFBP2, galectin-3-binding protein.
Integrative Personal Omics Profiles during Periods of Weight Gain and Loss	Piening, 2018	Luminex and ProSeek by Olink	Participants: 23 with overweight or obesity undergoing weight loss. Total proteins associated: 27/276. Overlapping findings: changes in leptin.
Protein quantitative trait locus study in obesity during weight-loss identifies a leptin regulator	Carayol, 2017	SomaLogic	Participants: 494 participants on low DiOGenes on low calorie diet. Total proteins associated: 104/1129. Overlapping findings: Changes in leptin, GHR, TIG2, IGFBP2, RET proto-oncogene (granulysin).
Analysis of 1508 plasma samples by capillary-flow data-independent acquisition profiles proteomics of weight loss and maintenance	Bruderer, 2019	Mass spectrometry	Participants: 477 in DiOGenes. Total proteins associated: 271/465. Overlapping findings: Changes in Apolipoprotein F, INHBC, SHBG.

1. Overall effect
2. Comparison to other studies? Caloric restriction and bariatric surgery?

- [116] Longitudinal weight loss (7 time points over 14 months) in 43 obese participants using mass spec. Weight loss led to increase in ApoF, increase in ITIH3 and an increase in SHBG, and increase in Corticosteroid-binding globulin.
- Figarska 130/263 olink proteins associated with changes in BMI. If protein decreased upon weight loss = positive beta coefficient. 118 decreased with weight loss, 12 increased. Supp Table 2. Those in agreement = igfb1/igfbp2, e-selectin, cd163.
- [117] Using LC-MS/MS. Compared before and after. 473 individuals with overweight/obesity. SHBG, Adiponectin, IGFBP2, galectin-3-binding protein.

- [118] Hypercaloric challenge followed by hypocaloric challenge in 23 subjects. Supplementary Tables have results but they are confusing.
- [119] 512 participants on low calorie diet for 8 weeks. DIOgenes SomaLogic. 1129 proteins - 104 associated with weight loss. Weight loss led to changes in leptin, GHR, TIG2, IGFBP2, RET proto-oncogene (granulysin)
- [120] DiOGenes. Low calorie for 8 weeks, followed by weight maintenance diet. Overlap with proteins (supplement). Apolipoprotein F, INHBC, SHBG.

3. Comparison of estimates across study designs.
4. Biology - inflammation?
5. Implications for platelet function / thrombosis. Stringdb. input beta coefficients with uniprot IDs and there is pathway enrichment for platelet function and platelet aggregation (256 genes).

6. limitations - not solely weight loss effect

- Apolipoprotein F (APOF) two smaple MR causal effect on myocardial infarction and coronary heart disease by Wald ratio.
- SHBG MR - increased shbg leads to reduction in cardioembolic stroke, platelet count.
- RET - cancers/Alzheimer's.
- FABP1 - increased levels increases MPV.
- Prolylcarboxypeptidase CHD and MI.
- Fumarylacetoacetase (FAAA/FAH) increased protein reduces risk of haemorrhage. NB not possible to test for pleiotropy - other diseases associated with these proteins too.

Chapter 6

Protein DVT MR, enrichment analyses???

Discussion

6.1 Overall findings - table or figure ???

6.2 Strengths and limitations

6.3 Future studies

6.4 Wider implications of findings (e.g. clinical)

6.5 Overal conclusion

Appendix A

The First Appendix

This first appendix includes all of the R chunks of code that were hidden throughout the document (using the `include = FALSE` chunk tag) to help with readibility and/or setup.

In the main Rmd file

In Chapter ??:

Appendix B

The Second Appendix, for Fun

References

1. Blüher, M. Obesity: global epidemiology and pathogenesis. *Nat Rev Endocrinol* **15**, 288–298 (2019).

2. The Lancet Public Health. Tackling obesity seriously: the time has come. (2018). doi:10.1016/S2468-2667(18)30053-7

3. Wade, K. H., Carslake, D., Sattar, N., Davey Smith, G. & Timpson, N. J. BMI and Mortality in UK Biobank: Revised Estimates Using Mendelian Randomization. *Obesity (Silver Spring)* **26**, 1796–1806 (2018).

4. (NCD-RisC), N. C. D. R. F. C. Trends in adult body-mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 192 million participants. *Lancet* **387**, 1377–1396 (2016).

5. Templin, T., Hashiguchi, T. C. O., Thomson, B., Dieleman, J. & Bendavid, E. The overweight and obesity transition from the wealthy to the poor in low- And middleincome countries: A survey of household data from 103 countries. *PLoS Medicine* (2019). doi:10.1371/journal.pmed.1002968

6. Bell, J. A. *et al.* Associations of Body Mass and Fat Indexes With Cardiometabolic Traits. *J Am Coll Cardiol* **72**, 3142–3154 (2018).

7. Garg, S. K., Maurer, H., Reed, K. & Selagamsetty, R. Diabetes and cancer: two diseases with obesity as a common risk factor. *Diabetes Obes Metab* **16**, 97–110 (2014).

8. Khan, S. S. *et al.* Association of Body Mass Index With Lifetime Risk of Cardiovascular Disease and Compression of Morbidity. *JAMA Cardiol* **3**, 280–287 (2018).

9. Kortt, M. & Baldry, J. The association between musculoskeletal disorders and obesity. *Aust Health Rev* **25**, 207–214 (2002).

10. Bhaskaran, K. *et al.* Body-mass index and risk of 22 specific cancers: a population-based cohort study of 524 million UK adults. *Lancet* **384**, 755–765 (2014).

11. Dixon, P., Davey Smith, G. & Hollingworth, W. The Association Between Adiposity and Inpatient Hospital Costs in the UK Biobank Cohort. *Appl Health Econ Health Policy* **17**, 359–370 (2019).

12. Dixon, P., Hollingworth, W., Harrison, S., Davies, N. M. & Smith, G. D. The causal effect of adiposity on hospital costs: Mendelian Randomization analysis of over 300,000 individuals from the UK Biobank. *bioRxiv* 589820 (2019). doi:10.1101/589820

13. Nordestgaard, B. G. *et al.* The effect of elevated body mass index on ischemic heart disease risk: causal estimates from a Mendelian randomisation approach. *PLoS Med* **9**, e1001212 (2012).

14. Dale, C. E. *et al.* Causal Associations of Adiposity and Body Fat Distribution With Coronary Heart Disease, Stroke Subtypes, and Type 2 Diabetes Mellitus: A Mendelian Randomization Analysis. *Circulation* **135**, 2373–2388 (2017).

15. Carreras-Torres, R. *et al.* The Role of Obesity, Type 2 Diabetes, and Metabolic Factors in Pancreatic Cancer: A Mendelian Randomization Study. *J Natl Cancer Inst* **109**, (2017).

16. Wing, R. R. *et al.* Benefits of modest weight loss in improving cardiovascular risk factors in overweight and obese individuals with type 2 diabetes. *Diabetes Care* **34**, 1481–1486 (2011).

17. Cercato, C. & Fonseca, F. A. Cardiovascular risk and obesity. (2019). doi:10.1186/s13098-019-0468-0

18. Kern, P. A., Ranganathan, S., Li, C., Wood, L. & Ranganathan, G. Adipose tissue tumor necrosis factor and interleukin-6 expression in human obesity and insulin resistance. *American Journal of Physiology - Endocrinology and Metabolism* (2001). doi:10.1152/ajpendo.2001.280.5.e745

19. Esser, N., Legrand-Poels, S., Piette, J., Scheen, A. J. & Paquot, N. Inflammation as a link between obesity, metabolic syndrome and type 2 diabetes. *Diabetes Res Clin Pract* **105**, 141–150 (2014).

20. Bell, J. A. *et al.* Associations of Body Mass and Fat Indexes With Cardiometabolic Traits. *J Am Coll Cardiol* **72**, 3142–3154 (2018).

21. Würtz, P. *et al.* Metabolic signatures of adiposity in young adults: Mendelian randomization analysis and effects of weight change. *PLoS Med* **11**, e1001765 (2014).

22. Imming, P., Sinning, C. & Meyer, A. Drugs, their targets and the nature and number of drug targets. *Nat Rev Drug Discov* **5**, 821–834 (2006).

23. Timpson, N. J. *et al.* C-reactive protein levels and body mass index: elucidating direction of causation through reciprocal Mendelian randomization. *Int J Obes (Lond)* **35**, 300–308 (2011).

24. Cominetti, O. *et al.* Obesity shows preserved plasma proteome in large independent clinical cohorts. *Sci Rep* **8**, 16981 (2018).

25. Gold, L., Walker, J. J., Wilcox, S. K. & Williams, S. Advances in human proteomics at high scale with the SOMAscan proteomics platform. *N Biotechnol* **29**, 543–549 (2012).

26. Goudswaard, L. J. *et al.* Higher body mass index raises immature platelet count: evidence from Mendelian randomization analyses. *medRxiv* (2021). doi:10.1101/2021.05.19.21257443

27. Bhaskaran, K., Dos-Santos-Silva, I., Leon, D. A., Douglas, I. J. & Smeeth, L. Association of BMI with overall and cause-specific mortality: a population-based cohort study of 36 million adults in the UK. *Lancet Diabetes Endocrinol* **6**, 944–953 (2018).

28. Wolk, R., Berger, P., Lennon, R. J., Brilakis, E. S. & Somers, V. K. Body mass index: a risk factor for unstable angina and myocardial infarction in patients with angiographically confirmed coronary artery disease. *Circulation* **108**, 2206–2211 (2003).

29. Koupenova, M., Kehrel, B. E., Corkrey, H. A. & Freedman, J. E. Thrombosis and platelets: an update. *Eur Heart J* **38**, 785–791 (2017).

30. Puurunen, M. K. *et al.* ADP Platelet Hyperreactivity Predicts Cardiovascular Disease in the FHS (Framingham Heart Study). *J Am Heart Assoc* **7**, (2018).

31. Barrachina, M. N. *et al.* GPVI surface expression and signalling pathway activation are increased in platelets from obese patients: Elucidating potential anti-atherothrombotic targets in obesity. *Atherosclerosis* **281**, 62–70 (2019).

32. Nardin, M. *et al.* Body Mass Index and Platelet Reactivity During Dual Antiplatelet Therapy With Clopidogrel or Ticagrelor. *J Cardiovasc Pharmacol* **66**, 364–370 (2015).

33. Goudswaard, L. J. *et al.* Effects of adiposity on the human plasma proteome: observational and Mendelian randomisation estimates. *International Journal of Obesity* (2021). doi:10.1038/s41366-021-00896-1

34. Würtz, M., Hvas, A. M., Kristensen, S. D. & Grove, E. L. Platelet aggregation is dependent on platelet count in patients with coronary artery disease. *Thromb Res* **129**, 56–61 (2012).

35. Slavka, G. *et al.* Mean platelet volume may represent a predictive parameter for overall vascular mortality and ischemic heart disease. *Arterioscler Thromb Vasc Biol* **31**, 1215–1218 (2011).

36. Gill, D., Monori, G., Georgakis, M. K., Tzoulaki, I. & Laffan, M. Genetically Determined Platelet Count and Risk of Cardiovascular Disease. *Arterioscler Thromb Vasc Biol* **38**, 2862–2869 (2018).

37. Unay Demirel, O., Ignak, S. & Buyukuysal, M. C. Immature Platelet Count Levels as a Novel Quality Marker in Plateletpheresis. *Indian J Hematol Blood Transfus* **34**, 684–690 (2018).

38. Ibrahim, H. *et al.* Association of immature platelets with adverse cardiovascular outcomes. *J Am Coll Cardiol* **64**, 2122–2129 (2014).

39. Lev, E. I. Immature Platelets: Clinical Relevance and Research Perspectives. *Circulation* **134**, 987–988 (2016).

40. Bernlochner, I. *et al.* Impact of immature platelets on platelet response to ticagrelor and prasugrel in patients with acute coronary syndrome. *Eur Heart J* **36**, 3202–3210 (2015).

41. Ibrahim, H., Nadipalli, S., DeLao, T., Guthikonda, S. & Kleiman, N. S. Immature platelet fraction (IPF) determined with an automated method predicts clopidogrel hyporesponsiveness. *Journal of Thrombosis and Thrombolysis* (2012). doi:10.1007/s11239-011-0665-7

42. Furuncuoğlu, Y. *et al.* How obesity affects the neutrophil/lymphocyte and platelet/lymphocyte ratio, systemic immune-inflammatory index and platelet indices: a retrospective study. *Eur Rev Med Pharmacol Sci* **20**, 1300–1306 (2016).

43. Han, S. *et al.* Associations of Platelet Indices with Body Fat Mass and Fat Distribution. *Obesity (Silver Spring)* **26**, 1637–1643 (2018).

44. Coban, E., Ozdogan, M., Yazicioglu, G. & Akcit, F. The mean platelet volume in patients with obesity. *Int J Clin Pract* **59**, 981–982 (2005).

45. Heffron, S. P., Marier, C., Parikh, M., Fisher, E. A. & Berger, J. S. Severe obesity and bariatric surgery alter the platelet mRNA profile. *Platelets* 1–8 (2018). doi:10.1080/09537104.2018.1536261

46. Vaduganathan, M. *et al.* Platelet reactivity and response to aspirin in subjects with the metabolic syndrome. *Am Heart J* **156**, 1002.e1–1002.e7 (2008).

47. Mijovic, R. *et al.* Reticulated platelets and antiplatelet therapy response in diabetic patients. *J Thromb Thrombolysis* **40**, 203–210 (2015).

48. Di Angelantonio, E. *et al.* Efficiency and safety of varying the frequency of whole blood donation (INTERVAL): a randomised trial of 45000 donors. *Lancet* **390**, 2360–2371 (2017).

49. Astle, W. J. *et al.* The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* **167**, 1415–1429.e19 (2016).

50. Akbari, P. *et al.* Genetic Analyses of Blood Cell Structure for Biological and Pharmacological Inference. *bioRxiv* 2020.01.30.927483 (2020). doi:10.1101/2020.01.30.927483

51. Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body mass index in 700000 individuals of European ancestry. *Hum Mol Genet* **27**, 3641–3649 (2018).

52. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559–575 (2007).

53. Team, R. C. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/. (2019).

54. Henningsen, A. & Hamann, J. D. Systemfit: A package for estimating systems of simultaneous equations in R. *Journal of Statistical Software* (2007). doi:10.18637/jss.v023.i04

55. Mumford, A. D. *et al.* Near-patient coagulation testing to predict bleeding after cardiac surgery: a cohort study. *Research and Practice in Thrombosis and Haemostasis* (2017). doi:10.1002/rth2.12024

56. Grove, E. L., Hvas, A. M., Mortensen, S. B., Larsen, S. B. & Kristensen, S. D. Effect of platelet turnover on whole blood platelet aggregation in patients with coronary artery disease. *J Thromb Haemost* **9**, 185–191 (2011).

57. Freynhofer, M. K. *et al.* Platelet turnover predicts outcome after coronary intervention. *Thromb Haemost* **117**, 923–933 (2017).

58. Cohen, A. *et al.* Immature platelets in patients hospitalized with Covid-19. *J Thromb Thrombolysis* (2020). doi:10.1007/s11239-020-02290-6

59. Klok, F. A. *et al.* Incidence of thrombotic complications in critically ill ICU patients with COVID-19. *Thrombosis Research* (2020). doi:10.1016/j.thromres.2020.04.013

60. Kaser, A. *et al.* Interleukin-6 stimulates thrombopoiesis through thrombopoietin: Role in inflammatory thrombocytosis. *Blood* (2001). doi:10.1182/blood.V98.9.2720

61. Nikolaou, C. K., Hankey, C. R. & Lean, M. E. J. Accuracy of on-line self-reported weights and heights by young adults. *Eur J Public Health* **27**, 898–903 (2017).

62. Marcucci, C. E. & Schoettker, P. *Perioperative hemostasis: Coagulation for anesthesiologists*. (2015). doi:10.1007/978-3-642-55004-1

63. Mannuß, S. Influence of different methods and anticoagulants on platelet parameter measurement. (2020). doi:10.1515/labmed-2020-0037

64. Briggs, C., Longair, I., Kumar, P., Singh, D. & Machin, S. J. Performance evaluation of the Sysmex haematology XN modular system. *Journal of Clinical Pathology* (2012). doi:10.1136/jclinpath-2012-200930

65. Chan, M. V., Armstrong, P. C. & Warner, T. D. 96-well plate-based aggregometry. (2018). doi:10.1080/09537104.2018.1445838

66. Rivera, J., Lozano, M. L., Navarro-Núñez, L. & Vicente, V. Platelet receptors and signaling in the dynamics of thrombus formation. *Haematologica* **94**, 700–711 (2009).

67. Yau, J. W., Teoh, H. & Verma, S. Endothelial cell control of thrombosis. *BMC Cardiovasc Disord* **15**, 130 (2015).

68. Badimon, L., Padró, T. & Vilahur, G. Atherosclerosis, platelets and thrombosis in acute ischaemic heart disease. *Eur Heart J Acute Cardiovasc Care* **1**, 60–74 (2012).

69. Gear, A. R. & Camerini, D. Platelet chemokines and chemokine receptors: linking hemostasis, inflammation, and host defense. *Microcirculation* **10**, 335–350 (2003).

70. Belin de Chantemele, E. J. & Stepp, D. W. Influence of obesity and metabolic dysfunction on the endothelial control in the coronary circulation. *J Mol Cell Cardiol* **52**, 840–847 (2012).

71. Maury, E. *et al.* Effect of obesity on growth-related oncogene factor-alpha, thrombopoietin, and tissue inhibitor metalloproteinase-1 serum levels. *Obesity (Silver Spring)* **18**, 1503–1509 (2010).

72. Blair, T. A., Moore, S. F. & Hers, I. Circulating primers enhance platelet function and induce resistance to antiplatelet therapy. *J Thromb Haemost* **13**, 1479–1493 (2015).

73. Pasquet, J. M. *et al.* Thrombopoietin potentiates collagen receptor signaling in platelets through a phosphatidylinositol 3-kinase-dependent pathway. *Blood* **95**, 3429–3434 (2000).

74. Falcinelli, E., Guglielmini, G., Torti, M. & Gresele, P. Intraplatelet signaling mechanisms of the priming effect of matrix metalloproteinase-2 on platelet aggregation. *J Thromb Haemost* **3**, 2526–2535 (2005).

75. Ezumi, Y., Uchiyama, T. & Takayama, H. Thrombopoietin potentiates the protein-kinase-C-mediated activation of mitogen-activated protein kinase/ERK kinases and extracellular signal-regulated kinases in human platelets. *Eur J Biochem* **258**, 976–985 (1998).

76. Willigen, G. van, Gorter, G. & Akkerman, J. W. Thrombopoietin increases platelet sensitivity to alpha-thrombin via activation of the ERK2-cPLA2 pathway. *Thromb Haemost* **83**, 610–616 (2000).

77. Safa, A. *et al.* Higher circulating levels of chemokines CXCL10, CCL20 and CCL22 in patients with ischemic heart disease. *Cytokine* **83**, 147–157 (2016).

78. Clemetson, K. J. *et al.* Functional expression of CCR1, CCR3, CCR4, and CXCR4 chemokine receptors on human platelets. *Blood* **96**, 4046–4054 (2000).

79. Gear, A. R. *et al.* Adenosine diphosphate strongly potentiates the ability of the chemokines MDC, TARC, and SDF-1 to stimulate platelet function. *Blood* **97**, 937–945 (2001).

80. Noori, F., Naeimi, S., Zibaeenezhad, M. J. & Gharemirshamlu, F. R. CCL22 and CCR4 Gene Polymorphisms in Myocardial Infarction: Risk Assessment of rs4359426 and rs2228428 in Iranian Population. *Clin Lab* **64**, 907–913 (2018).

81. Zheng, J. *et al.* Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nature Genetics* (2020). doi:10.1038/s41588-020-0682-6

82. Sun, B. B. *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).

83. Smith, G. D. & Ebrahim, S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* **32**, 1–22 (2003).

84. Rohloff, J. C. *et al.* Nucleic Acid Ligands With Protein-like Side Chains: Modified Aptamers and Their Use as Diagnostic and Therapeutic Agents. *Mol Ther Nucleic Acids* **3**, e201 (2014).

85. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).

86. Pain, O., Dudbridge, F. & Ronald, A. Are your covariates under control? How normalization can re-introduce covariate effects. *European Journal of Human Genetics* (2018). doi:10.1038/s41431-018-0159-6

87. Huang, da W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44–57 (2009).

88. Becker, K. G., Barnes, K. C., Bright, T. J. & Wang, S. A. The genetic association database. *Nat Genet* **36**, 431–432 (2004).

89. Zaghlool, S. B. *et al.* Revealing the role of the human blood plasma proteome in obesity using genetic drivers. *Nature Communications* (2021). doi:10.1038/s41467-021-21542-4

90. Klok, M. D., Jakobsdottir, S. & Drent, M. L. The role of leptin and ghrelin in the regulation of food intake and body weight in humans: a review. *Obes Rev* **8**, 21–34 (2007).

91. Millard, L. A. *et al.* MR-PheWAS: hypothesis prioritization among potential causal effects of body mass index on many outcomes, using Mendelian randomization. *Sci Rep* **5**, 16645 (2015).

92. Gruzdeva, O., Borodkina, D., Uchasova, E., Dyleva, Y. & Barbarash, O. Leptin resistance: underlying mechanisms and diagnosis. *Diabetes Metab Syndr Obes* **12**, 191–198 (2019).

93. Nakata, M., Yada, T., Soejima, N. & Maruyama, I. Leptin promotes aggregation of human platelets via the long form of its receptor. *Diabetes* **48**, 426–429 (1999).

94. Wallace, A. M. *et al.* Plasma leptin and the risk of cardiovascular disease in the west of Scotland coronary prevention study (WOSCOPS). *Circulation* **104**, 3052–3056 (2001).

95. Furuhashi, M., Saitoh, S., Shimamoto, K. & Miura, T. Fatty Acid-Binding Protein 4 (FABP4): Pathophysiological Insights and Potent Clinical Biomarker of Metabolic and Cardiovascular Diseases. *Clin Med Insights Cardiol* **8**, 23–33 (2014).

96. Xu, A. *et al.* Adipocyte fatty acid-binding protein is a plasma biomarker closely associated with obesity and metabolic syndrome. *Clin Chem* **52**, 405–413 (2006).

97. Gudmundsdottir, V. *et al.* Circulating protein signatures and causal candidates for type 2 diabetes. *Diabetes* (2020). doi:10.2337/db19-1070

98. Nakamura, R. *et al.* Serum fatty acid-binding protein 4 (FABP4) concentration is associated with insulin resistance in peripheral tissues, A clinical study. *PLoS One* **12**, e0179737 (2017).

99. Yeung, D. C. *et al.* Serum adipocyte fatty acid-binding protein levels were independently associated with carotid atherosclerosis. *Arterioscler Thromb Vasc Biol* **27**, 1796–1802 (2007).

100. Wallace, I. R., McKinley, M. C., Bell, P. M. & Hunter, S. J. Sex hormone binding globulin and insulin resistance. *Clin Endocrinol (Oxf)* **78**, 321–329 (2013).

101. Cooper, L. A., Page, S. T., Amory, J. K., Anawalt, B. D. & Matsumoto, A. M. The association of obesity with sex hormone-binding globulin is stronger than the association with ageing–implications for the interpretation of total testosterone measurements. *Clin Endocrinol (Oxf)* **83**, 828–833 (2015).

102. Goto, A. *et al.* Age, body mass, usage of exogenous estrogen, and lifestyle factors in relation to circulating sex hormone-binding globulin concentrations in postmenopausal women. *Clin Chem* **60**, 174–185 (2014).

103. Baglietto, L. *et al.* Circulating steroid hormone concentrations in postmenopausal women in relation to body size and composition. *Breast Cancer Res Treat* **115**, 171–179 (2009).

104. Ritchie, S. C. *et al.* Integrative analysis of the plasma proteome and polygenic risk of cardiometabolic diseases. *bioRxiv* (2021). doi:10.1101/2019.12.14.876474

105. Brion, M. J. A., Shakhbazov, K. & Visscher, P. M. Calculating statistical power in Mendelian randomization studies. *International Journal of Epidemiology* (2013). doi:10.1093/ije/dyt179

106. Burgess, S., Davies, N. M. & Thompson, S. G. Bias due to participant overlap in two-sample Mendelian randomization. *Genetic Epidemiology* (2016). doi:10.1002/gepi.21998

107. Rosengren, A. *et al.* Socioeconomic status and risk of cardiovascular disease in 20 low-income, middle-income, and high-income countries: the Prospective Urban Rural Epidemiologic (PURE) study. *Lancet Glob Health* **7**, e748–e760 (2019).

108. Tillmann, T. *et al.* Education and coronary heart disease: mendelian randomisation study. *BMJ* **358**, j3542 (2017).

109. Davies, N. M., Holmes, M. V. & Davey Smith, G. Reading Mendelian randomisation studies: A guide, glossary, and checklist for clinicians. *BMJ (Online)* (2018). doi:10.1136/bmj.k601

110. Lawlor, D. A., Tilling, K. & Smith, G. D. Triangulation in aetiological epidemiology. *International Journal of Epidemiology* (2016). doi:10.1093/ije/dyw314

111. Lean, M. E. *et al.* Primary care-led weight management for remission of type 2 diabetes (DiRECT): an open-label, cluster-randomised trial. *The Lancet* (2018). doi:10.1016/S0140-6736(17)33102-1

112. Leslie, W. S. *et al.* The Diabetes Remission Clinical Trial (DiRECT): protocol for a cluster randomised trial. *BMC Fam Pract* **17**, 20 (2016).

113. Lean, M. *et al.* Feasibility and indicative results from a 12-month low-energy liquid diet treatment and maintenance programme for severe obesity. *British Journal of General Practice* (2013). doi:10.3399/bjgp13X663073

114. Hughes, D. A. *et al.* metaboprep: an R package for pre-analysis data description and processing. *bioRxiv* 2021.07.07.451488 (2021). doi:10.1101/2021.07.07.451488

115. Figarska, S. M. *et al.* Proteomic profiles before and during weight loss: Results from randomized trial of dietary intervention. *Sci Rep* **10**, 7913 (2020).

116. Geyer, P. E. *et al.* Proteomics reveals the effects of sustained weight loss on the human plasma proteome. *Mol Syst Biol* **12**, 901 (2016).

117. Oller Moreno, S. *et al.* The differential plasma proteome of obese and overweight individuals undergoing a nutritional weight loss and maintenance intervention. *Proteomics - Clinical Applications* (2018). doi:10.1002/prca.201600150

118. Piening, B. D. *et al.* Integrative Personal Omics Profiles during Periods of Weight Gain and Loss. *Cell Systems* (2018). doi:10.1016/j.cels.2017.12.013

119. Carayol, J. *et al.* Protein quantitative trait locus study in obesity during weight-loss identifies a leptin regulator. *Nat Commun* **8**, 2084 (2017).

120. Bruderer, R. *et al.* Analysis of 1508 plasma samples by capillary-flow data-independent acquisition profiles proteomics of weight loss and maintenance. *Molecular and Cellular Proteomics* (2019). doi:10.1074/mcp.RA118.001288

Abbreviations

```{r abbreviations, echo=FALSE} library(readxl) library(tidyr) library(kableExtra) library(knitr) abbreviations <- read_xlsx("figure/Abbreviations/Abbreviations.xlsx" abbreviations <- abbreviations[order(abbreviations$Abbreviation),]

kinitr::kable(abbreviations, longtable = T, booktabs = T, format = "latex", escape = F, caption = "Full list of abbreviations") %>% kable_classic_2(full_width = F) %>% kable_styling(full_width = TRUE) %>% column_spec(1, width = "3cm") ```
```