

Causal relevance of inflammation to cardiovascular disease



Mohd Anisul Karim

St Edmund Hall

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2021

To my parents, Mohd Anwarul Islam and Nasima Shahin.

Acknowledgements

“The world breaks everyone and afterwards many are strong at the broken places” -
Ernest Hemingway, A Farewell to Arms

First, I would like to express my immense gratitude to my supervisors, Michael Holmes and Robin Walters, who have made me strong. They worked hard on me, promoted my growth at every step of my DPhil journey, and, most importantly, supported me to great lengths when it felt like nothing made sense. They went above and beyond to provide me with opportunities to travel to conferences, present my work, and advise on my clinical endeavours. Their wisdom and support enriched my learning experience and placed me on strong grounds for the progression of my clinical academic career.

Second, I would like to remember and convey my sincere thanks to those who kept me sane. Completion of this dissertation would have been impossible without the patient love and relentless support of all my friends, without the laughs and crude jokes of Shamir Montazid, without the practical wisdom of Monzilur Rahman, without the inspirations from Hassan Saad Ifti, without the food of Nazmus Sakib Tareque and Nusrat Priya, without the company of Raysa Tareque, without the homes of Oli and Katarzyna Jahid, without my friends back home (Shahed Sarj, Touhid Rahman, Rocky Ahmed, Mohammad Sayem, Sajia Islam, and Kamrun Koly) and, most certainly, without Amita Farzana - my wife and the beacon of all good things in my life. She put up with me when I made no sense and remained kind when I was not. She made me a better person and I am always in awe of her.

Third, I remain deeply indebted and grateful to all my colleagues at Big Data Institute, Nuffield Department of Population Health. They tolerated my naivete, supported me when I was stressing for no reason, and helped me navigate the surprising twists and turns of a DPhil student life. I am especially thankful to Iona Millwood, Robert Clarke, Kuang Lin (who performed the genome-wide association study and fine-mapping work of plasma C-reactive protein that supported my work in this thesis), Zammy Fairhurst-Hunter, Katerina Lind, Sarasvati Bahadursingh,

Huaidong Du, Christiana Kartsonaki, Peter Chan, Muriel Levy, Shajedur Rahman Shawon, Rob Smith, among many others, who have always advised and supported me without hesitation.

Finally, my parents, my sister, and all my extended family back home. If I had an ounce of ‘stickability’, perseverance, and grit, it was because of what I observed in my father and mother while growing up in a lower-middle-class household in a foreign country. They had provided me and my sister with all the joys in the world. I am especially proud to have a strong sister - Asma Anwar. She may be younger than me but I have always looked up to her. I rejoice now at the sight of my nieces - Alina and Alisha - every day, seeing them grow up and take the first steps makes it all worth it.

Mohd Anisul Karim
St Edmund Hall, Oxford

Declaration

This thesis is submitted for the degree of Doctor of Philosophy in Population Health at the University of Oxford. No part of this thesis has been accepted or is currently being submitted for any degree, diploma, certificate or other qualification in this University or elsewhere. This thesis is my own work, except where stated.

Mohd Anisul Karim
St Edmund Hall, Oxford
June 2021

My COVID-19 Log

Event date	Brief description	Level of impact	Additional details
4th April 2020	Family member affected	Low	
10th April 2020	Family member affected	Low	
8th May 2020	Family members affected	Medium	
10th May 2020	Family members affected	Medium	
10th June 2020	Family member affected	High	1 family member passed away
24th July 2020	Family member affected	High	
10th Oct 2020	Family member affected	High	Admitted to ICU

Abstract

Cardiovascular disease (CVD) is a major cause of premature deaths and disability globally. Inflammation, the response of the immune system to damaging stimuli, was reported to contribute to CVD risk, with key inflammation-associated proteins (e.g. interleukin-6 [IL-6] signalling) targeted in recent clinical trials. I investigated whether components (IL-6 and IL-6 receptor [IL-6R]) and markers (C-reactive protein [CRP] and fibrinogen) of the IL-6 signalling were associated with CVD risk, with the aim of identifying relevant therapeutic targets for CVD.

First, using a nested case-control study design in the prospective China Kadoorie Biobank (CKB), I reported that in Chinese adults higher plasma CRP concentrations, but not plasma fibrinogen levels, were log-linearly associated with higher risks of major coronary events, ischaemic stroke (IS), and intracerebral haemorrhage (ICH). However, prior genetic studies have reported null associations of plasma CRP-associated genetic variants with CVD risk, precluding CRP as a therapeutic target for CVD.

To investigate potential therapeutic targets, I used genome-wide association studies of CRP in CKB and in Europeans to derive genetic instruments representing the IL-6R and the IL-6 ligand. The IL-6R instrument was a non-synonymous genetic mutation in the *IL6R* locus (rs2228145-C) and reportedly mimicked the effects of tocilizumab (an IL-6R antagonist). The IL-6 instrument was composed of three genetic variants in CKB and a single genetic variant in Europeans (rs1880241-G) near the *IL6* locus. The CKB/EUR IL-6 genetic instruments were, therefore, constructed to broadly represent effects of lower IL-6 signalling. Using a two-sample Mendelian randomization approach, I showed that although the IL-6R instrument, associated with lower plasma CRP, was associated with a lower risk of coronary heart disease (CHD) phenotypes in both CKB and Europeans, there was no evidence of an association of the IL-6 instrument with CHD in either population. Additionally, on trans-ethnic meta-analysis, there was weak evidence that the IL-6R instrument was associated with a lower risk of IS and a higher ICH risk but with consistent estimates observed with the IL-6 instrument.

In conclusion, although the observational study supported the role of inflammation in CVD, the null association of the IL-6 instrument with CHD in both the Chinese and Europeans with weak evidence of lower and higher risk of IS and ICH, respectively, questions the net therapeutic value of targeting IL-6 signalling for CHD.

Contents

List of Figures	xiv
List of Tables	xvii
List of Abbreviations	xix
1 Introduction	1
1.1 Global burden of premature deaths and disability	1
1.2 Cardiovascular Disease	2
1.2.1 Coronary heart disease	3
1.2.2 Stroke	5
1.3 Risk factors of cardiovascular disease	6
1.3.1 Non-modifiable risk factors	6
1.3.2 Modifiable risk factors	8
1.3.3 Relevance of CVD risk factors to low-grade persistent inflammation	14
1.4 Atherosclerosis and its immunological origins	19
1.4.1 Atherosclerosis	19
1.4.2 Inflammation	20
1.4.3 Role of immune cells in the atherosclerotic process	24
1.4.4 Role of immune cell secreted proteins (‘cytokines’)	25
1.4.5 Role of cytokines in non-atherosclerotic CVD	26

1.5	CRP, Fibrinogen, and Interleukins: components of systemic inflammation	27
1.5.1	Discovery of CRP, fibrinogen and interleukins	27
1.5.2	Observational studies	29
1.5.3	Randomised controlled trials	31
1.5.4	Mendelian randomisation studies	37
1.5.5	Interleukin-6 signalling and its components	44
1.6	Evidence and knowledge gaps	46
1.7	Thesis Aims & Objectives	49
2	Data	51
2.1	Non-CKB data	52
2.1.1	Data sources	52
2.1.2	CVD outcome definitions in non-CKB data sources	53
2.2	CKB data	54
2.2.1	Study population	54
2.2.2	Data Collection	55
2.2.3	Follow up for fatal and non-fatal outcome	55
2.2.4	Measurement of biomarkers	56
2.3	Statistical analysis plan	57
3	CRP, Fibrinogen and Cardiovascular Disease	59
3.1	Introduction	59
3.2	Methods	62
3.2.1	Participant selection for the nested case-control study	62
3.2.2	Clinical chemistry measurements	62
3.2.3	Case-control definitions for the present study	63

3.2.4	Participants for the present study	64
3.2.5	Statistical Analysis	67
3.3	Results	68
3.3.1	Baseline characteristics of participants	68
3.3.2	Cross-sectional associations of CRP and fibrinogen with selected CVD risk factors	72
3.3.3	Associations of CRP and fibrinogen with CVD	81
3.3.4	Prespecified subgroup analyses of C-reactive protein and fibrinogen	92
3.4	Discussion	96
3.4.1	Associations of CRP and fibrinogen with baseline traits	96
3.4.2	Associations of CRP and fibrinogen with CVD outcomes	100
3.4.3	Strength and limitations	104
3.5	Conclusion	106
4	Interleukin-6 signalling and Cardiovascular Disease	107
4.1	Introduction	107
4.2	Methods	111
4.2.1	Study population	111
4.2.2	Selection of cases and controls	116
4.2.3	Statistical analysis	116
4.3	Results	119
4.3.1	Selection of SNP to construct the IL-6R genetic instrument	119
4.3.2	Selection of SNPs to construct the IL-6 genetic instrument	120
4.3.3	Association of the IL-6(R) instruments with CRP and IL-6 protein	124
4.3.4	Summary and interpretation of genetic instruments selected	127

4.3.5	Baseline characteristics of study participants	127
4.3.6	Associations of interleukin-6 signalling with cardiovascular diseases	132
4.3.7	Associations of interleukin-6 signalling with CVD risk factors	137
4.3.8	Associations of interleukin-6 signalling with lipids	140
4.3.9	Pre-specified subgroup analysis of the IL-6(R) instruments with CVD	143
4.4	Discussion	146
4.4.1	Genetic instruments representing interleukin-6 signalling . .	146
4.4.2	Interleukin-6 signalling and coronary heart disease	149
4.4.3	Interleukin-6 signalling and stroke types	150
4.4.4	Interleukin-6 signalling and cardiovascular disease risk factors	152
4.4.5	Interleukin-6 signalling and lipids	153
4.4.6	Strength and limitations	155
4.5	Conclusion	161
5	Phenome-wide associations of interleukin-6 signalling	162
5.1	Introduction	162
5.2	Methods	164
5.2.1	Selection of cases and controls	164
5.3	Results	166
5.3.1	Associations of the IL-6(R) instruments with prespecified non-vascular diseases in CKB	166
5.3.2	Associations of the IL-6(R) instruments with clinical chemistry measurements in CKB	169
5.3.3	Proteomic associations of IL-6/IL-6R instruments	174
5.3.4	Exploratory associations of IL-6R/IL-6 instrument with non-prespecified disease outcomes in CKB	176

5.4	Discussion	178
5.4.1	Interleukin-6 signalling and cancers	179
5.4.2	Interleukin-6 signalling and prespecified non-cancer outcomes	181
5.4.3	Interleukin-6 signalling and blood biomarkers	182
5.4.4	Phenome-wide and proteomic associations of interleukin-6 signalling	183
5.4.5	Strength and limitations	183
5.5	Conclusion	185
6	Discussion	186
6.1	CRP, Fibrinogen, and CVD outcomes	186
6.2	Interleukin-6 signalling and CVD outcomes	189
6.3	Phenome-wide associations of interleukin-6 signalling	195
6.4	Research implications	196
Appendices		
A	APPENDIX	200
A.1	Supplemental contents for Chapter 3	201
A.2	Supplemental contents for Chapter 4	203
A.3	Supplemental contents for Chapter 5	209
REFERENCES		214

List of Figures

1.1	Global Burden of Cardiovascular Disease	2
1.2	Central role of immune cells and their secreted proteins in the pathogenesis of atherosclerosis	23
1.3	Effect of statins on risk of recurrent vascular events in different patient groups	35
1.4	Comparison of cumulative risk of recurrent CVD outcomes in the CANTOS and CIRT studies	36
1.5	Comparison of study designs evaluating causality	38
1.6	Mendelian randomisation concept	39
1.7	Mendelian randomisation using proteins as exposure traits	41
1.8	Drug discovery and development pipeline	42
1.9	Different modes of interleukin-6 (IL-6) signalling	45
3.1	Flowchart of participant selection	66
3.2	Cross-sectional associations of C-reactive protein with selected traits	74
3.3	Cross-sectional associations of CRP and Fibrinogen with biomarkers	77
3.4	Risk of Major Coronary Events, Ischaemic Stroke, and Intracerebral haemorrhage by quintiles of C-reactive protein (CRP) and fibrinogen	84
3.5	Risk of CVD outcomes per standard deviation (SD) of usual log-transformed high-sensitivity plasma CRP	87
3.6	Risk of CVD outcomes per standard deviation (SD) of usual log-transformed plasma fibrinogen	88

3.7	Progressive adjustment of CRP and Fibrinogen for Major Coronary Events, Ischaemic Stroke and Intracerebral haemorrhage	91
3.8	Risk of CVD outcomes per standard deviation (SD) of usual log-transformed plasma CRP by pre-specified subgroups	94
3.9	Risk of CVD outcomes per standard deviation (SD) of usual log-transformed plasma fibrinogen by pre-specified subgroups	95
4.1	Drug-target Mendelian randomisation	111
4.2	Summary of genotyped participants in CKB, genotyped with 384-SNP panel or 800k SNP array	113
4.3	Selection of <i>IL6</i> SNPs	122
4.4	Associations of the IL-6(R) instruments with markers of inflammation in CKB	125
4.5	Per-allele associations of the IL-6(R) instruments with C-reactive protein	126
4.6	Trans-ethnic association of the IL-6(R) instruments with cardiovascular disease traits and outcomes	136
4.7	Trans-ethnic association of the IL-6(R) instruments with selected cardiovascular disease risk factors	139
4.8	Trans-ethnic association of the IL-6(R) instruments with lipids	142
4.9	Subgroup analysis of the IL-6R instrument with cardiovascular disease outcomes	144
4.10	Subgroup analysis of the IL-6 instrument with cardiovascular disease outcomes	145
5.1	Association of the IL-6R and IL-6 instruments with prespecified non-vascular outcomes	168
5.2	Associations of the IL-6(R) instruments with additional clinical chemistry measurements	173
5.3	Proteomics associations of the IL-6R instrument	175
5.4	Proteomics associations of the IL-6 instrument	175

5.5	Phenome-wide associations of the IL-6R and the IL-6 instruments	177
A.1	Distributions of biomarkers studied in this thesis before and after log-transformation	201
A.2	Associations of CRP and fibrinogen with MCE, IS and ICH on adjustments for individual risk factors	202
A.3	LosusZoom plots of inflammation loci from CRP GWAS	204
A.4	Association of the IL-6(R) instruments and <i>IL6</i> regulatory SNP (rs13306436-A) with CRP and IL-6 in all and random subsets	205
A.5	Association of <i>IL6</i> SNPs with vascular and non-vascular outcomes	206
A.6	Association of <i>IL6</i> SNP rs13306436 with CVD traits and outcomes	207
A.7	Associations of <i>IL6R</i> , <i>IL6</i> gene score with lipidomic traits	208
A.8	Association of <i>IL6</i> SNP rs13306436 with non-vascular outcomes	210
A.9	Proteome-wide associations of the IL-6(R) instruments	211
A.10	<i>IL6R</i> and <i>IL6</i> gene expression across post-mortem tissues (GTEx v8)	212
A.11	Effect per rs7529229-C (EUR/CKB IL-6R instrument) on <i>IL6R</i> gene expression across tissues (GTEx v8)	213

List of Tables

1.1	Evidence from observational studies: Association of CRP, fibrinogen and interleukin-6 with CVD outcomes	30
1.2	Evidence from Mendelian randomisation studies: Association of CRP, fibrinogen and interleukin-6 with CVD outcomes	49
2.1	Selected outcomes from non-CKB data sources for trans-ethnic meta-analysis	52
3.1	Clinical chemistry measurements in China Kadoorie Biobank	63
3.2	Case-control definitions	64
3.3	Baseline characteristics of study participants	71
3.4	Geometric mean and 95% confidence interval (CI) of C-reactive protein (CRP) and Fibrinogen by key baseline characteristics	80
3.5	Evidence from observational studies: Association of CRP and fibrinogen with CVD outcomes	101
4.1	Case-control selection for <i>IL6R</i> Mendelian randomisation analyses .	115
4.2	Candidate SNPs in <i>IL6R</i> locus associated with CRP in CKB	120
4.3	Multivariable regression models of first 6 fine-mapped <i>IL6</i> SNP sets from FINEMAP	123
4.4	Comparison of IL6R and IL6 Instruments between CKB and UKB .	124
4.5	Summary of the IL-6(R) instruments for CKB and Europeans	127
4.6	Baseline traits of the study population	130

4.7	Baseline traits of the study population according to genotype subgroups of IL6R and tertiles of IL6 gene score	131
5.1	Summary of comparisons of clinical chemistry measurements between effect of tocilizumab and IL-6R/IL-6 instruments	172
5.2	Look-up and comparison of IL-6R instrument (rs7529229-C) signal in other colorectal cancer case-control cohorts	180
6.1	Summary of trans-ethnic (East Asians and Europeans) and inter-instrument (IL6(R)) consistency of associations	194
A.1	Tissue-specific associations of rs1880241-G (EUR IL-6 instrument) on <i>IL6</i> gene expression	203
A.2	Power estimates for primary vascular endpoints for IL6R analyses .	203
A.3	Power estimates for primary vascular endpoints for IL6 analyses . .	203
A.4	Power estimates for pre-specified non-vascular endpoints with an $\alpha = 0.05/10$ (10 is the number of secondary non-vascular endpoints) .	209

List of Abbreviations

BMI		Body mass index.
CHD		Coronary heart disease
CI		Confidence interval
CKB		China Kadoorie Biobank
hsCRP		high-sensitivity C-reactive protein
CVD		Cardiovascular disease
HDL-C		High-density lipoprotein cholesterol
HS		Haemorrhagic stroke
ICH		Intracerebral haemorrhage
IS		Ischaemic stroke
IL-1β		Interleukin-1 beta
IL-6		Interleukin-6
LRT		Likelihood ratio test
LDL-C		Low-density lipoprotein cholesterol
MR		Mendelian randomisation
MI		Myocardial infarction
RCT		Randomised controlled trial
SBP		Systolic blood pressure
TNF-α		Tumour necrosis factor-alpha
TG		Triglycerides.

1

Introduction

1.1 Global burden of premature deaths and disability

According to the Global Burden of Disease (GBD) 2019, there were 56.5 million deaths recorded globally in 2019¹. Of the 56.5 million deaths, 42 million deaths were due to non-communicable diseases (NCDs). Cardiovascular disease (CVD), musculoskeletal and mental disorders, and cancer were among the largest contributors to deaths due to NCDs; of these, deaths due to CVD comprised the majority of NCD-related deaths (18.6 million out of 42 million NCD-related deaths). Cardiovascular diseases are also a major cause of global morbidity. A metric used by GBD to quantify the number of years lost due to morbidity, disability or premature death is the disability-adjusted life years (DALYs). Cardiovascular diseases were among the largest contributors to DALYs, accounting for 393 million DALYs out of 1.62 billion DALYs due to NCDs in 2019 (Figure 1.1), representing an increase of 12.5% since 2010. The increase in global DALY rates due to CVD does not necessarily suggest an absolute increase in global morbidity due to CVD. The higher CVD DALY rates may be due to global population growth and longer average life expectancies. Hence, when DALY rates are standardised for age, age-standardised DALY (ASDALY) rates

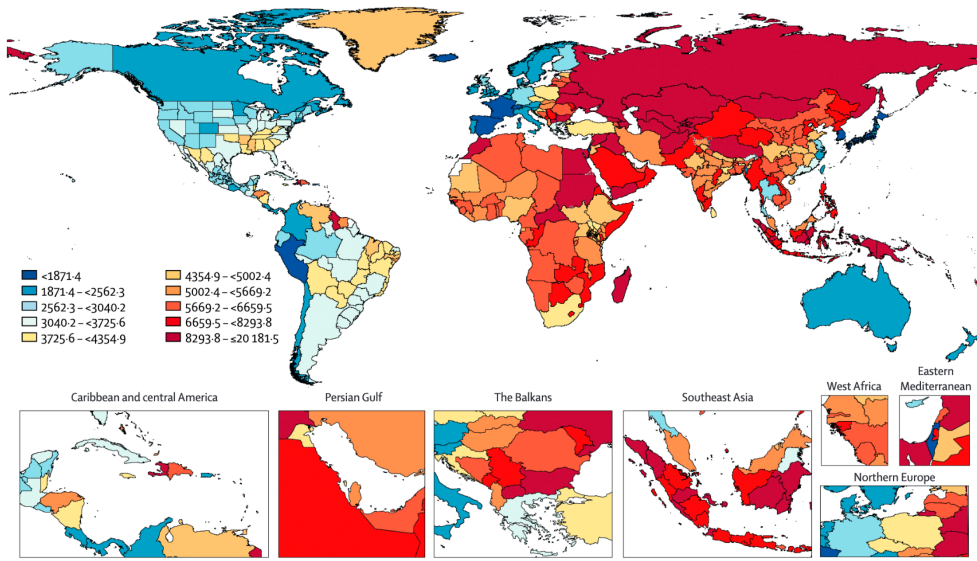


Figure 1.1: Global Burden of Cardiovascular Disease. A global map of age-standardized disability-adjusted life years due to CVD in 2019, showing the disproportionate effects of CVD in mostly low-and-middle-income countries. Reproduced from Vos et al [1]

for CVD since 2010 were reported to have actually declined by 11.2%. The higher life expectancy is likely primarily due to advances in secondary prevention measures for CVD (i.e. better therapies and therapeutic procedures after the first CVD event) and this is evidenced by the disproportionate ASDALYs in developing countries compared to developed nations due to CVD in 2019 (Figure 1.1). Nevertheless, although the global CVD ASDALY rates have reduced over the years, CVD remains the largest cause of premature deaths in the world. Additionally, the steady population growth and higher proportions of the older population forecasted for the next few decades is likely to translate to higher absolute burdens of CVD in the coming years and, therefore, warrants further research to advance our understanding of effective therapies.

1.2 Cardiovascular Disease

The World Health Organisation (WHO) defines CVD as a group of disorders that affects the heart or the circulatory system². The major groups that comprise

CVD include coronary heart disease (CHD), peripheral arterial disease (PAD) and cerebrovascular disease. The underlying pathology in many of these CVD groups is a restriction of blood supply to their respective organs, such as the heart (CHD), limbs (PAD), or the brain (cerebrovascular disease). In most cases, the restricted circulation results from internal occlusion of the lumen of blood vessels due to progressive growth inside the lining of the vessels of a plaque composed of fibrous and fatty tissue (Section 1.4). The abnormal growth of the fibrofatty plaque in blood vessels of the respective organs is slow and undulating, with affected individuals reporting symptoms spanning from mild to progressively severe depending on the extent of the plaque growth³. When blood circulation supporting the heart (via coronary blood vessels) is affected, mild symptoms include chest pain on physical exertion that is relieved at rest (stable angina) that, if untreated, becomes progressively severe enough to remain even at rest (unstable angina)⁴. When blood supply is restricted in brain tissues, it can initially lead to only a temporary (typically less than 24 hours) focal disturbance of brain function characterised by blindness or limb weakness affecting one-half of the body (transient ischaemic attack). In both the heart and brain, when the plaque is ruptured (usually from high-pressure turbulent blood flow), it typically leads to clot formation at the site of rupture. The resulting clot can potentially occlude blood vessels causing permanent damage and death (necrosis) of underlying tissues; when this occurs in the heart, it causes myocardial infarction (MI) and, in the brain, it leads to a stroke. Among CVD groups, CHD and stroke were the top two leading causes of global mortality and morbidity in individuals over 50 years of age in 2019¹.

1.2.1 Coronary heart disease

There are several terms used interchangeably in the literature to refer to symptoms and signs arising from blockage of arteries around the heart (coronary arteries), such as coronary artery disease (CAD), CHD, ischaemic heart disease (IHD), and acute coronary syndromes (ACS). However, these terms have distinct meanings⁴. Both CHD and IHD are near-equivalent terms and collectively refer to the clinical signs

and symptoms that arise from the pathological occlusion of coronary arteries. When the occlusion inside coronary vessels is confirmed by imaging (e.g. by computed tomography angiography), it is referred to as CAD. If the occlusion is sufficient to cause necrosis of heart muscles, the acute symptoms that result are one of the acute coronary syndromes, either unstable angina or acute MI⁵. These acute symptoms can be classified as obstructive CAD, non-obstructive CAD (NOCAD), or coronary microvascular dysfunction (CMD)⁶. The extent of stenosis or narrowing in the coronary arteries required to angiographically distinguish between obstructive and non-obstructive CAD varies between different guidelines or studies⁷. The most recent American College of Cardiology/American Heart Association guidelines considers stenosis in the coronary arteries to be significant when there is an angiographically-determined luminal narrowing of more than 70% of any coronary artery or more than 50% narrowing of the left main coronary artery or a 40% to 70% narrowing of any coronary artery if there is associated inducible ischaemia⁸. If the coronary artery narrowing is not significant stenosis, it is generally classified as non-obstructive CAD. Both NOCAD and CMD can present with symptoms of angina (i.e. chest pain with or without exertion) and can potentially underlie the persistence of ischaemic symptoms even after coronary revascularisation for obstructive CAD⁶.

In 2019, CHD (mostly representing obstructive CAD) resulted in approximately 9 million global deaths, contributing 182 million DALYs to global DALYs, representing an increase of 50% since 1990¹. Prevalence of non-obstructive CAD varies between studies, with one study reporting 16% to 22% of patients undergoing coronary angiography had no obstructive CAD⁹. The prevalence of NOCAD was reported to be higher in women (65%) than men (32%)¹⁰. In both men and women, almost 60% of patients with NOCAD had evidence of CMD (defined as endothelium-dependent or endothelium-independent microvascular dysfunction)¹¹. Irrespective of the diagnosis of specific CAD type, patients with NOCAD or CMD are at a higher risk of obstructive CAD, major adverse cardiovascular events (MACE), heart failure with preserved ejection fraction, and ACS^{7,10}.

1.2.2 Stroke

Stroke is defined by WHO as “a clinical syndrome consisting of rapidly developing clinical signs of focal (or global) disturbance of cerebral function, lasting more than 24 hours or leading to death, with no apparent cause other than that of vascular origin”¹². About 6 million deaths in 2019 were attributed to stroke representing a contribution of 143 million DALYs to global DALYs¹. Depending on the underlying pathology (occlusion of cerebral arteries or bleeding in brain tissues), stroke is classified as either ischaemic or haemorrhagic stroke type. Ischaemic stroke (IS) describes the occlusive pathology that occurs inside blood vessels supplying the brain and results in signs and symptoms of focal brain disturbances that arise from the occlusion. Haemorrhagic stroke (HS) occurs due to rupture of weakened or malformed blood vessels or iatrogenically induced thinning of the blood (e.g. inappropriate use of anticoagulants). While anticoagulant therapy may be instituted in IS, it is contraindicated in HS. The determination of whether a stroke is ischaemic or haemorrhagic, therefore, is crucial as a treatment for one can exacerbate the other. The determination of IS or HS is also necessary from a prognostic point of view. Although IS is the prevalent stroke type (~80%¹), IS generally has a better prognosis than HS¹³ with lower case-fatality rates (CFR) than ICH - the most common subtype of HS (CFR: 15-20% for IS¹⁴ vs. 40% for ICH in the first month¹⁵). The prognosis of SAH - the less common type of HS caused commonly by a ruptured malformed vessel - is similar to ICH with CFRs between 32% and 67%¹⁶.

Therapeutic and prognostic differences also exist within IS stroke subtypes. The Trial of Org 10172 in Acute Stroke Treatment (TOAST)¹⁷ classified IS into five IS subtypes: 1) large-artery atherosclerosis, 2) cardioembolism, 3) small-vessel occlusion, 4) stroke of other determined aetiology, and 5) stroke of undetermined aetiology. Of these, recurrent strokes with higher mortality rates were observed frequently in cardioembolic stroke (where a clot from the heart lodges and occludes cerebral arteries) than other IS subtypes¹⁷. The therapeutic modalities may also

differ depending on the IS subtype. For example, while long-term anticoagulation is the preferred mode of treatment for cardioembolic stroke¹⁸, carotid endarterectomy (the surgical opening and removal of plaque from carotid arteries) is recommended for the management of large-artery atherosclerosis where one or both the carotid arteries are narrowed by plaque deposition¹⁹.

Both IS and ICH had DALY contributions of 64 million and 69 million, respectively, to the 143 million global DALYs accrued due to stroke in 2019. Despite declines in CVD ASDALY rates over the years, both CHD and stroke remained the top two causes of premature death globally in 2019¹. Identification and further mitigation of risk factors of CHD and stroke remain the cornerstone of the prevention of CVD and its complications.

1.3 Risk factors of cardiovascular disease

A number of modifiable and non-modifiable risk factors contribute to the risk of CVD. The non-modifiable factors include age, male sex, and ethnicity.

1.3.1 Non-modifiable risk factors

Aging

As described in Chapter 1.1, the ASDALYs due to CVD have declined since the 1990s and, therefore, much of the attributed increase of DALYs due to CVD since 1990 have been due to population growth and ageing. The association of non-modifiable factors to CVD is due, at least in part, to the correlation of non-modifiable factors to potentially modifiable CVD risk factors. Ageing is a good example of this model as the higher CVD risk due to ageing can potentially be the result of the cumulative exposure to modifiable CVD risk factors over time²⁰. This was shown to be the case in the Framingham Heart Study which examined differences in lifetime risk at age 50 by modifiable risk factor burden. The study demonstrated substantially low lifetime risks of CVD and longer median survival in both men and women at age 50 with the lowest risk factor burden²¹.

Sex

Similar to age, the higher CVD risk in males than females can be partially accounted for by differences in modifiable risk factor burden between the sexes (for e.g. males, on average, tend to drink and smoke more than females²²). However, whether biological sex is an independent risk factor remains unclear. Several lines of epidemiological evidence suggest that men tend to have CVD events at younger ages with higher mortality rates due to CVD than females^{23–25}. It was postulated that this sex-specific difference in CVD risk was due to a protective effect of female hormones prior to menopause but there is limited epidemiological and experimental evidence to support this notion²⁶. A more recent study showed that mortality due to CHD in women increased exponentially with age, with no evidence of acceleration during menopause²⁷. With men, the investigators observed a higher risk of death due to CHD in younger ages with the deceleration of risk after 45 years of age compared to age-matched women²⁷. Indeed, at age 55, men and women tended to have similar lifetime risks of CVD²⁴, further weakening the case of female hormones playing a cardioprotective role.

Ethnicity

The trans-ethnic differences in CVD risk (i.e. a higher CVD risk in blacks and South Asians than the Europeans) may, at least partially, stem from ethnic disparities in the burden of potentially modifiable social risk factors of socio-economic status, childhood nutrition, and equitable access to quality healthcare²⁸. Like sex, whether ethnicity is an independent risk factor for CVD remains widely debated. It was observed by several investigators that migrants of South Asian descent in the UK and African Caribbeans were at higher risk of CHD²⁹ and stroke³⁰ than their White-ancestry European peers. Disparities of equitable access to healthcare were postulated to explain part of the differences in risk; for example, among patients where revascularisation surgery was indicated for obstructive CAD, a study showed that South Asian patients were less likely to receive revascularisation surgery than white patients³¹. Large-scale studies that have systematically examined inequities

in healthcare access based on ethnic groups are scarce. Several other studies have postulated metabolic derangements in South Asians such as insulin resistance³² and higher levels of lipoprotein (a)³³ to explain the higher CVD risks. Markers of systemic inflammation such as C-reactive protein (CRP)³⁴ and fibrinogen³⁵ were also found to be disproportionately elevated in South Asians than Europeans. South Asians, Blacks, and other non-European ethnic groups are currently under-represented in genetic studies³⁶, and hence, evidence of genetic predisposition to CVD risk of one ethnic group over another is limited. The scarcity of non-European genetic studies also greatly limits investigations on mechanisms that would confer higher CVD risk to South Asians or Blacks compared to Europeans. However, efforts to increase genetic diversity in GWAS are currently underway³⁷, and the current thesis, which aims to understand the relevance of inflammation to CVD risks in the Chinese, represents a small step towards the goal of inclusive genetic epidemiological investigations.

1.3.2 Modifiable risk factors

Several modifiable risk factors were identified and extensively studied over the years (Box 1.1)³⁸. This section briefly describes the epidemiological evidence for the associations of modifiable risk factors with CVD and then discusses the biological relevance of these risk factors to inflammation.

Blood pressure

The pressure of blood in the circulatory system rises (systolic blood pressure, SBP) and falls (diastolic blood pressure, DBP) on each heartbeat and is commonly referred to as blood pressure. There is strong epidemiological evidence that higher SBP and higher DBP predispose to higher risks of most CVD phenotypes in both Europeans^{39,40} and East Asians⁴¹. The Prospective Studies Collaboration meta-analysis of 61 European cohorts (aged between 40 and 89 years of age) reported log-linear associations of usual SBP and DBP with risks of CVD mortality, without any evidence of a threshold down to an SBP/DBP of 115/75 mm Hg. The lifetime

risk of CVD (a measure of absolute risk representing the probability that an individual will develop CVD over the course of their lifetime) was reported to be 17.2% higher in individuals at 30 years of age with hypertension (i.e. blood pressure more than or equal to 140/90 mm Hg) compared to age-matched individuals without hypertension⁴⁰. Hence, lowering blood pressure is associated with a reduced risk of mortality from CVD and the relative risk reductions are proportional to the magnitude of the BP reductions achieved⁴². A meta-analysis of 123 trials with 613,815 participants showed that every 10 mm Hg reduction in SBP was associated with approximately 20% to 30% relative risk reductions of major CVD events, CHD, stroke, and heart failure⁴².

Obesity

Obesity is defined by WHO as abnormal or excessive fat accumulation that may impair health⁴³. Body-mass index (BMI, weight in kilograms/(height in metres)²) is frequently used to quantify obesity. Individuals with BMI between 25.0 and 29.9 kg/m² are classified as overweight, and a BMI over 30.0 kg/m² is considered obese⁴⁴. Analyses in 57 prospective studies with 894,576 participants (mostly Europeans, 35-89 years of age) showed higher risks of death due to CHD or stroke with higher BMI; for each 5 kg/m² higher BMI than baseline BMI associated with a higher relative risk of CHD or stroke by approximately 40% (HR: 1.41 [95% CI: 1.37, 1.45])⁴⁵. A similar association of BMI with CHD was also observed in East Asians, with one large prospective cohort (n = 222,000 Chinese men, 40-79 years of age) reporting that each 2 kg/m² higher usual BMI, over 20 kg/m², was associated with a 12% (95% CI: 6%,19%) higher CHD mortality⁴⁶. In the same prospective East Asian cohort but with a larger sample size (n = 512,891 adults, aged 30-79 years), the investigators showed that each 5 kg/m² higher BMI was associated with a 30% higher risk of IS (HR: 1.30 [95% CI 1.28-1.33]) and 11% higher risk of ICH (1.11 [1.07,1.16])⁴⁷. However, the associations of BMI with IS in East Asians were substantially attenuated on adjustment for usual SBP (HR: 1.05 [1.03-1.07] per 5 kg/m² higher BMI)⁴⁷, with associations of BMI with ICH reversed (i.e. higher BMI

was associated with lower ICH) on statistical adjustment of SBP. The association of BMI with stroke types pre- and post-adjustment for SBP suggests that at least part of the risk of BMI on IS and ICH may be potentially mediated via the effect of BMI on SBP.

Blood Cholesterol

The Framingham Heart Study provided the first prospective evidence of elevated blood total cholesterol levels and CHD risk in a cohort of approximately 5000 men and women⁴⁸. The Prospective Studies Collaboration subsequently combined 61 prospective observational studies comprising of almost 900,000 adults (mostly Europeans) between the ages of 40 and 89 years and showed that each mmol/L lower total cholesterol was associated with lower risk of CHD and the risk was lowest in both sexes at ages 40–49 years (HR: 0.44 [95% CI: 0.42, 0.48])⁴⁹. However, total cholesterol is a composite trait and represents the sum of cholesterol concentrations in different lipoprotein particles. Two of the major lipoproteins studied in relation to CVD risk are low-density lipoprotein (LDL, which transport cholesterol as LDL-C to peripheral tissues) and high-density lipoprotein (HDL, which carry cholesterol as HDL-C away from tissues to transport them back to the liver for excretion). Several epidemiological studies have provided strong evidence supporting the causal role of LDL-C as a risk factor for CHD and stroke^{50–54}. The specific role of HDL-C remains unclear. Higher plasma HDL-C was observationally associated with lower CHD risk⁵⁵, but there was no evidence that genetically predicted plasma HDL-C was associated with CHD risk⁵¹. However, both genetically predicted plasma LDL-C and HDL-C were associated with stroke types and subtypes. Genetically predicted lower plasma LDL-C was associated with a lower risk of IS⁵⁴, specifically the large-artery atherosclerotic subtype of IS⁵⁶ but higher risk of ICH⁵⁴. Genetically predicted lower plasma HDL-C was only associated with a lower risk of small-artery occlusion⁵⁶. Neither genetically predicted plasma LDL-C nor plasma HDL-C were, however, associated with cardioembolic stroke⁵⁶.

Blood glucose

Epidemiological studies have indicated that irrespective of diabetes, both fasting and postprandial glucose levels were associated with CVD risk^{57–59}. For random blood glucose, while a preliminary study showed largely neutral associations with CVD risk⁶⁰, a larger prospective investigation using a large East Asian cohort showed a log-linear relationship of elevated random plasma glucose levels and risk of heart disease and stroke^{61,62}. The mechanism linking excessive plasma glucose with CVD risks is unclear. There is some experimental evidence to suggest a role of advanced glycation end-products (AGE, these are products of chemical transformations of amine-containing molecules, e.g. proteins, lipids or nucleotides, by reducing sugars such as glucose)^{63,64}; AGE products are triggered by high oxidative stress that typically results from long-term blood glucose elevations⁶⁵ and is reported to contribute to both the aetiology and complications (e.g. CVD⁶⁶) of diabetes⁶⁷.

Smoking and Alcohol

Smoking and alcohol are among the major lifestyle factors associated with CVD risk. An extensive body of studies spanning several decades has provided strong evidence for the aetiological role of cigarette smoking in CHD, stroke, aortic aneurysm, and peripheral arterial disease^{68–72}. The observations of association of smoking with CVD risks are not limited to smokers but also extend to individuals exposed to secondhand smoke⁷³. Several mechanisms were postulated linking tobacco smoke with CVD, including sustained elevations of heart rate and blood pressure, derangements of lipid metabolism, endothelial dysfunction, and low-grade persistent chronic inflammation⁷⁴. Unlike smoking which had strong support from conventional epidemiological studies that linked smoking with CVD risk, the interpretation of the association of alcohol with CVD risk using traditional epidemiological study designs was not as straightforward. While observational studies have shown a U-shaped or J-shaped association of alcohol consumption with CVD risks, genotype-predicted alcohol consumption showed clear log-linear associations of higher alcohol intake

with a higher risk of most CVD phenotypes with no evidence of a lower safe threshold of drinking^{75–77}.

Diet

According to the GBD 2019 risk factors collaboration, among risk factors with the highest attributable deaths in 2019, dietary risks ranked 2nd and 3rd for women and men, respectively⁷⁸. In absolute numbers, dietary risks were responsible for 11 million deaths globally in 2019, of which 10 million deaths were due to CVD⁷⁹. Specific diets considered as risk factors for CVD by the GBD study include diets low in fruits and vegetables, and high in red meat, processed meat, sodium, sugar, and trans-fats. These dietary risk factors collectively referred to as ‘Western’ pattern diet, were nominated by the GBD investigators because there was extensive evidence from several epidemiological studies that implicated these diets as major risk factors of CVD^{80–83}. Many of these diets mediate CVD risk through the aforementioned risk factors. For example, high salt intake is linked to high blood pressure which contributes to CVD risk. Consumption of trans-fats and sugar is linked to obesity and elevated plasma glucose levels, both of which are well-established CVD risk factors.

Air pollution

In 2015, approximately nine million premature deaths were due to diseases linked to air pollution⁸⁴. Of these nine million premature deaths, CVD deaths accounted for 21%, with 26% of CVD deaths due to CHD and 23% due to stroke⁸⁴. The adverse CVD effects of air pollution are not limited to ambient air pollution; household air pollution due to (self-reported) burning of solid fuels were also associated with higher risks of CVD mortality^{85,86}.

Environmental enteropathy

Environmental enteric dysfunction (EED) is a subclinical intestinal abnormality that is reported to be prevalent among children growing in impoverished resource-limited areas of developing countries⁸⁷. EED in children is postulated to result from the

lack of access to clean water and inadequate sanitation⁸⁷. As a consequence, there is constant exposure to environmental non-commensal pathogens which disrupts the absorption of essential nutrients from the gut lining, accelerates low-grade chronic intestinal inflammation and causes childhood growth faltering⁸⁸. EED in early life is increasingly being recognised as an emerging risk factor for CVD in adulthood^{88–90}.

Low-grade persistent inflammation

Inflammation represents the response of the immune system to foreign antigens and tissue damage in living tissues⁹¹ (Section 1.4.2). Markers of dysregulated systemic inflammation, plasma C-reactive protein (CRP) and fibrinogen were observationally associated with risks of heart disease and stroke^{92,93}. Although the observational associations of plasma CRP and fibrinogen with CVD risks were not observed with genetically predicted plasma concentrations of CRP and fibrinogen^{94,95}, the genotype-predicted concentration of soluble interleukin-6 receptor (sIL-6R) which associated with circulating CRP and fibrinogen levels were associated with risks of CHD^{96,97}. The observational associations of plasma CRP and fibrinogen along with the roles of components of IL-6 signalling to CVD, are among the major themes of my thesis and discussed in-depth in subsequent sections and the next few chapters.

Box 1.1: Modifiable risk factors for cardiovascular disease (CVD)**Established CVD risk factors**

- Smoking [98]
- Alcohol [76, 99]
- Adiposity [47]
- 'Western'-style diet [80]
- Sedentary lifestyle [100]
- Cholesterol [101]
- Blood Glucose [61]
- Blood Pressure [102]

Emerging CVD risk factors

- Low-grade persistent inflammation [103, 104]
- Air pollution [105]
- Environmental enteropathy [106]

1.3.3 Relevance of CVD risk factors to low-grade persistent inflammation

Dysregulated inflammation that persists at low-grade underlies the mechanisms for many of the aforementioned well-established modifiable and non-modifiable CVD risk factors.

Aging and sex

Ageing is associated with the development of a low-grade persistent inflammatory state¹⁰⁷. Age-associated increases in persistent systemic inflammation are associated with higher CVD risk¹⁰⁸. Using parental lifespan as a proxy for the longevity of study participants, a genome-wide association study (GWAS) in Europeans of parental life-span found a number of GW significant signals in loci harbouring genes associated with inflammation¹⁰⁹. The precise mechanisms or inflammation pathways associated with aging remains unclear but there is some emerging experimental evidence to suggest a role of the nucleotide-binding oligomerization domain (NOD)-like receptor family pyrin domain containing 3 (NLRP3) inflammasome^{110–112}. The NLRP3 inflammasome activates a series of cellular inflammation pathways that result in the synthesis of interleukin-1 beta (IL-1 β) and interleukin-6 (IL-6) among other inflammation-associated molecules (these inflammation-associated proteins are described in Section 1.4.2).

Men have a higher risk of mortality due to CVD at younger ages than women (Section 1.3.1). Higher levels of testosterone - the male sex hormone - was

postulated to increase risks of CVD. However, there was conflicting evidence from observational studies for the relevance of testosterone to CVD risk^{113–115} but a genetic epidemiological study supported the role of testosterone in MI¹¹⁶. Further, an experimental study showed altered levels of testosterone was observed to induce reactive oxygen species (ROS) which activates the NLRP3 inflammasome¹¹⁷.

Obesity and Blood pressure

As defined in Section 1.3.2, obesity is the abnormal or excessive accumulation of fat that impairs health. A mechanism linking excessive fat accumulation to suboptimal health is persistent low-grade inflammation characterised by higher plasma CRP concentrations^{118,119}. Evidence to support this obesity-inflammation hypothesis emerged from both conventional observational and genetic epidemiological studies, where the latter also showed it was genetically predicted adiposity that was associated with higher plasma CRP concentrations and not the other way round^{120,121}. Furthermore, a clinical trial to test the effects of a diet- and exercise-induced weight loss program on markers of persistent low-grade inflammation showed lower plasma CRP and plasma IL-6 concentrations in the diet-induced weight loss group compared to the comparison group that only received health education¹²², further supporting the role of obesity in persistent low-grade inflammation. Emerging evidence also suggested *where* fat accumulates is linked to higher inflammation; the fat tissue around organs was reported to have higher numbers of macrophages than other tissues¹²³, secrete more cytokines than subcutaneous adipose tissue¹²⁴, and contribute to the development of atherosclerosis¹²⁵.

Association of persistent low-grade inflammation (represented by plasma concentrations of CRP and IL-6) with blood pressure was reported mainly by conventional observational studies^{126,127}. These preliminary reports were tested using genetically predicted CRP concentrations and the genetic epidemiological study did not find evidence of association of CRP with any BP phenotype (SBP or DBP), suggesting that observational associations of inflammation and BP may have been due to confounding and/or reverse causation¹²⁸. Obesity was, in particular, noted as a

major confounder that distorted the association between inflammation and BP indices¹²⁸. Conversely, whether blood pressure plays a causal role in persistent low-grade inflammation remains unclear. There is some observational evidence suggesting associations of higher BP with plasma IL-6 concentrations¹²⁹. Genetic epidemiological studies to clarify the associations of genetically predicted BP phenotypes on plasma CRP and IL-6 concentrations are warranted.

Smoking and alcohol

Smoking, specifically cigarette smoking, is a known causal risk factor for heart disease and stroke^{69,130}. Among the effects of cigarette smoking reported, are an increase in absolute monocyte count¹³¹, an increase in the adhesion of monocytes to blood vessel walls¹³², the enhanced migration of monocytes in the subendothelial space¹³³ and increases in concentrations of CRP¹³⁴ and IL-6¹³⁵.

With alcohol, conventional observational studies found an inverse relationship between light alcohol consumption and markers of inflammation such as IL-6 and CRP^{136,137}. Genetically predicted alcohol consumption, however, showed that lower alcohol consumption was causally associated with lower levels of IL-6, in contrast to traditional epidemiological studies, and a lower risk of CHD (with no minimum safe threshold of drinking)⁷⁵.

Cholesterol, glucose, and Western diet

Among the CVD risk factors listed in Box 1.1, two of them are biological molecules or biomarkers: cholesterol and glucose. High concentrations of specific types of cholesterol and high glucose levels are strongly associated with CVD risk^{62,101}.

High cholesterol mainly refers to low-density lipoprotein cholesterol (LDL-C) and triglycerides (TG). As illustrated in Figure 1.2, ingestion of lipids is an essential step for macrophages to become foam cells that ultimately form the core of the atherosclerotic plaque. Similarly, dysregulated glucose levels (dysglycaemia), a hallmark of type 2 diabetes, are correlated to CRP concentrations¹³⁸. Both cholesterol and glucose are disproportionately high in the ‘Western pattern’ diet.

The Western diet triggers systemic inflammation via activation of the NLRP3 inflammasome complex⁸⁰ that, as described in Section 1.4.4 and as shown in Figure 1.2, drives the IL-1 β stimulated IL-6 signalling pathway.

Unravelling the causal role of the IL-6 signalling pathway for CVD is, thus, a major aim of the present thesis. Indeed, the different types of balanced diet have been recommended by nutritionists to prevent CVD. Many of these balanced diet recommendations are associated with lower levels of inflammation when consumed on a regular basis. For example, the Mediterranean diet (that consists of fish, fruits, nuts, seed, whole-grains and olive oil) was shown to reduce markers of inflammation including CRP, IL-6, and lower concentrations of adhesion molecules and chemokines - factors that promote the atherosclerotic process (Section 1.4.3)^{139,140} (Figure 1.2). The PREDIMED (Prevención con Dieta Mediterránea) trial that investigated the effect of the Mediterranean diet on CVD, found that compared to a control diet, the Mediterranean diet with extra-virgin olive oil reduced the risk of a major cardiovascular event (myocardial infarction, stroke, or death from cardiovascular causes) by more than 30% (Hazard ratio: 0.69 [95% confidence interval, 0.53 to 0.91])^{82,141}. Such was the importance of inflammation to diet, that researchers constructed a Dietary Inflammatory Index (DII) to assess the inflammatory potential of different diets. The DII score was a literature-derived score that summarized the effect of dietary parameters on six inflammatory biomarkers (IL-1 β , IL-4, IL-6, IL-10, TNF- α and CRP)¹⁴². The Mediterranean diet was shown to be inversely associated with the DII score¹⁴³, i.e. adherence to the diet lowers concentrations of markers of inflammation, including CRP and IL-6¹⁴⁰.

Physical activity

Other lifestyle-based risk factors include sedentary activity or lack of physical activity and lack of sleep. There are a number of studies that have provided clear evidence that regular physical activity lowers CVD risk¹⁴⁴⁻¹⁴⁶. Some of these studies suggested that lower inflammation may be one mechanism mediating the cardio-protective benefits of regular physical activity^{147,148}. Among the various markers of

inflammation assessed with physical activity, levels of CRP and IL-6 stand out; a study that investigated participants over a 10-year follow-up showed that, compared to self-reported physically inactive participants, participants who reported to have engaged in at least 2.5 h/week of moderate to vigorous physical activity displayed lower CRP and lower IL-6 levels¹⁴⁹. The association of optimal sleep to inflammation and CVD is similar to physical activity and balanced diets. Shorter (< 7 hours) or longer sleep (> 9 hours) duration are linked to risk of CHD^{150–152} and stroke¹⁵³ compared to individuals who reported sleep for 7-9 hours. Longer sleep duration, in particular, has also been associated with higher levels of IL-6 and CRP¹⁵⁴. Several sleep researchers have suggested that the higher inflammation may contribute to the CVD risk of suboptimal sleep patterns^{155,156}.

Emerging risk factors: air pollution and environmental enteropathy

In addition to contributing to the adverse effects of ‘established’ risk factors on CVD, chronic low-grade inflammation is an independent risk factor for CVD^{103,157}. ‘Chronic Inflammation’ is, thus, listed as an ‘emerging’ risk factor in Box 1.1. Like the established risk factors, many non-inflammatory emerging risk factors converge on inflammation to affect CVD risk. For example, chronic childhood undernutrition results in linear growth faltering of the growing child; the impeded linear growth (‘stunting’) is reported to be partially due to impaired gut function¹⁰⁶. The impaired gut function results from the constant exposure to sources of faeco-oral contamination disrupting the gut barrier function, and allowing passage of microorganisms and/or their bioproducts from the intestinal lumen to the gut wall leading to mucosal inflammation, ultimately leading to damage of the digestive and absorptive functions of the gut. This phenomenon termed “Environmental Enteropathy” (EE)¹⁰⁶ is among the emerging risk factors for adult-onset CVD. The linear growth faltering that occurs as a consequence of EE is correlated to a systemic increase in markers of inflammation including CRP¹⁵⁸ and is observed to be associated with adult-onset CVD risk factors including dyslipidemia, high blood pressure and dysglycaemia⁸⁹. An additional pervasive risk factor that is known to

influence CVD risk is ambient air pollution¹⁰⁵. The main mechanism of ambient pollution contributing to CVD risk is by deposition of ambient particulate matter (PM) in the air sacs of the lungs; the finer the PM, the deeper they deposit in the walls of the air sacs and the more they trigger systemic inflammation. Higher levels of markers of systemic inflammation including CRP, IL-6, and IL-1 β were observed in healthy subjects even on short-term exposure to ambient PMs¹⁵⁹.

1.4 Atherosclerosis and its immunological origins

1.4.1 Atherosclerosis

The pathological processes which cause abnormal changes to occur inside blood vessel walls and typically lead to occlusion of vessels, restriction of oxygenated blood flow and subsequent hypoxic damage to tissues are collectively referred to as ‘atherosclerosis’¹⁶⁰. The atherosclerotic process involves the gradual formation of a fatty streak that is eventually surrounded by fibrous tissue (the ‘fibrofatty’ plaque). The resulting plaque protrudes into the vessel lumen and obstructs the laminar flow of blood. Atherosclerotic CVD (ASCVD) events generally result from rupture of these atherosclerotic plaques, the rupture triggers the clotting cascade and clot formation at the site leads to occlusion of the vessel lumen and hypoxic damage to the living tissue¹⁶¹.

However, the development of atherosclerotic plaque is not the only route to causing CVD symptoms. Many other CVDs such as haemorrhagic stroke (a stroke type that causes bleeding inside brain tissues) can occur on a background of the atherosclerotic process (e.g. haemorrhagic transformation of ischaemic stroke¹⁶²) or can be primary in origin (i.e. occur de novo¹⁶³). There are other types of CVD that do not depend on the lumen-occlusive effects of atherosclerosis at all, such as Prinzmetal’s angina that causes CVD symptoms due to the spasmodic occlusion of the blood vessels supplying the heart¹⁶⁴.

Regardless, atherosclerosis underlies the pathology of most ischaemic CVDs. Therefore, an understanding of atherosclerosis, including the cellular and molecular processes that contribute to its pathology, is crucial to appreciate the contribution of modifiable risk factors to atherosclerosis and to CVD risk as a result.

The atherosclerotic process is widely described as a ‘chronic inflammatory process’^{165–167} as immune cells play important roles at every stage of the atherosclerotic plaque development, from the early fatty streak stage to late-stage fibrofatty plaque to ultimately rupture of the plaque (Figure 1.2). Hence, prior to discussion of the atherosclerotic process in the context of immune cells, I will describe what inflammation is and how it works to aid in understanding the role of inflammation in the atherosclerotic process.

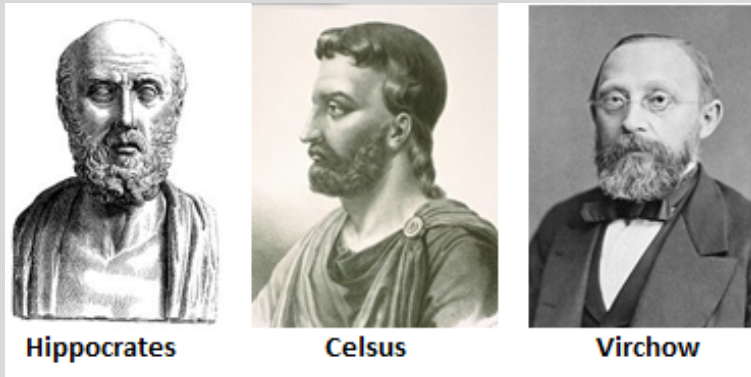
1.4.2 Inflammation

Inflammation has a number of definitions. In ancient times, it simply represented the classic symptoms of redness, swelling, heat, pain, and loss of function, in any body part due to any cause (Box 1.2)⁹¹. A more recent definition suggested inflammation was “the succession of changes which occurs in a living tissue when it is injured provided that the injury is not of such a degree as to at once destroy its structure and vitality” or “the reaction to injury of the living microcirculation and related tissues”¹⁶⁸.

As the recent definitions suggest, inflammation is a complex biological phenomenon that represents the response of the immune system to tissue damage in living tissues. Based on the intensity and duration of the immune response, inflammation can be acute or chronic. While acute inflammation is intense and lasts for a short duration (few days to few weeks), chronic inflammation is low-grade, persistent (few months to years), and not as easily resolved as acute inflammation.

Acute and chronic inflammation also differ in the type of immune cells recruited for their respective activities. For example, immune cells called ‘neutrophils’ play major roles in acute inflammation while ‘monocytes’ and ‘lymphocytes’ play important

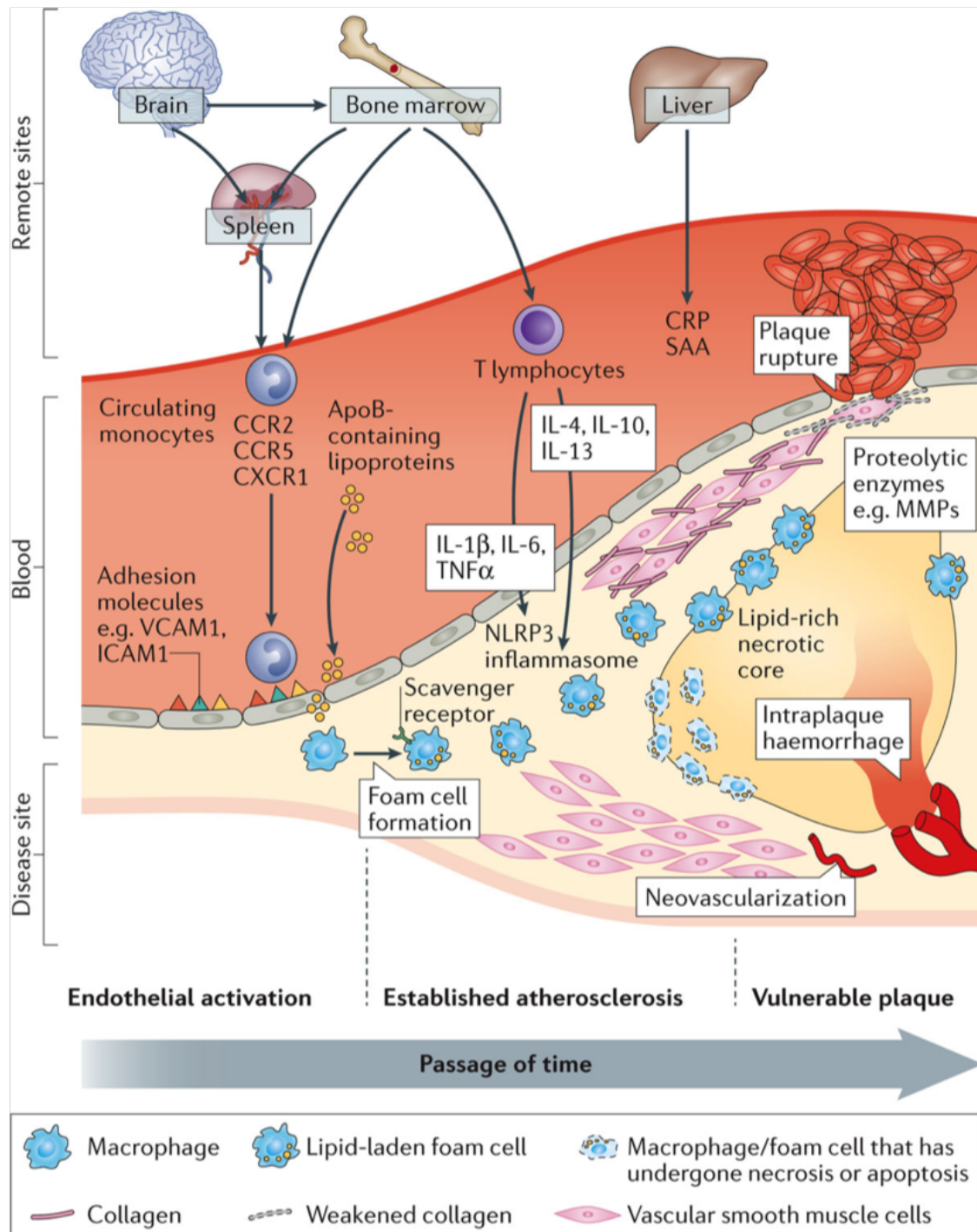
roles in chronic inflammation. Unsurprisingly, monocytes and lymphocytes have been demonstrated to drive the inflammatory process inside blood vessels that form the basis of atherosclerosis (Figure 1.2)¹⁶⁹.

Box 1.2: Inflammation and atherosclerosis

The concept of inflammation is ancient; the earliest records date back to 2000 BC when the sixth king of Babylon – Hammurabi – enacted laws for surgeons dealing with ocular abscess drainage. In the 5th century, BC Hippocrates of Cos first began to use terms like *edema* and *erysipelas* to describe the signs of inflammation. Roman medical writer Aulus Cornelius Celsus, in the 1st century CE, defined the four cardinal signs of inflammation: redness, swelling, heat, and pain. Galen (A.D 130–200) later added 'loss of function' as a fifth sign. Rudolf Virchow, a German physician – widely known as the father of modern pathology – in the 19th century was the first person to link inflammation with atherosclerosis [170] with his observation of the two processes that occur for the development of atherosclerotic plaques:

“The state of the matter here also is more or less very simply this, that two processes must be distinguished in the vessels, which are very analogous in their ultimate results; first, the simple fatty metamorphosis, which sets in without any discoverable preliminary stage, and in which the existing histological elements pass directly into a state of fatty degeneration and are destroyed, so that a larger or smaller proportion of the constituents of the walls of the vessel perishes; and, in the next place, a second series of changes, in which we can distinguish a stage of irritation preceding the fatty metamorphosis, comparable to the stage of swelling, cloudiness, and enlargement which we see in other inflamed parts. I have therefore felt no hesitation in siding with the old view in this matter, and in admitting an inflammation of the inner arterial coat to be the starting point of the so-called atheromatous degeneration”.

Experimental data in the 20th and 21st centuries led to a more explicit formulation of the inflammation hypothesis of atherosclerotic CVD – that persistently elevated mediators of systemic inflammation are associated with risk of CVD [104].



Notes: IL-6: Interleukin-6; gp130: glycoprotein 130; IL-6R α : Interleukin-6 receptor alpha; sIL-6R α : soluble Interleukin-6 receptor alpha;

Figure 1.2: Central role of immune cells and their secreted proteins in the pathogenesis of atherosclerosis. The figure illustrates the roles and activities of endothelial cells, lymphocytes, smooth muscle cells, monocytes, and macrophages, along with their secreted proteins, at every stage of the atherosclerotic process. Reproduced from Ruparel et al [171]

1.4.3 Role of immune cells in the atherosclerotic process

The atherosclerotic process typically starts when shear stress is exerted at the endothelium (inner lining of the blood vessel wall) by disrupted laminar flow of blood inside the vessel lumen (e.g. due to high blood pressure)¹⁷². The endothelial cells secrete adhesion molecules (e.g. vascular cell adhesion protein 1, VCAM-1) that promote the adhesion of monocytes along the endothelium¹⁷³. The mobilisation of monocytes inside the subendothelial space (space between endothelium and the smooth muscle cell layer of the vessel wall) is facilitated by other molecules called ‘chemokines’ (e.g. monocyte chemotactic protein 1, MCP-1) secreted by monocytes and smooth muscle cells (SMCs)¹⁷⁴. A postulated role of CRP is to increase the responsiveness of monocytes to the chemokine MCP-1¹⁷⁵.

When monocytes are inside the subendothelial space, they become large and transform into ‘macrophages’. These macrophages secrete more chemokines to attract more monocytes at the lesional site. The macrophages possess a receptor known as CD36 or ‘scavenger receptor’ that bind to and ingest oxidized lipids from apolipoprotein-B (apo-B) containing lipoprotein particles (because lipids are insoluble in blood plasma, they require proteins such as apo-B to carry it around circulation)^{173,176,177}. The lipid-laden macrophages, also known as ‘foam cells’¹⁷⁸, progressively increase in number in the lesional site, coalesce and constitute the core of the developing atherosclerotic plaque (Figure 1.2).

This growing plaque is made susceptible to rupture by proteolytic enzymes (matrix metalloproteinases) secreted by endothelial cells¹⁷⁹. Rupture of the plaque initiates the physiological clotting cascade that attempts to ‘heal’ the ruptured area by aggregating platelets, clotting factors, blood cells in a meshwork, the meshwork made of a protein called fibrin derived from fibrinogen¹⁶¹; both fibrinogen and CRP are products of the IL-6 signalling pathway (these are discussed in detail in Section 1.5).

1.4.4 Role of immune cell secreted proteins ('cytokines')

The cholesterol crystals in the ingested lipid within macrophages activates the NLRP3 (NACT, LRR and PYD domains-containing protein 3) inflammasome complex (a multi-protein complex) inside the cytoplasm of macrophage that activates a molecule called caspase-1 (also known as interleukin-1 converting enzyme, ICE)¹⁸⁰. Caspase-1 converts interleukin-1 beta (IL-1 β) to its active form; active IL-1 β is central to the activation of other 'cytokines' (cytokines are proteins that act as signalling molecules of the immune system)¹⁸¹. Activation of cytokines contributes to the initiation and progression of atherosclerotic plaque.

For example, in atherosclerotic lesions, IL-1 β induces its own expression and that of adhesion molecules in vascular cells, attract monocytes, stimulates proliferation of SMCs, and then stimulates SMCs to produce IL-6¹⁸²⁻¹⁸⁵. Like IL-1 β , IL-6 is thought to play crucial roles in SMC proliferation, SMC migration and activation of monocytes^{186,187}. As well as pro-inflammatory actions, including stimulating the hepatic synthesis of C-reactive protein (CRP) and fibrinogen^{188,189}, it is postulated that IL-6 has anti-inflammatory properties through feedback regulation of IL-1 β and up-regulation of IL-1 receptor antagonist, which blocks the action of IL-1 β by binding to its receptor¹⁹⁰. Interleukin-6 performs these activities either by binding to its membrane-bound IL-6R ('classic signalling') or soluble IL-6R ('trans-signalling') - that is reported to promote IL-6 pro-inflammatory effects¹⁹¹; binding to either receptor transduces signal inside the cell via IL-6 signal transducing (IL-6ST) protein glycoprotein 130 (gp130) and this activates the Janus kinase/signal transducer and activator of transcription 1 and 3 (JAK-STAT) inflammation pathway. Both types of IL-6 signalling - trans-signalling and classic signalling - are discussed in-depth in Section 1.5.5 as they are of particular relevance to the central aims of the present thesis.

1.4.5 Role of cytokines in non-atherosclerotic CVD

In contrast to the clear evidence for roles of cytokines in induction and progression of atherosclerosis for ASCVD, the evidence for cytokines conferring risk for non-atherosclerotic CVDs, such as intracerebral haemorrhage (ICH, a subtype of haemorrhagic stroke), is scarce.

For primary ICH, most studies implicate cytokines in the brain injury that follows ICH^{192–195} and not in the initiation of ICH. The haematoma (collection of blood) formed inside the brain in ICH activates microglia (connective tissues of the brain) and affected neurons (brain cells) to produce $\text{TNF-}\alpha$, $\text{IL-1}\beta$ and IL-6, that were reported to mediate secondary brain injury after an ICH event¹⁹⁶. However, there was conflicting evidence of clinical benefits for the amelioration of secondary brain injury following ICH with therapies that targeted the cytokines in animal models of primary ICH¹⁹⁷. Further, a clinical trial showed no overall clinical benefit on the surgical evacuation of hematoma, the source of inflammation-induced brain injury, compared to initial conservative treatment in patients with ICH¹⁹⁸. If anything, ICH was shown to induce neurogenesis in the adult human brain¹⁹⁹, and cytokines such as IL-6 were reported to play roles in axonal regeneration and neuronal survival²⁰⁰. Additionally, cytokines such as IL-6 tend to create pro-coagulant states that promote clot formation²⁰¹, biologically relevant for atherosclerotic CVD like ischaemic stroke. Indeed, the rates of ischaemic events were observed to be 6 times higher at 5 years following deep ICH²⁰².

However, there is a clear lack of studies implicating IL-6 signalling on the risk of primary ICH. The present thesis will attempt to fill this knowledge gap by exploring the observational and causal associations of components of IL-6 signalling with ICH.

1.5 CRP, Fibrinogen, and Interleukins: components of systemic inflammation

Plasma concentrations of several proteins increase during acute inflammation episodes and remain persistently elevated in chronic low-grade systemic inflammation²⁰³. These ‘acute-phase proteins’^{189,204} include CRP, fibrinogen and interleukins. The interleukins include IL-1 β and IL-6 that stimulate the hepatic synthesis of CRP and fibrinogen²⁰⁵.

1.5.1 Discovery of CRP, fibrinogen and interleukins

C-reactive protein is a ring-shaped protein produced by the liver. It binds to dead or dying cells (and some types of bacteria) and facilitates ingestion of these cells by larger immune cells such as neutrophils²⁰⁶. The discovery of CRP and fibrinogen began when Paul Ehrlich and Ilya Metchnikoff, two experimentalists, discovered and experimentally demonstrated the ‘antigen-antibody’ complementarity at the start of the 20th century. ‘Antigens’ are proteins, peptides, or polysaccharides (long carbohydrate molecules) on the surface of cells or free-flowing in blood plasma that can elicit the production of specific ‘antibodies’ that it can bind to. The antibodies are produced by cells of the immune system tailored for a particular antigen. When antibodies bind to their antigen, they ‘precipitate’, i.e. form long insoluble solid products sometimes visible to the naked eye. In the 1930s, it was observed that serum (blood plasma without proteins) taken from acutely ill patients with bacterial pneumonia precipitated when mixed with a purified extract of C-polysaccharide (the ‘antigen’) obtained from the pneumococcal bacterial cell wall²⁰⁷. This precipitation reaction was not observed in healthy individuals or patients recovering from pneumonia. This suggested that patients with acute pneumonia possess disproportionate amounts of a certain protein (the ‘antibody’) that reacts with the C-polysaccharide of the bacteria; this protein was named the *C-reactive protein*. It was later found that levels of CRP can be raised in almost any acute or chronic inflammatory events²⁰⁸ and not limited to inflammation triggered by bacteria.

Using precipitation methods to detect new proteins was not new. Half a century earlier from the discovery of CRP, Hammarsten used precipitation methods in horse plasma to isolate *fibrinogen* – a rod-like glycoprotein complex that is essential for the formation of blood clots in humans²⁰⁹. Soluble fibrinogen, like CRP, is produced by the liver and accumulates at sites of tissue injury where it is cleaved to insoluble fibrin. The insoluble fibrin forms a mesh that traps platelets and other blood cells and plugs the site of injury. This process is central to the sequence of events that occur post-atherosclerotic plaque rupture (Section 1.4.3 and Figure 1.2).

Both CRP and fibrinogen synthesis are regulated by signalling molecules called *interleukins* that play essential roles in orchestrating the immune system response. It was in the early 1980s when several independent groups began identifying these signalling molecules, isolating their complementary DNA (cDNA) (i.e. DNA complementary to the messenger RNA that produces the protein) from monocytes in humans. These signalling molecules were termed ‘*interleukins*’ given their role in signalling between leukocytes (i.e. white blood cells). Starting from 1984, Dinarello et al published the cDNA sequence of Interleukin-1 β (IL-1 β) precursor²¹⁰, Aggarwal and colleagues characterized Tumour Necrosis Factor-alpha (TNF- α) in 1985²¹¹, followed by 3 independent groups discovering the cDNA sequence of IL-1-induced 26-kDa protein or (later named) ‘*Interleukin-6*’ (IL-6) in 1986^{212–214}. In 1987, Gauldie et al first showed the regulation of fibrinogen in liver cells by IL-6²¹⁵. Several other groups demonstrated regulation of CRP synthesis by IL-6 also in liver cell lines^{188,216}. Around the same time, Libby et al²¹⁷ reported the induction of IL-1 gene expression in human endothelial cells and smooth muscle cells by TNF- α and Van Damme et al²¹⁸ demonstrate the induction of IL-6 by IL-1 β and TNF- α in human fibroblasts. It was later shown that IL-1 β also induces IL-6 synthesis in endothelial cells²¹⁹, epidermal cells²²⁰ and monocytes²²¹.

These proteins, CRP, fibrinogen, IL-1 β and IL-6 have been the subject of a number of studies that aimed to assess whether they would be predictive of or causal for CVD^{92,93,96,222–224}.

1.5.2 Observational studies

Many of the evidence in humans that suggested the role of inflammation in CVD were generated by observational studies. Several prospective observational studies in humans and their meta-analyses consistently demonstrated strong associations of CRP, fibrinogen and IL-6 protein with both atherosclerosis and incident CVD events^{92,93,225–228} (Table 1.1). In the largest meta-analyses aggregating data from mostly Europeans, clear log-linear associations were observed for log CRP and fibrinogen with atherosclerotic CVD such as IHD and ischaemic stroke (IS)^{92,93}, and log IL-6 with CHD^{225,229}. The large sample sizes of Europeans in the meta-analyses ensured that these studies were well-powered, i.e. they had a low probability of observing associations of CRP and IL-6 with CVD if there was none (type II error). The large sample sizes also enabled a high precision of risk estimates (i.e. lower standard errors) contributing to the internal validity of the findings. Internal validity refers to the extent to which observations in the study represent the truth in the population of study participants. However, a couple of questions remained:

1. Are these findings generalizable to other non-European population groups (i.e. external validity)?, e.g. East Asians?

Indeed, only a few studies included populations of East Asian origin for assessment of associations of CRP and fibrinogen with stroke types and CHD^{230,231}. Associations observed in Europeans may not necessarily be the same (either in terms of direction or magnitude) in other population groups. Indeed, even inferences from an internally valid study in a single study sample that is not a random sample of the target population being studied cannot be generalizable to the target population of interest²³² and one reason for this could be due to differences in the distribution of potential confounders and effect modifiers between the study population and target population or between different populations in general²³³. For example, levels of LDL-C are inherently much lower in the East Asian population than the Europeans (mean ~140 mg/dL vs ~90 mg/dL)^{54,92,93}. Given that LDL-C is a well-established risk factor of CVD⁵⁵, is it possible that differences in the distribution

of LDL-C between these population groups limit generalisability of associations (for example, association of CRP with CVD) observed only in one population? For these reasons, external validity is as important as internal validity of studies and neither should be considered in isolation when appraising a study²³⁴. Hence it is important to understand whether there are differences in the direction or magnitude of associations of markers of inflammation with CVD in other population groups that have different risk factor distributions.

Table 1.1: Observational Evidence

Tabulating observational studies of markers of inflammation showing the association CRP, fibrinogen and interleukin-6 with ischaemic stroke, intracerebral haemorrhage and coronary heart disease

Protein	Study reference	Study design	Population	IS		ICH		CHD	
				N	RR (95% CI) ^a	N	RR (95% CI) ^a	N	RR (95% CI) ^a *
CRP	[92]	Meta-analysis	Europeans	1931	1.27 (1.15-1.40)	469	1.07 (0.86-1.32)	5373	1.36 (1.29-1.45)
	[230]	Meta-analysis	East Asians	1105	1.40 (1.08 - 1.81)	738	1.04 (0.66 - 1.65)	529	1.75 (0.96 - 3.19)
	[231]	Prospective study	Chinese	1049	1.15 (1.06 - 1.24)	383	0.97 (0.86 - 1.08)	-	-
	[235]	Prospective study	European women	517	1.69 (1.26 - 2.29)	111	0.83 (0.44 - 1.57)	-	-
Fibrinogen	[93]	Meta-analysis	Mostly Europeans	596	1.93 (1.47 - 2.52)	215	1.45 (0.91 - 2.29)	2534	1.82 (1.60 - 2.06)
	[235]	Prospective study	European women	517	1.01 (0.77 - 1.34)	111	1.14 (0.64 - 2.00)	-	-
Interleukin-6	[229]	Meta-analysis	Europeans	-	-	-	-	5730	3.34 (2.45 - 4.56)
	[225]	Meta-analysis	Europeans	-	-	-	-	7982	1.25 (1.19 - 1.32)
	[236]	Case-control study	Europeans	-	-	-	-	931	2.53 (1.86 to 3.43)

N: Number of cases; IS: Ischaemic stroke; ICH: Intracerebral haemorrhage; CHD: Coronary heart disease; CRP: C-reactive protein; RR: Relative risk; CI: Confidence Interval

^aAll studies reported RR per SD increase of natural log-transformed biomarker except [235] - top vs bottom quartile, and [236] - top vs bottom third.

2. Are markers of systemic inflammation associated with the risk of ICH?

As noted before, one of the CVD phenotypes that have been understudied in all population groups is ICH - a subtype of haemorrhagic stroke that has been increasing in incidence and prevalence in China²³⁷. Sample sizes for investigation of association of CRP and fibrinogen with ICH in East Asians were particularly low (largest $N_{\text{cases}} = 738$ ²³⁰) compared to sample sizes assessing associations with other atherosclerotic CVDs (Table 1.1).

It is for these reasons, one of the aims of the present thesis is to investigate associations of CRP and fibrinogen with CVDs, including ICH, in a large cohort of East Asians.

1.5.3 Randomised controlled trials

Two major problems in observational studies are the problems of bias due to confounding and reverse causation. Both of these problems are solved, to a large extent, by random allocation of study participants to intervention and placebo groups, making the two groups comparable by ensuring equal distribution of confounders between the groups. The prospective nature of randomised controlled trials (RCTs) also ensures the study design is robust to reverse causation bias. RCTs, therefore, remain the gold standard for establishing causation. Despite being the gold standard, they are not as widely used for epidemiological queries as observational studies. This is because:

1. RCTs are expensive

Randomised controlled trials are ‘controlled’ by design. From the random assignment of study participants to treatment/placebo arms to investigator-blinded evaluations of study participants, every step requires careful thought, planning and management. The logistics and human resources thus required to execute a trial are vastly different from conducting an observational study. In clinical trials that evaluate the efficacy and safety of drugs, there is an additional cost incurred to manufacture the drug at

its appropriate dose. Further, not all drugs that are investigated in trials eventually make it to the market. When studying the proportion of drugs approved during a ten-year period (1991–2000) for ten big pharmaceutical companies in the United States and Europe, it was shown that only 1 in 5 drugs investigated for CVD was approved²³⁸. For CVD drugs, 44% of these drug failures were attributed to poor efficacy and 24% due to safety concerns²³⁹. On average, it is estimated to cost \$2.5 billion and require 12.5 years per approved drug for CVD²⁴⁰. These cost considerations have prompted investigators to look for better ways to reduce this high attrition rate observed in costly clinical trials. One approach that I will discuss in detail in my thesis is the use of genetic data from observational studies for causal inference (Mendelian randomisation); this approach is reported to increase the odds of success of a drug in clinical trials^{241–243}.

2. RCTs are not always appropriate to address an epidemiological query

Randomised trials are not always an appropriate design for evaluating ‘high-risk’ exposures (e.g. smoking or alcohol consumption); high-risk exposures are those where there is a reasonable consensus among large numbers of well-powered observational studies that it is likely that exposures can cause addiction or harm. Prime examples of these exposures include alcohol consumption and smoking. But this means that causal questions related to smoking or alcohol consumption on important health outcomes such as CVD remain unresolved. Here, again, genetic data from large population biobanks can be leveraged to understand whether there is a causal relationship between the exposure and outcome. For example, as discussed in Section 1.3, when carriers of the A-allele of rs1229984 genetic variant representing a non-synonymous mutation of the alcohol dehydrogenase 1B (*ADH1B*) gene which encodes the *ADH1B* enzyme (these individuals flush when they drink due to their physiological inability to metabolise acetaldehyde which accumulates in the blood and cause flushing; these individual therefore tend to avoid alcohol intake²⁴⁴) were compared with individuals carrying the G-allele of the same genetic variant (these individuals have a functioning alcohol dehydrogenase 1B enzyme, can metabolise

acetaldehyde, and therefore can drink without any physiological reactions), it was shown that lower alcohol consumption (with no minimum safe threshold of drinking) was causally associated with a lower risk of CHD⁷⁵.

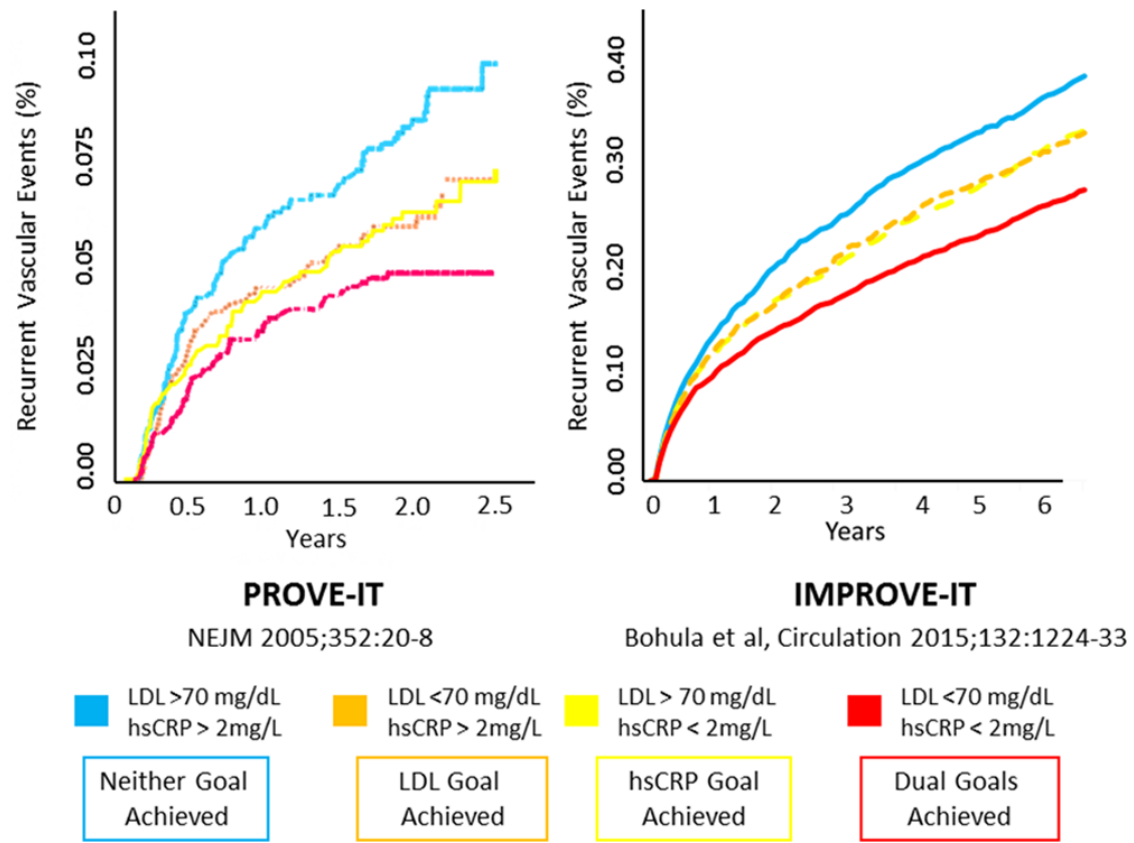
In spite of these limitations, to establish the efficacy and safety of drugs, RCTs remain the gold standard and produce the strongest evidence (or lack of evidence) of causal associations. Randomised controlled trials although primarily designed to test hypotheses generated from observational and experimental studies can also generate new hypotheses. But this hypothesis generation by trials often breaks the randomisation because it involves the examination of a subgroup of individuals with a combination of traits. For example, subgroup analyses of early RCTs of statins provided the first insights into the anti-inflammatory effects of statins that were independent of its lipid-lowering effect^{245,246}. Three early RCTs evaluating the efficacy of statins on CHD (REVERSAL²⁴⁷, PROVE-IT²⁴⁸ and IMPROVE-IT²⁴⁹) demonstrated maximal clinical benefit for CHD in patients who achieved low levels of both LDL-C and CRP concentrations (Figure 1.3). Prospective analysis of the JUPITER (The Justification for the Use of Statins in Primary Prevention: an Intervention Trial Evaluating Rosuvastatin) trial²⁵⁰ that randomised patients with low LDL-C concentrations but high levels of CRP to receive Rosuvastatin or placebo showed greater relative reductions in CHD in those achieving lower post-treatment LDL-C and CRP levels, compared with lowering of LDL-C only. The anti-inflammatory effect of statins was also supported by human cell line studies; statins reduced IL-6-induced CRP production directly in hepatocytes via inhibition of protein geranylgeranylation and, additionally, by inhibition of STAT3 phosphorylation^{251,252}.

However, the causal role of inflammation on CVD was unclear as subgroup analyses of RCTs break randomisation and, therefore, these were essentially observational analyses of trials that evaluated the relevance of the lipid-lowering effect of statins to CVD. To test whether inhibition of the IL-6-CRP inflammation pathway, independent of lipid-lowering, leads to cardiovascular disease risk reductions, Ridker

et al set up the canakinumab Anti-inflammatory Thrombosis Outcome Study (CANTOS)²²³ and Cardiovascular Inflammation Reduction Trial (CIRT)²⁵³.

CANTOS and CIRT studies The CANTOS trial used a human monoclonal antibody – canakinumab – that binds to and inhibits the activity of IL-1 β ²⁵⁴ while low-dose methotrexate (LDM) in CIRT was hypothesized to exert athero-protective effects via broad-spectrum anti-inflammatory mechanisms^{255,256}. In ~10,000 patients with prior MI (>90% on statins, median baseline LDL-C levels in the placebo group: 83 mg/dL), canakinumab in the CANTOS trial lowered CRP and IL-6 levels, with no effect on lipids, and reduced risk of recurrent CVD represented by their primary endpoint (non-fatal MI, non-fatal stroke or CVD death) compared to placebo (HR [95% CI]: 0.85 [0.74, 0.98], $p = 0.021$)²⁵⁷; this clinical benefit was proportional to the magnitude of reduction in CRP levels²⁵⁸. By contrast, LDM in CIRT did not lower CRP or IL-6 levels and did not significantly reduce recurrent cardiovascular events compared to placebo²⁵⁹.

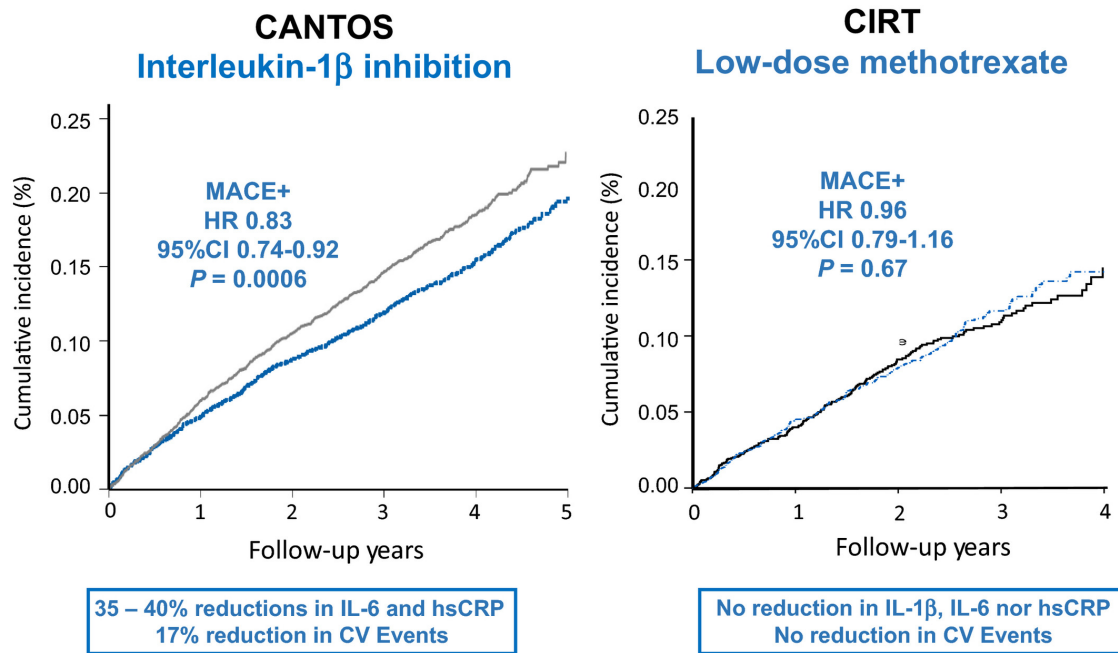
The discordant findings from CANTOS and CIRT highlighted the importance of targeting the IL-1 β –IL-6–CRP pathway of innate immunity in cardiovascular disease²⁶⁰ (Figure 1.4).



LDL-C: low-density lipoprotein cholesterol; hsCRP: high-sensitivity C-reactive protein
Reproduced from Ridker et al [261]

Figure 1.3: Post-hoc analysis of PROVE-IT and IMPROVE-IT trials, that investigated effect of statins on risk of recurrent vascular events, in different patient groups

Illustrating that the lowest risk of recurrent vascular events were achieved in patients when both LDL-C and hsCRP were lowered



MACE+: recurrent myocardial infarction, stroke, need for urgent revascularization, or cardiovascular death.
Reproduced from Ridker et al [260]

Figure 1.4: Comparison of cumulative risk of recurrent CVD outcomes in the CANTOS and CIRT

Illustrating that interleukin-1 β inhibition that lowered interleukin-6 (IL-6) protein and high-sensitivity C-reactive protein (hsCRP) reduced risk of major adverse cardiovascular events (MACE), compared to low-dose methotrexate that did not lower IL-6 or hsCRP and did not lower the risk of MACE

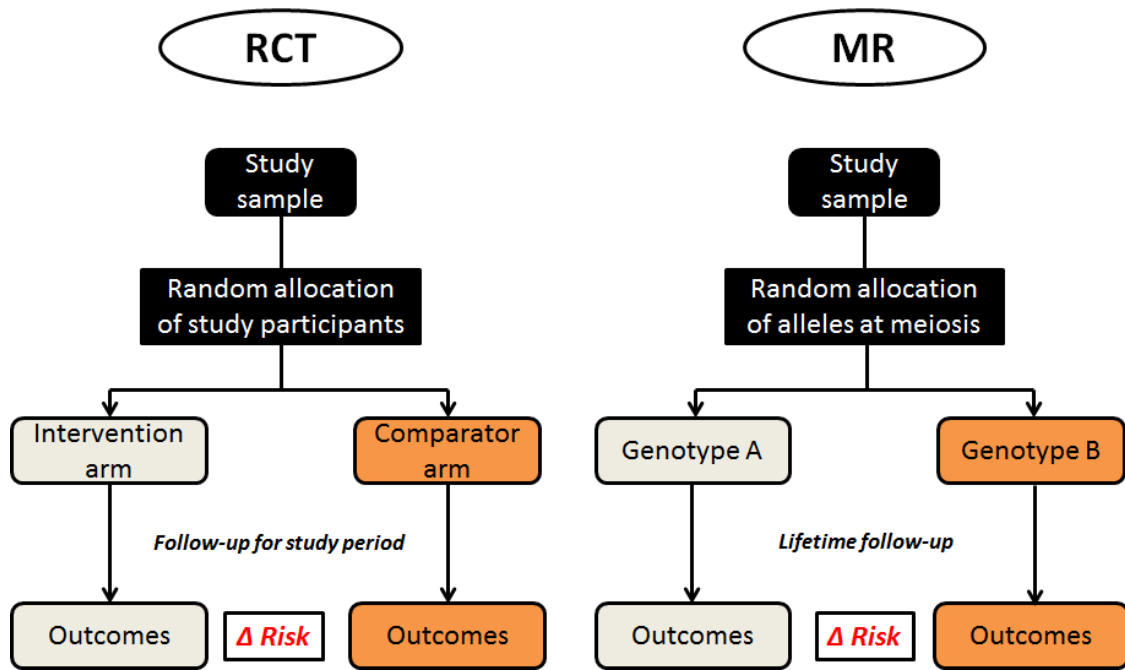
The CANTOS trial was the first RCT demonstrating the relevance of inflammation to CVD risk. However, the success of CANTOS is not without caveats. For example, although canakinumab reduced the risk of a composite CVD endpoint by 15%, this was mainly driven by the reduction in risk of MI by 24% with no effect observed for other atherosclerotic CVD such as stroke. The question here is whether this was due to the lack of power of the CANTOS trial to detect an effect on stroke or that inflammation plays a disproportionately larger role on some atherosclerotic CVD more than others? Further, there was a slight increase in the risk of fatal infections with canakinumab compared to placebo, consistent with the role of immuno-modulatory role of IL-1 β ²⁶².

For both the therapeutic and adverse effects of canakinumab, a query that looms is whether these effects are due to blockade of IL-1 β or mediated by lower concen-

trations of proteins regulated by IL-1 β , such as IL-6? One reliable way to answer these questions is to undertake another RCT with an agent that only blocks the IL-6 signalling pathway (e.g. tocilizumab) powered to investigate and compare causal effects of blocking IL-6 signalling on CVD, MI, and whether adverse effects observed in CANTOS are also seen in this trial. In the absence of such an RCT, a separate approach that can generate causal inferences similar to (but not equivalent to) RCTs is the use of genetic variants in humans (if available) that can reliably represent effects of drug target modulation.

1.5.4 Mendelian randomisation studies

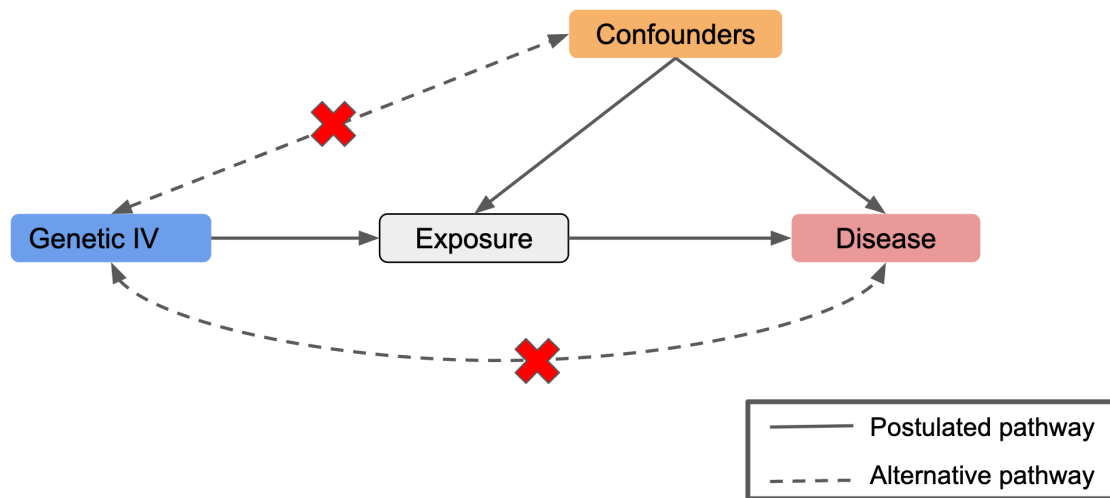
Analogous to RCTs, it is possible to exploit the naturally randomised allocation of germline genetic variants that encode biological traits in a population to draw causal inferences for the lifetime impact of an exposure trait (e.g. levels or activity of proteins) on the risk of an outcome (Figure 1.5). The approach of using genetic variants as instruments for exposure as formalised by George Davey Smith and Shah Ebrahim was termed ‘Mendelian randomisation’ (MR)²⁶³. The principle of MR is based on what has become known as Mendel’s first and second laws of genetic inheritance, the Law of Segregation and the Law of Independent Assortment, respectively²⁶⁴. The Law of Segregation implies that, at conception, the offspring will randomly inherit one of two of their mother’s and one of two of their father’s allele, at any given chromosomal locus. The Law of Independent Assortment states that, during the formation of gametes in an individual, alleles will segregate to gametes independently of each other. Although MR was initially conceived in parent-offspring trios in 2003, subject to consistent family formations in a population²⁶⁵, one can still use the MR approach with population-level data of unrelated individuals to obtain approximate and reliable epidemiological associations between exposures and outcomes.



RCT: randomised controlled trial; MR: Mendelian randomisation; δ - change

Figure 1.5: Comparison of study designs evaluating causality
Illustrating the conceptual similarity of a randomised controlled trial (RCT) with Mendelian randomisation (MR) study to investigate causal associations

A key aspect of MR analyses is the derivation of genetic instruments representing the exposure trait. Genome-wide association studies of the exposure trait are the chief source of genetic instruments that are associated with the exposure trait (relevance). The selected genetic instruments for the exposure should, additionally, not be associated with potential confounders of the exposure-outcome relationship (independence) and not affect the outcome trait conditional on exposure and confounders (exclusion restriction). Although these three requirements, relevance, independence, and exclusion restriction, constitute the instrumental variable (IV) assumptions²⁶⁶ (Figure 1.6), some of these assumptions are testable. While the relevance assumption is required for the selection of IVs, providing evidence to reliably prove independence and exclusion restriction assumption remains impossible. Nevertheless, to assess if the genetic IVs are at least plausible, independence and exclusion restriction assumption can be tested by investigating the association of the genetic instrument with several traits, analogous to testing the balance of baseline characteristics between treatment and placebo arms in an RCT²⁶⁷. Additional



IV: instrumental variable

Figure 1.6: Mendelian randomisation

Illustrating the general concept of Mendelian randomisation (MR) approach. For a genetic instrument to reliably represent the exposure, it must meet three IV assumptions: 1) the genetic instrument must be associated with the exposure of interest, 2) the genetic instrument must not be associated with confounders of exposure-outcome association, and 3) the genetic instrument must affect outcome via the exposure

tests are required when the genetic IVs represent molecular traits such as protein abundance and these are detailed in the next section.

MR analysis using proteomic traits as exposures

There are two broad approaches to selection of genetic instruments for MR analysis where molecular traits are exposure traits. The first approach entails selection of genetic variants associated with the molecular trait from across the genome ('polygenic' instrument). The second approach involves use of locus-specific genetic instruments, either cis-acting genetic variants (i.e. genetic variants located near the region of the genome encoding the molecular trait) or trans-acting genetic variants (i.e. genetic variants distantly located from the molecular trait-encoding genetic region). When the molecular trait is a protein, the respective genetic instruments are referred to as cis-pQTL (protein-quantitative trait locus) or trans-pQTL.

There are a couple of advantages of the polygenic approach to genetic instrument selection. A key reason to consider polygenic instruments is power - the larger

number of genetic variants are likely to explain more of the variation in plasma protein concentrations than use of a few locus-specific instruments. Furthermore, the larger number of instruments permit use of additional analytic methods (e.g. MR-Egger that typically requires more than three instruments) to explore whether the observed MR association is robust to violations of the instrumental variable assumptions. However, the inclusion of trans-pQTL in the genetic instrument agnostic to prior biological knowledge (i.e. evidence linking the trans-pQTL associated genes with the protein under investigation) risks the association between the protein and outcomes driven by pathways that are not necessarily mediated through the protein (i.e. horizontal pleiotropy). Additionally, while the use of MR-Egger and other post-MR sensitivity analyses can assess the extent of and potentially correct for the horizontal pleiotropy, the corrected estimates typically have wider confidence intervals reflecting the low power of these sensitivity tests²⁶⁸. For these reasons, it is not ideal to use the polygenic approach to appraise the proteins as therapeutic targets.

In contrast to the polygenic approach, the use of locus-specific genetic instruments provides a greater degree of specificity of the genetic instrument (and, therefore, a lower chance of MR associations driven by horizontal pleiotropy) for the protein under investigation (Figure 1.7). The high degree of specificity comes at a cost of power, as the number of locus-specific instruments associated with a protein at a given threshold of significance generally tends to be lower than genome-wide instruments and, as a consequence, often explains less of the variation of the protein abundance. The lower number of locus-specific instruments also prevents assessment of validity of MR associations using MR-Egger or weighted median approaches but permits use of other methods such as genetic colocalisation to examine whether the MR estimate is confounded by linkage disequilibrium. The major advantage of the locus-specific approach is that, unlike the polygenic approach, the high specificity of the instruments enables researchers to interpret the locus-specific MR estimates as evidence (or no evidence) of likely therapeutic or adverse effects of pharmacologically

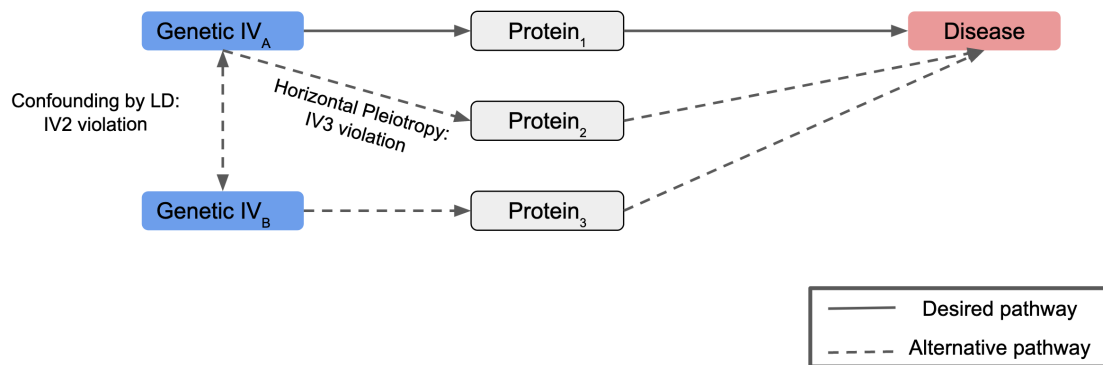


Figure 1.7: Mendelian randomisation using proteins as exposure traits. The figure illustrates the desired pathway (solid arrows) and alternative pathways (dashed arrows) that can potentially explain the association of genetically predicted levels of protein₁ with disease. The alternative pathway can be due to the genetic instrument (IV_A) being in linkage disequilibrium (LD) with a second genetic instrument (IV_B) that associates with disease via protein₃ or association of IV_A with protein₂ that associates with disease.

perturbing a protein²⁶⁹. There are two types of locus-specific instruments: cis-pQTL and trans-pQTL. In MR analysis of drug targets (in which more than 90% are proteins), it is more straightforward to infer that the likely biological pathway from a protein to a disease or risk factor of a disease is via the protein under investigation when cis-pQTL instruments are used for the MR analysis as opposed to trans-pQTL instruments. Nevertheless, trans-pQTL instruments are often informative when there is evidence of biological plausibility linking the trans-pQTL associated genes with the protein (e.g. IL6R genetic instruments for CRP) and the protein can then be used as a marker of target modulation (e.g. using CRP as a marker of genetically mimicking IL-6R blockers, as has been done in my thesis).

Informing drug development pipelines using MR analysis of proteins

Drug development pipelines are incredibly expensive, due to only a fraction of drugs translated successfully from early stage clinical trials to clinical approval for specific indications (Figure 1.8). MR studies of proteins can provide important pre-trial on-target efficacy and adverse drug event data that can potentially influence decisions to invest in large-scale clinical trials for specific drugs. It has been demonstrated that drugs with genetic support are twice more likely to be approved than drugs

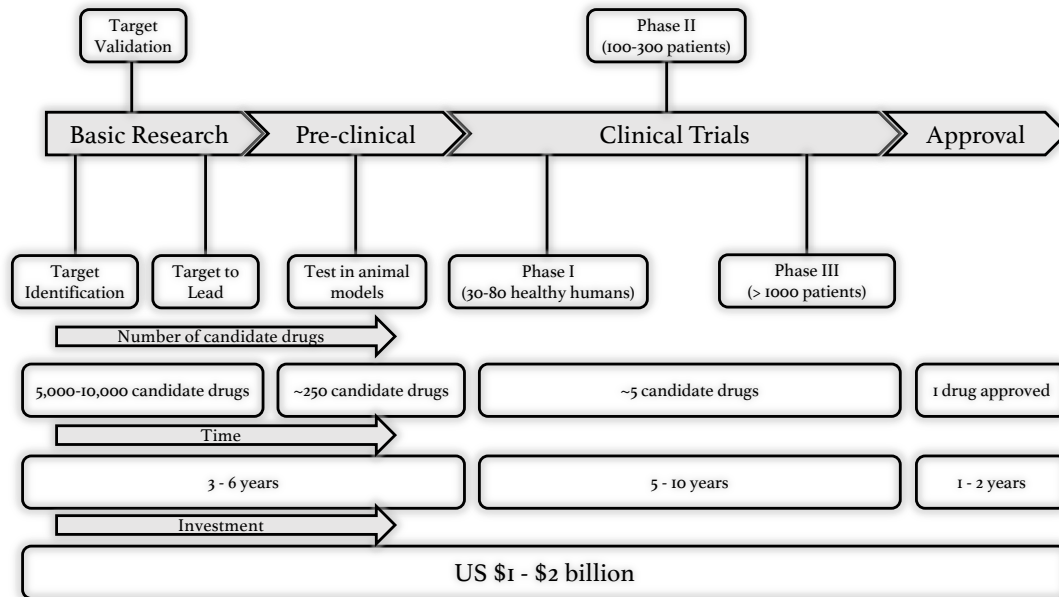


Figure 1.8: Drug discovery and development pipeline

Illustrating the processes spanning from drug discovery to approval. There is a high attrition rate of compounds that show promise in preclinical studies in animal models. Indeed, only 1 in 5,000 compounds screened in early-stage discovery successfully makes it through to market. Data for the figure was obtained from Hughes et al[271]

without genetic support²⁴². The most recent estimates suggest that drugs with human genetic evidence increase the probability of drug approval for a given disease by more than two-fold²⁷⁰.

To aid drug development, MR analyses using proteins can potentially discover:

1. Novel druggable targets for existing diseases

For example, Chong et al performed two-sample MR analyses for ischaemic stroke deriving *cis*-acting genetic instruments from summary data of 5 proteomic GWAS datasets and stroke data from MEGASTROKE and UK Biobank²⁷². Using these genetic instruments, they sought to discover any druggable target that would confer protection against any IS subtypes. They discovered *SCARA5* that encodes the scavenger receptor class A5 as a novel drug target for the cardio-embolic type of IS, with no predicted adverse effects.

2. Potential adverse effects of druggable or existing drugs

Potential adverse effects can be assessed by conducting phenome-wide association studies (‘PheWAS’) with given a drug or druggable target. For example, Gill et al performed a two-sample MR-PheWAS in UK Biobank using *cis*-acting genetic instruments for the blood pressure lowering effects of five classes of antihypertensive medications: angiotensin-converting enzyme (ACE) inhibitors, angiotensin receptor blockers, beta-blockers, calcium channel blockers (CCB) and thiazide diuretics deriving instruments from a GWAS of systolic blood pressure²⁷³; the study predicted a higher risk of diverticulosis (a condition which causes the inner layer of the gut wall to protrude outward through the outer layer) with a specific class of CCBs²⁷³. Similarly, other MR studies have also shown adverse effects of existing drugs. A notable example is statin use and risk of diabetes²⁷⁴ and ICH⁵⁴.

3. Novel indications for existing drugs (‘therapeutic repurposing’)

A good example where MR has informed of a potential therapeutic repurposing opportunity is the drug-target MR study of interleukin-6 receptor (IL-6R) signalling. Swerdlow et al⁹⁷ and Sarwar et al⁹⁶ used a drug-target MR approach with a functional variant (Asp358Ala, Section 1.5.5) to demonstrate the potential of repositioning anti-inflammatory drugs such as IL-6R blockers (e.g. tocilizumab) typically indicated for chronic inflammatory diseases like rheumatoid arthritis, for CHD. The genetic variant Asp358Ala is a functional genetic variant that alters the concentrations of soluble and membrane-bound forms of IL-6R, increasing concentrations of the soluble form over the membrane-bound form, whilst IL-6R blockers like tocilizumab blocks both forms of the IL-6R. However, the effects of IL-6R blockade, such as lower CRP and fibrinogen (along directionally consistent effects between Asp358Ala and tocilizumab on several biomarkers documented by Swerdlow et al⁹⁷), is observed in both tocilizumab-treated patients and individuals harbouring the Asp358Ala variant, providing the empirical basis for the claim that Asp358Ala variant mimics the effect of tocilizumab. The present thesis investigates the Asp358Ala variant in greater detail, comparing effects between populations and with effects of genetic instruments

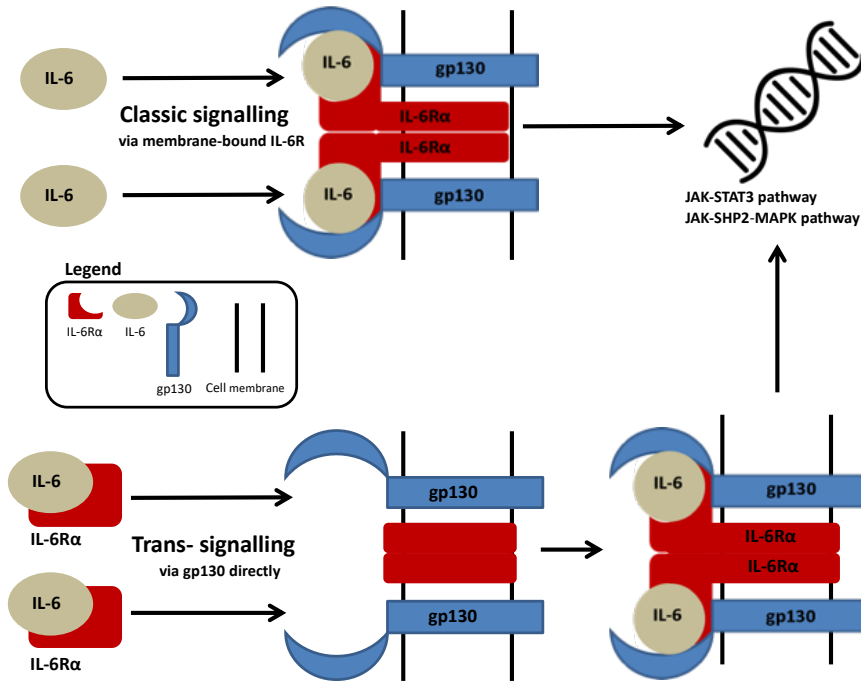
representing lower IL-6 plasma protein concentrations, to understand whether effects of Asp358Ala are equivalent to effects of lower IL-6 protein concentration. However, to understand the effects of the Asp358Ala genetic variant, it is essential to first understand the modes of IL-6 signalling and this is described next.

1.5.5 Interleukin-6 signalling and its components

Modes of IL-6 signalling

The IL-6R is a receptor complex made of two molecular components: IL-6R alpha (IL-6R α) which interacts with the IL-6 protein, and glycoprotein 130 (gp130) which transduces the signal from the IL-6R α -IL-6 interaction to within the cell. Two modes of IL-6 signalling - classic signalling and trans-signalling - govern localised and broader systemic effects of IL-6 signalling, respectively. Classic signalling occurs when IL-6 binds to membrane-bound IL-6 receptors (mIL-6R α), present in limited cells types including liver cells and some immune cells²⁷⁵. The extracellular domain of mIL-6R α can be cleaved from the IL-6R receptor complex by proteases to the soluble IL-6 receptor (sIL-6R). Trans-signalling occurs when IL-6 binds to sIL-6R; the IL-6:sIL-6R complex can bind to gp130 membrane-bound receptor that is present in most cell types, and exert the broader systemic effects²⁷⁶ (Figure 1.9). In healthy physiological states, the relatively higher molar concentrations of sgp130 are thought to preferentially bind to the IL-6:sIL-6R complex and prevent trans-signalling²⁷⁷. However, this sgp130-mediated physiological inhibition is lifted in pathological inflammatory states when levels of IL-6 overwhelm the buffer capacity of sgp130²⁷⁸ and allow trans-signalling to occur. The downstream intracellular consequences of classic and trans-signalling are the same. In either signalling, Janus kinase (JAK) signal transducer and activator of transcription 3 (STAT3) pathway, and the JAK- Src homology region 2 (SH2) domain tyrosine phosphatase 2 (SHP2)-mitogen-activated protein (MAP) kinase pathway are activated and increase transcription of IL-6 responsive genes. The IL-6 responsive genes transcribed will depend on what kind of cells are activated, e.g. in liver cells, genes encoding CRP

will be expressed more than other genes²⁷⁹; in endothelial, genes encoding MCP-1 will be preferentially expressed²⁸⁰.



Notes: IL-6: Interleukin-6; gp130: glycoprotein 130; IL-6Rα: Interleukin-6 receptor alpha; JAK: Janus kinase; SHP2: Src homology region 2 domain tyrosine phosphatase 2; MAPK: mitogen-activated protein kinase.

Figure 1.9: Different modes of interleukin-6 (IL-6) signalling

Classic signalling occurs when IL-6 binds to the membrane-bound IL-6Rα (present in limited tissues such as liver cells and immune cells); trans-signalling occurs when IL-6 first binds to cleaved soluble IL-6Rα. In both cases, this IL-6:IL-6Rα complex induces signal via gp130 expressing cells and two downstream pathways are activated: JAK-STAT3 and JAK-SHP2-MAPK pathways

Asp358Ala genetic variant alters the ratio of mIL-6R to sIL-6R

The functional variant (rs2228145-C) causes a non-synonymous mutation in the *IL6R* gene, changing the aspartic acid to alanine in the 358th position of the IL-6R polypeptide chain (hence known as Asp358Ala *IL6R* variant)²⁸¹. The Asp358Ala mutation makes the mIL-6Rα susceptible to cleavage by circulating proteases and is associated with higher concentrations of sIL-6Rα. While a reduction of mIL-6Rα would plausibly reduce classic signalling (as indicated by lower CRP concentrations

with Asp358Ala), it remains unclear whether the higher sIL-6R α increases trans-signalling. As stated earlier, a number of investigators hypothesized the role of sgp130 that buffers the systemic effects of trans-signalling under physiological conditions^{277,278,281,282}. However, so far, strong experimental evidence of this buffering phenomenon is lacking. For this reason, for the aims of the present thesis, I avoided making strict comments on the direction of the effect of Asp358Ala on IL-6 signalling and instead refer to it as ‘altered’ IL-6R signalling throughout the thesis. Although Swerdlow et al⁹⁷ and Sarwar et al⁹⁶ have shown that Asp358Ala associations with biomarkers such as CRP were directionally consistent with effects of tocilizumab (an IL-6R blocker indicated for chronic inflammatory disorders like rheumatoid arthritis and blocks both classic and trans-signalling), it remains unclear whether the lower risk of CVD also observed with Asp358Ala was due to lower classic signalling or higher trans-signalling (or a combination of both). The direction of classic and trans-signalling of the Asp358Ala variant can help us interpret whether tocilizumab (which is reportedly mimicked by the Asp358Ala variant) is likely to have cardioprotective benefits as demonstrated by the association of the Asp358Ala variant with a lower risk of CHD in Europeans.

1.6 Evidence and knowledge gaps

Observational associations of plasma CRP and fibrinogen with CHD and IS have been extensively studied in Europeans, with very limited studies in East Asians (Table 1.1). Furthermore, underpowered studies have reported largely null associations of CRP and fibrinogen with ICH - a subtype of haemorrhagic stroke, that is becoming increasingly prevalent in China²³⁷.

Additionally, compared to MR studies on lipids for CVD, MR studies investigating the association of components of systemic inflammation with CVD risk have been understudied (Table 1.2). Over the years, drug-target MR studies investigated the effects of lipid-lowering therapies using genetic variants to represent lifelong lower LDL-C levels; these genetic variants are associated with lower LDL-C levels at

various drug-target relevant loci such as *LDLR*, *PCSK9*, *HMGCR*, and *NPC1L1*²⁸³. However, only a few studies explored causal associations of inflammatory markers with CVD (Table 1.2); this is mainly due to the lack of reliable genetic instruments representing different components of systemic inflammation.

In addition to being understudied, there were discrepant associations between observational associations of markers of inflammation and their corresponding MR studies. Comparing observational and MR studies of CRP and fibrinogen, the MR studies (Table 1.2) reported that observational associations of CRP and fibrinogen with IS and CHD (Table 1.1) were largely non-causal. By contrast, while observational studies showed higher levels of IL-6 protein was associated with a higher risk of CVD (Table 1.1), MR studies of altered IL-6R signalling using the Asp358Ala variant which increased IL-6 concentrations, was associated with a lower risk of CVD (Table 1.2). As mentioned, previous studies assumed that the Asp358Ala variant lowers both classic and trans-signalling, and this assumption was based on the directional consistency of the Asp358Ala variant's phenotypic associations (mainly circulating IL-6 and sIL-6R) with effects of tocilizumab reported in clinical trials, raising the possibility that tocilizumab (indicated for chronic inflammatory diseases like rheumatoid arthritis) could be repurposed for CVD outcomes⁹⁷. However, the biological mechanism of, for example, higher circulating IL-6 and sIL-6R observed in healthy carriers of the Asp358Ala variant is unlikely to be the same as the biological mechanism of elevation of the same circulating markers in patients receiving tocilizumab. For instance, Nishimoto et al's²⁸⁴ pharmacokinetic analyses in patients who had received tocilizumab found that the higher sIL-6R and IL-6 protein observed with tocilizumab was due to tocilizumab competing with free IL-6 protein to bind to sIL-6R and form tocilizumab:sIL-6R complexes; this prolongs the half-life of sIL-6R and results in higher circulating IL-6 protein, without increasing IL-6 trans-signalling. However, the Asp358Ala variant is associated with higher sIL-6R:IL-6 complexes which can, if not neutralised by circulating sgp130, trigger IL-6 trans-signalling - the mode of IL-6 signalling that underlies the pathology of a number of chronic inflammatory diseases. A proper assessment of the contributions

of classic and trans-signalling will ideally require data on the ratio of circulating sIL-6R:IL-6 complex and sIL-6R:IL-6:sgp130. As of writing this thesis, there were no data on sIL-6R or sgp130 in CKB or in external publicly available datasets. Therefore, to represent both lower classic and lower trans-signalling in my thesis, I used an IL-6 ligand-based genetic instrument.

Furthermore, although two PheWAS were carried out to assess potential therapeutic repurposing opportunities or adverse effects of the Asp358Ala variant, a limited number of phenotypes were used^{285,286}. For example, Rosa et al assessed associations of the Asp358Ala variant with only 17 traits²⁸⁶. Although Cai et al explored more than 1000 traits, 90% of the participants were men²⁸⁵. And in both of these PheWAS, associations were assessed only in Europeans. The external validity of epidemiological studies (i.e. whether associations observed in the study population are generalisable to other population groups) are as important as internal validity (explained in Section 1.5.2), and, often regarded as part of internal validity²³⁴. Hence, it is important to understand whether the biological and epidemiological associations observed in Europeans are portable to other population groups such as East Asians²³².

The present thesis will carry out an MR analysis and MR-PheWAS in a densely phenotyped cohort of more than 100,000 East Asians who are part of the half a million strong China Kadoorie Biobank, who possess the Asp358Ala variant at a similar allele frequency (39%) as Europeans. Additionally, I will also use a separate instrument representing altered *IL-6* signalling (the altered IL-6 signalling represented by genetic variants in the *IL6* region associated with lower plasma CRP concentrations), to investigate the consistency of effects between instruments representing the receptor and ligand components of the IL-6 signalling pathway.

The next section outlines the aims and objectives of the present DPhil thesis.

Table 1.2: Mendelian randomisation studies of inflammation markers in European population. The table shows studies that have investigated the associations of genetically predicted CRP, fibrinogen and IL-6R using cis-acting genetic instruments on ischaemic stroke and coronary heart disease.

Protein	MR Study design	IS		CHD		Reference
		N	RR (95%CI) ^a	N	RR (95%CI) ^a	
CRP	One sample MR (gene score) ^b	-	-	37,336	1.00 (0.90 - 1.13)	[94]
	Two-sample MR (gene score) ^b	3,548	1.19 (0.93 to 1.53)	60,801	0.98 (0.91 - 1.06)	[287]
	Two-sample MR (gene score) ^b			60,601	0.96 (0.84 - 1.09)	[288]
Fibrinogen	One sample MR (gene score) ^b	-	-	4,685	1.00 (0.95 - 1.04)	[289]
	Meta-analysis	4,712	1.03 (1.00 - 1.07)	40,695	1.00 (0.97 - 1.03)	[290]
Interleukin-6 receptor	Two-sample MR (functional SNP)	6,904	0.98 (0.95 - 1.00)	25,458	0.95 (0.93 - 0.97)	[97]
	Two-sample MR (functional SNP)	-	-	51,441	0.966 (0.950 - 0.982)	[96]

CHD: Coronary heart disease; IS: Ischaemic stroke; N: Number of people; RR: Relative risk; CI: Confidence Interval
^aStudy [94] reported risks for 1 SD higher genetically predicted log CRP; Studies [287] and [288] reported risks for 1 mg/L increase in genetically predicted log CRP. The rest of the studies reported risks per minor allele.

^bGene score refers to multiple cis-acting genetic instruments used by each of these studies

1.7 Thesis Aims & Objectives

The overarching aim of my DPhil research is to understand the observational and causal roles of different components of systemic inflammation in cardiovascular disease.

To represent systemic inflammation, I will use data from the China Kadoorie Biobank for CRP, fibrinogen, IL-6 protein and IL-6R. As outlined in Table 1.1, there is an evidence gap with regards to the associations of CRP in East Asian populations (traditionally having lower LDL-C than Europeans) with both stroke types, including ICH, due to low sample sizes. Further, as shown in Table 1.2, to date the largest IL-6 signalling studies have used *IL6R* to represent IL-6 signalling; however, *IL6R* strictly represents IL-6R mediated signalling (Fig 1.9). Whether associations of altered IL-6R signalling are equivalent to lower IL-6 protein activity remains unclear.

Hence, the specific objectives of my thesis are the following:

-
1. To investigate observational associations of CRP and fibrinogen with major cardiovascular disease in East Asians (Chapter [3](#)).
 2. To assess associations of altered IL-6R and IL-6 signalling on cardiovascular diseases using separate genetic instruments for IL-6R and IL-6 protein (Chapter [4](#)).
 3. To evaluate associations of altered IL-6R and IL-6 signalling on pre-specified non-vascular disease outcomes and phenome-wide using separate instruments for IL-6R and IL-6 protein (Chapter [5](#)).

2

Data

To address the research aims outlined in Chapter 1.7, I used the following data sources for analyses:

1. Individual-participant data from China Kadoorie Biobank (CKB)
2. Summary data from the following sources (non-CKB)
 - UK Biobank (UKB, as processed by MRC-Integrative Epidemiology Unit, University of Bristol and Neale Lab)
 - CARDIoGRAMplusC4D Consortium
 - MEGASTROKE
 - International Stroke Genetics Consortium
 - GIANT consortium

Brief descriptions of the non-CKB data sources and outcome definitions used are provided in Table 2.1 and Section 2.1.2 below. Detailed descriptions of these data sources are provided in their respective papers^{291–296}.

2.1 Non-CKB data

2.1.1 Data sources

Where possible, I assessed and used traits from publicly available data sources with the largest sample sizes for that trait. In most cases, the quantitative traits were from UKB and disease traits were obtained from studies that meta-analysed several large studies, including, in some cases, using data from UKB (Table 2.1). The UKB is a large ($n = \sim 500,000$), population-based, prospective study, that enables detailed investigations of the genetic and nongenetic determinants of chronic diseases in the European population. Follow-up of UKB study participants was via electronic linkages to routinely available national datasets. Details of the UKB study are provided elsewhere²⁹³. The summary statistics of UKB were processed by MRC-IEU and Neale lab, which were used for analyses in the present thesis.

Table 2.1: Selected outcomes from non-CKB data sources for trans-ethnic meta-analysis

Phenotype	Data source	Population	Sample size	Cases	Controls	Units
CVD outcomes						
CHD	CARDIoGRAMplusC4D [292]	Europeans	-	113,937	339,115	Binary
MI	CARDIoGRAMplusC4D [297]	Europeans	-	43,676	128,199	Binary
IS	MEGASTROKE [291]	Europeans	-	34,217	406,111	Binary
ICH	ISGC [294]	Europeans	-	3,226	3,742	Binary
CVD risk factors^a						
SBP	UKB [293]	Europeans	436,419	-	-	SD (mmHg)
DBP	UKB [293]	Europeans	436,424	-	-	SD (mmHg)
HR	UKB [293]	Europeans	436,424	-	-	SD (beats/min)
BMI	GIANT + UKB [295]	Europeans	656,398	-	-	SD (kg/m ²)
Lipid outcomes^b						
Lp(a)	UKB [296]	Europeans	273,896	-	-	SD (mg/dL)
HDL-C	UKB [296]	Europeans	315,133	-	-	SD (mg/dL)
Apo-A1	UKB [296]	Europeans	313,387	-	-	SD (mg/dL)
LDL-C	UKB [296]	Europeans	343,621	-	-	SD (mg/dL)
Apo-B	UKB [296]	Europeans	342,590	-	-	SD (mg/dL)
TG	UKB [296]	Europeans	343,992	-	-	SD (mg/dL)
TC	UKB [296]	Europeans	344,278	-	-	SD (mg/dL)

Apo-A1: Apolipoprotein-A1; Apo-B: Apolipoprotein-B; BMI: Body-mass index; CHD: Coronary Heart Disease; CVD: Cardiovascular disease; DBP: Diastolic Blood Pressure; HDL-C: high density lipoprotein – cholesterol; HR: Heart rate; ICH: Intracerebral haemorrhage; IS: Ischaemic stroke; LDL-C: low density lipoprotein – cholesterol; Lp(a): Lipoprotein (a); TC: Total cholesterol; TG: Triglycerides.

^aUKB estimates as processed by MRC-Integrative Epidemiology Unit, Bristol University (<http://gwas.mrcieu.ac.uk/>)

^bUKB estimates as processed by Neale Lab [296]

2.1.2 CVD outcome definitions in non-CKB data sources

Coronary Heart Disease

Coronary Heart Disease phenotype was variably defined by participating studies in CARDIoGRAMplusC4D but broadly identified participants who reported ischaemic heart disease, myocardial infarction, angina, and underwent any coronary revascularization procedure. The data used for CHD meta-analysis with equivalent phenotype - Major Coronary Events (MCE) - was obtained from meta-analysed estimates reported by Nelson et al²⁹² and were downloaded from www.CARDIOGRAMPLUSC4D.ORG.

Myocardial infarction

Myocardial infarction phenotype was variably defined by participating studies in CARDIoGRAMplusC4D but included participants who have had an episode of ST-segment elevated MI or non-ST-segment elevated MI with elevated troponin levels. The data used for MI meta-analysis with equivalent CKB phenotype was obtained from meta-analysed estimates reported by Nikpay et al²⁹⁷ and were downloaded from www.CARDIOGRAMPLUSC4D.ORG.

Ischaemic Stroke

Stroke was defined according to the World Health Organization (WHO), i.e. rapidly developing signs of focal (or global) disturbance of cerebral function, lasting more than 24 hours or leading to death with no apparent cause other than that of vascular origin.

Ischaemic stroke was variably defined (using clinical/imaging criteria) by participating studies in MEGASTROKE but broadly identified individuals who have neurological symptoms consistent with the WHO definition of stroke that is supported by ischaemic changes observed in magnetic resonance imaging (MRI) of the brain or computer tomographic (CT) scan of the head. The data used for ischaemic stroke meta-analysis with equivalent CKB phenotype was obtained from meta-analysed estimates reported by Malik et al²⁹¹ and were downloaded from www.MEGASTROKE.ORG.

Intracerebral haemorrhage (a subtype of haemorrhagic stroke)

Intracerebral haemorrhage was variably defined (using clinical/imaging criteria) by participating studies in MEGASTROKE but broadly identified individuals who have neurological symptoms consistent with the WHO definition of stroke that is supported by haemorrhagic changes observed in MRI of the brain or CT scan of the head. The data used for ischaemic stroke meta-analysis with equivalent CKB phenotype was obtained from meta-analysed estimates reported by Woo et al²⁹⁴ and were downloaded from International Stroke Genetics Consortium www.CEREBROVASCULARPORTAL.ORG.

A brief description of CKB, the primary data source for individual-level data, is provided below along with details of the study population and statistical methods used for observational and Mendelian randomisation analyses.

2.2 CKB data

2.2.1 Study population

China Kadoorie Biobank was established to enable the investigation of the role of genetic and environmental factors of chronic diseases. It is a prospective study of 512,715 Chinese adults, between the ages of 30-79 years at enrolment. For the baseline survey, eligible study participants from 10 diverse regions (5 urban and 5 rural) of China identified through public registry records were sent invitation letters by health workers or local community leaders. The overall response rate was about 30%. Recruitment began on 25 June 2004 and ended on 15 July 2008. The CKB cohort has been extensively described elsewhere²⁹⁸. Ethical approval for CKB was obtained from the University of Oxford, the Chinese Centre for Disease Control and Prevention (CCDC) and the regional CDC from the 10 study areas. Release 15 of the CKB database was used for all analyses.

2.2.2 Data Collection

Following informed written consent, study participants (at baseline and resurveys of approximately 5%) were interviewed on socio-demographics (e.g. age, occupation), lifestyle (e.g. smoking, alcohol consumption) and past medical history (e.g. diabetes, hypertension). Anthropometric (e.g. height, weight, hip and waist circumferences) and physiological measurements (e.g. blood pressure) were obtained by trained health workers using calibrated instruments and following a standard protocol. A 10 mL non-fasting blood sample (with last meal time recorded) was obtained from all participants at baseline and a subset of study participants during resurvey. The blood samples were kept in a cool box for a few hours before transfer to the local study laboratory for immediate processing. Four vials (three plasma and one buffy coat) from each blood sample were stored in a -40°C freezer for 3–4 months, before being couriered on dry ice to the central blood repository in Beijing and cryopreserved at -80°C . Clinical biochemistry, metabolomics and proteomics data were obtained for subsets of participants. The second resurvey was carried out between August 4, 2013, and September 18, 2014, and included ultrasound measurements (GM-72P00A; Panasonic Healthcare Co, Ltd) of carotid intima-media thickness and plaque. Further descriptions of survey and data collection methods are provided elsewhere²⁹⁹.

2.2.3 Follow up for fatal and non-fatal outcome

Study participants were followed up for incident events, using their unique national identification numbers, by electronic linkage to mortality and disease (CHD, stroke, cancer, and diabetes) registries and to a nationwide health insurance system. These events were coded by trained staff blinded to baseline information using the *International Classification of Diseases and Related Health Problems, Tenth Revision* (ICD-10). Specific case-control definitions are provided in Chapter 3.2.3. Medical records for hospitalization due to CVD events have been systematically reviewed by clinicians in Oxford and China; for example, approximately 90% of

stroke diagnoses were supported by computed tomography (CT) scan or magnetic resonance imaging (MRI)⁴¹.

2.2.4 Measurement of biomarkers

Clinical biochemistry, including markers of inflammation (high-sensitivity CRP [hsCRP], and fibrinogen), lipids (total cholesterol, LDL-C, triglycerides [TG], high-density lipoprotein cholesterol [HDL-C]), lipoproteins (lipoprotein(a) [Lp(a)], apolipoprotein A1 [ApoA1] and apolipoprotein B [ApoB]), markers of liver function (albumin, aspartate transaminase [AST], alanine transaminase [ALT] and gamma-glutamyl transferase [GGT]), markers of renal function (creatinine, uric acid and cystatin-C), and total vitamin D, were assayed at the NDPH Wolfson Laboratory, Oxford. Assays were performed using an AU680 clinical chemistry analyser (Beckman-Coulter, UK) with the following exceptions: plasma fibrinogen and cystatin C were measured using a BN Prospec nephelometer analyser (Siemens, UK); and vitamin D was measured using an Access 2 Immunoanalyser (Beckman-Coulter, UK). All assays used standard manufacturers' reagents, calibrators, and settings.

For measurement of metabolites, a high-throughput proton nuclear magnetic resonance (¹H NMR) spectroscopy platform was used to obtain quantitative molecular information on the plasma samples for 147 metabolic measures and 78 derived measures (225 in total), which included lipoprotein subclasses, fatty acids, amino acids, intermediate substrates of glycolysis and ketone bodies³⁰⁰.

Measurements of 92 proteins were carried out using proximity extension assays (PEA) (Olink Proteomics); in PEA, pairs of oligonucleotide-labelled antibodies are used as 'probes' to bind to the target protein. When the two probes are in close contact, and in the presence of DNA polymerase, the oligonucleotides hybridize in a pair-wise manner ('proximity-dependent DNA polymerization event') and generate a unique polymerase chain reaction (PCR) target sequence, that can be detected and quantified using real-time PCR³⁰¹.

2.3 Statistical analysis plan

Details of statistical analyses conducted are provided in individual Chapters 3, 4, and 5. Here I will briefly describe the analysis plan for each of the Chapters.

For analyses in Chapter 3, I used differentially adjusted logistic regression models to estimate associations of usual plasma concentrations of CRP and fibrinogen with risks of major coronary events (MCE), ischaemic stroke (IS), and intracerebral haemorrhage (ICH) using individual-level data from CKB.

In Chapter 4, I used both individual-level (genotype and trait) data from CKB and summary statistics data for corresponding traits in Europeans from publicly available external sources. With CKB data, I first investigated appropriate genetic instruments for the interleukin-6 receptor (IL-6R) and interleukin-6 (IL-6) protein. The sample size of study participants with both genetic data and measurements of plasma IL-6R and IL-6 was extremely small, compared to the number of genotyped participants with measurements of CRP. As plasma concentration of CRP is regulated by IL-6 signalling, I examined for genetic variants in *IL6R* and *IL6* loci that associated with plasma CRP concentrations in CKB to identify appropriate genetic instruments representing altered IL-6R and IL-6 signalling, respectively. In *IL6R* locus, the genetic variant which showed the strongest association with CRP in CKB was in close linkage disequilibrium (LD) at an r^2 of 0.99 with a missense genetic variant (rs2228145) that was known to alter soluble IL-6R abundance. The aim of the IL-6R instrument construction was to find if there were additional SNPs in the *IL6R* locus that associated with CRP independently of rs2228145. To construct the IL-6 instrument (which did not have any functional variants like IL-6R instrument), the *IL6* region was fine-mapped and credible sets of likely causal SNPs were generated; of these, the set of SNPs that explained the most variance in plasma CRP concentrations was used to construct the IL-6 gene score. The IL-6 gene score was used as the IL-6 instrument. To compare associations between the IL-6R and IL-6 instrument in CKB, associations of the IL-6 instrument were compared with the IL-6R instrument for halving of genetically predicted CRP concentrations. To

compare associations between CKB and Europeans, I constructed equivalent IL-6R and IL-6 instruments using summary data of CRP from UKB. In Chapter 4, linear and logistic regression models were used to assess the association of the IL-6R and IL-6 instruments with quantitative and binary CVD traits, respectively.

In Chapter 5, the same methods were used to perform phenome- and proteome-wide associations of the IL-6R and IL-6 instruments, to identify additional repurposing opportunities or potential adverse events. Unless otherwise stated, in all results Chapters, regression models were fitted in each of the 10 CKB regions and adjusted for age, age² and sex, with estimates meta-analysed by fixed effect inverse-variance weighted method.

3

CRP, Fibrinogen and Cardiovascular Disease

3.1 Introduction

Evidence from randomised clinical trials (RCTs) for interleukin-1 β (IL-1 β) and from human genetic studies for interleukin-6 receptor (IL-6R) signalling suggest that IL-1 β and IL-6 signalling play causal roles in cardiovascular disease (CVD)^{96,222,223}. Both IL-1 β and IL-6 also demonstrably induce hepatic synthesis of C-reactive protein (CRP) and fibrinogen^{205,302}. This makes CRP and fibrinogen useful indicators for IL-6 and IL-1 β signalling (discussed in Chapter 1.5.5). This biology has been leveraged, for example, in clinical trials that investigated inhibition of IL-1 β (where direct measurement of IL-1 β can be expensive), assessing IL-1 β activity via measurements of CRP and fibrinogen levels³⁰³ (see Chapter 1.5.3).

Interleukin-1 β /IL-6 led systemic inflammation as represented by CRP and fibrinogen was associated with increased risks of atherosclerosis and atherosclerotic CVDs (mainly CHD)^{92,93,304–310}. However, these studies that investigated risks of CHD with CRP and fibrinogen were largely conducted in European populations, with only a few studies in East Asians^{230,231,311}. As stated in Chapter 1.5.2, differences in

distributions of effect modifiers between population groups can, in principle, affect the generalisability of associations of a risk factor with an outcome observed in only one of the population groups^{233,234}. For example, one potential effect modifier that can affect trans-ethnic generalisability of associations of CRP and fibrinogen with CVD is low-density lipoprotein cholesterol (LDL-C). Higher levels of LDL-C are associated with higher risks of CHD⁵⁵ and levels of LDL-C are much lower in East Asians compared to Europeans (mean ~140 mg/dL vs ~90 mg/dL)^{54,92,93}. This raises the question, if there are trans-ethnic differences in associations of CRP or fibrinogen with CVD, are these differences due to trans-ethnic differences in levels of potential effect modifiers like LDL-C?

Furthermore, compared to studies that have examined associations of CRP and fibrinogen with risk of CHD, observational evidence for the relevance of systemic inflammation to individual stroke types is limited. In particular, there is limited evidence on the associations of CRP and fibrinogen with risk of intracerebral haemorrhage (ICH). Reliable assessment of the association of systemic inflammation with stroke types is of particular relevance to China, where rates of stroke, particularly ICH (incidence/prevalence of ICH in China: 24% and 16% respectively²³⁷), are much higher than for Western populations. Additionally, whether there are any systematic differences in the pattern of association between CRP and fibrinogen with CHD and stroke types in East Asians is largely unknown.

In this Chapter, I investigated the associations of CRP and fibrinogen measured at baseline with incident CVD outcomes in a large cohort of East Asians. For the present study, I used observational data on CRP and fibrinogen assayed in a subset of East Asians individuals ($n = 18,255$) selected from a larger nested CVD case-control study ($n = 22,936$) in CKB. The logic of participant selection for the present study from the nested case-control CVD study is illustrated in Figure 3.1. The CVD outcomes selected are major coronary events (MCE), and both stroke types - ischaemic stroke (IS) and ICH. The definitions of cases and controls are provided in Table 3.2. To assess the magnitude and direction of association of CRP

and fibrinogen with the selected CVD outcomes, I fitted logistic regression models initially adjusting only for age, age², and sex ('base model'). I used age as a linear and quadratic term to account for any non-linearity of association between age and CVD (as has been done in previous studies of CVD^{40,312}). However, the observed associations in the base model may still be confounded by risk factors beyond age and sex (see Chapter 1.3 and Chapter 1.5.2) that are independently associated with CRP/fibrinogen and CVD outcomes. To identify additional risk factors, I first tabulated the distribution of key baseline traits by case-control status (Chapter 3.3.1). The objective of this analysis was to generate insights on whether there were any systematic differences in cases and controls with respect to key baseline traits that have been associated with CVD risk in previous studies (Chapter 1.3). Second, I examined variations of baseline CRP and fibrinogen concentrations adjusted for case status by groups of selected risk factors measured at baseline (Chapter 3.3.2). On identifying the additional risk factors, I compared the shape and strength of the crude association of CRP and fibrinogen with each CVD outcome in the base model with the shape and strength of associations in the adjusted model (i.e. the model that accounts for the additional risk factors) (Chapter 3.3.3). These comparisons were aimed to illustrate the impact of potential confounding by risk factors in the base model. Separately, I also investigated the impact of individual risk factors and physiologically similar risk factor groups (e.g. lipids - LDL-C and triglycerides) on the base model. Finally, I examined whether the adjusted associations of CRP and fibrinogen with MCE, IS and ICH vary by pre-specified subgroups, that would suggest potential effect modification (Chapter 3.3.4).

The work in this Chapter was published in *Scientific Reports* on 2020¹.

¹Karim, M.A., Kartsonaki, C., Bennett, D.A. et al. Systemic inflammation is associated with incident stroke and heart disease in East Asians. *Sci Rep* 10, 5605 (2020). <https://doi.org/10.1038/s41598-020-62391-3>

3.2 Methods

For the present observational study, I used individual-level data from CKB. I have described the study participants in CKB, including details of collection of relevant baseline and follow-up data in Chapter 2 (Data) of this thesis. I have divided this section into five inter-related parts that describe the reasoning I used to select participants for the present study and methods I used for statistical analyses.

3.2.1 Participant selection for the nested case-control study

Participants in the present study were a subset of a larger nested case-control study of incident stroke (IS and ICH) and incident ischaemic heart disease (IHD) for which participants were selected until June 2014 (Figure 3.1). Cases for incident IHD in the nested case-control study included study participants who had fatal MI (i.e. resulting in death due to the CVD event) or non-fatal MI (ICD-10: I21-I23) or fatal IHD (ICD-10: I20-I25). Cases for incident stroke in the nested case-control study were assigned IS if the first stroke event was IS (ICD-10: I63 ‘Cerebral infarction’, I69.3 ‘Sequelae of cerebral infarction’). Stroke cases were assigned ICH if the first stroke event was ICH (ICD-10: I61 ‘Intracerebral haemorrhage’, I69.1 ‘Sequelae of intracerebral haemorrhage’). The controls in the nested case-control study were frequency matched to ICH cases (matched for age, sex and region). To minimize confounding by medication use or prevalent disease status, individuals with a history of CVD, cancer, or statin use at baseline were excluded from among incident cases and controls from the nested case-control cohort.

3.2.2 Clinical chemistry measurements

Clinical chemistry measurements were obtained from a cryopreserved 10 mL non-fasting blood sample (with last meal time recorded) (details provided in Chapter 2.2.2). Although the 10 ml blood samples were obtained from all CKB study participants at baseline, clinical chemistry measurements were conducted for only a subset of the study participants. Furthermore, not all participants in the nested case-control study had clinical chemistry measurements.

Table 3.1: Clinical chemistry measurements for the nested case-control study in China Kadoorie Biobank

Type of biomarker	Biomarker	Measurement
Inflammatory biomarkers	C-reactive protein	AU680 clinical chemistry analyser (Beckman-Coulter, UK)
	Fibrinogen	BN Prospec nephelometer analyser (Siemens, UK)
Lipids	Total cholesterol	AU680 clinical chemistry analyser (Beckman-Coulter, UK)
	Low-density lipoprotein cholesterol	
	Triglycerides	
	High-density lipoprotein cholesterol	
	Lipoprotein(a)	
	Apolipoprotein A1	
	Apolipoprotein B	
Markers of liver function	Albumin	AU680 clinical chemistry analyser (Beckman-Coulter, UK)
	Alanine transaminase	
	Aspartate transaminase	
Markers of renal function	Gamma-glutamyl transferase	AU680 clinical chemistry analyser (Beckman-Coulter, UK)
	Creatinine	
	Uric acid	
Vitamins	Cystatin-C	BN Prospec nephelometer analyser (Siemens, UK)
	Total vitamin D	Access 2 Immunoanalyser (Beckman-Coulter, UK)

Details of clinical biochemistry measurements are provided in Table 3.1. All assays used standard manufacturers' reagents, calibrators, and settings.

3.2.3 Case-control definitions for the present study

In the present study, cases were identified by first CVD event: IS (ICD-10 codes: I63, I69.3), ICH (ICD-10 codes: I61, I69.1) or Major Coronary Event (MCE, defined as defined as acute myocardial infarction [ICD-10: I21-I23] or fatal IHD [ICD-10: I20, I24 or I25] or undergoing a coronary revascularization procedure such as stent placement or coronary artery bypass graft) before 1st January 2017. Because relevant cases were identified by the first CVD event, individuals who had MCE, IS and/or ICH events on the same day or had non-MCE/IS/ICH CVD events (e.g. rheumatoid heart disease) earlier than MCE, IS or ICH during follow-up, were excluded. Only the first CVD events were chosen to minimize the possibility of events arising due to treatment of the first event. For example, ICH can result from anti-thrombolytic or anti-platelet treatment for IS or MCE. Using the first CVD event as ICH ensured that assessment of association of systemic inflammation with ICH was robust to adverse events arising from treatment of MCE or IS.

To minimize selection bias, controls were selected from the same source population as cases (CKB source population described in Chapter 2.2.1) and were common to

Table 3.2: Definitions of Case/Control and exclusions used with corresponding ICD:10 codes (where appropriate)

Case/Control/Exclusion	Disease (ICD-10 code)/Procedure
Major Coronary Events (MCE)	1. Ischaemic heart disease (ICD-10: I20-I25) (when fatal); 2. Acute myocardial infarction (ICD-10: I21); 3. PCI/CABG/stent procedure
Ischaemic stroke (IS)	Cerebral infarction, including sequelae (ICD-10: I63, I69.3)
Intracerebral haemorrhage (ICH)	Non-traumatic intracerebral haemorrhage, including sequelae (ICD-10: I61, I69.1)
Controls	Not a case of MCE, IS, or ICH as defined above
Exclusions	1. Other CVD events that are not MCE, IS, or ICH (ICD-10: I00-I09, I16-I19, I27-I60, I62, I64-I69 [except I69.1 and I69.3], I70-I88, I95-I99; I10-I15); 2. Same day MCE & IS or MCE & ICH or IS & ICH event

MCE, IS and ICH cases; study participants identified as controls had no CVD event before 1st January 2017. All endpoint definitions are provided Table 3.2.

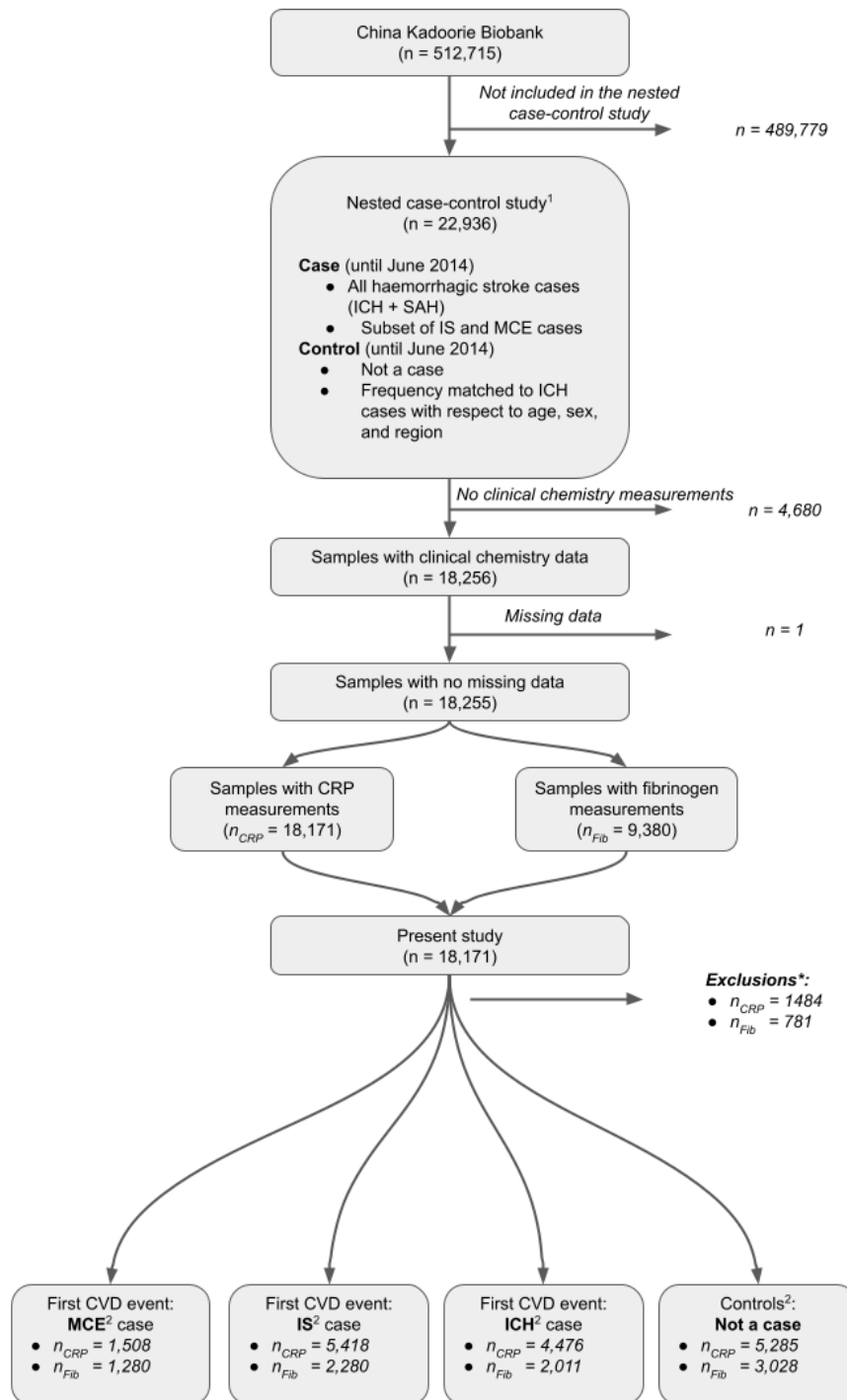
3.2.4 Participants for the present study

For the present study, I selected participants from the nested case-control study that had clinical chemistry measurements. From the subset that had clinical chemistry measurements, I used the case status of study participants as of 1st Jan 2017, not the time of ascertainment for the nested case-control (June 2014). Hence, some controls in the nested case-control study became cases in the present study if an event occurred prior to 1st Jan 2017.

Among study participants, baseline CRP (fibrinogen) measurements were available for 1,508 (1,280) participants with MCE as the first CVD event, 5,418 (2,280) participants with IS as the first CVD event, 4,476 (2,011) participants with ICH as the first CVD event, and 5,285 (3,028) participants with no CVD event before 1st January 2017. About 1,484 (781) individuals who had CVD events other than the ones defined in Table 3.2 or had same day MCE/IS/ICH events were excluded from the present analyses. One individual with missing data for BMI was excluded from all analyses. (Figure 3.1).

Because single CRP and fibrinogen measurements may not accurately reflect the usual or average levels of the CRP or fibrinogen (due to random measurement errors or biological fluctuations over time), I used data on repeat measurements of CRP and fibrinogen from a randomly selected subset of individuals for whom plasma

samples were collected during the second resurvey, approximately 10 years after the baseline survey, between 2013-2014 (hsCRP, n=1,296; fibrinogen, n=639).



CRP: C-reactive protein; CVD: Cardiovascular disease; ICD: International Classification of Diseases; ICH: Intracerebral haemorrhage; IHD: Ischaemic heart disease; IS: Ischaemic stroke; MCE: Major Coronary Event; TIA: transient ischaemic attack

Figure 3.1: Flowchart of participant selection.

Flowchart shows the selection of study participants for the present analysis, starting with the set-up of the nested case-control study in China Kadoorie Biobank to selecting individuals with clinical chemistry data on CRP and fibrinogen to selection of relevant cases and controls after necessary exclusions.

¹The nested study excluded individuals with history of CVD, cancer or statin use at baseline

²Defined in Table 3.2

3.2.5 Statistical Analysis

For all analyses, data for CRP, fibrinogen, and selected biomarkers (TG, AST, ALT, GGT, Lp(a), creatinine, cystatin-C) were log-transformed to obtain near-Gaussian distributions (Appendix Figure A.1). The standard deviation for baseline $\log(e)$ -transformed CRP and fibrinogen concentration was 1.19 and 0.23 respectively, equivalent to a 3.3 fold (i.e. $e^{1.19}$) and 1.3 fold (i.e. $e^{0.23}$) differences in mg/L and g/L on the original scale of CRP and fibrinogen measurements respectively.

For assessment of baseline characteristics, adjusted means and proportions of baseline characteristics for cases and controls, and cross-sectional mean levels of CRP and fibrinogen for categories (or quintiles) of selected CVD risk factors, were standardized (as appropriate) to age, sex, study area, and case status using R package *lsmeans*³¹³ with P values for association derived from linear regression models (Wald test). To assess cross-sectional associations of quantitative traits (standardized using the R `scale` function) per standard deviation (SD) of log-transformed CRP and fibrinogen, I used linear regression, adjusted for age and sex, and stratified by study area, combining the area-specific estimates with inverse variance-weighted fixed-effect meta-analysis using function `rma` from R package *metafor*³¹⁴.

To estimate odds ratios (ORs) for the associations of CRP and fibrinogen with risk of incident MCE, IS and ICH, logistic regression models were used with adjustment for age, age², and sex (base model), and stratification by study area with inverse variance-weighted fixed effect meta-analysis. The base model was further adjusted (individually or in combination) for BMI, SBP, LDL-C, TG, and additional CVD risk factors: household income (in yuan: <2500, 2500-4999, 5000-9999, 10,000-19,999, 20,000-34,999, >35,000); occupation (agriculture, factory worker, administrator/manager, professional/technical, sales and service, retired, house-wife/husband, self-employed, unemployed, other or not stated); education (no formal school, primary school, middle school, high school, technical school/college, university); standing height (meters); self-reported or screen-detected diabetes

status; physical activity (metabolic equivalent of task hours per day, MET h/day); smoking (never smoker, occasional smoker, ex-regular smoker, smoker); and alcohol consumption (never regular drinker, ex-regular drinker, occasional or seasonal drinker, monthly drinker, reduced intake, weekly drinker)⁹⁹. Where models failed to converge due to very low numbers of cases, those study area(s) or strata were excluded from the analysis. The Wald χ^2 statistic (with one degree of freedom) was calculated using function *anova.rma* from R package *metafor* to assess the improvement in model fit following inclusion of additional covariates. Heterogeneity and trend tests were conducted by χ^2 tests applied to the log ORs and their SEs. For quintiles of CRP and fibrinogen, floating CIs were estimated using the Plummer method³¹⁵ to provide 95% CI for all quintiles including the reference quintile, enabling inter-quintile comparisons.

Regression calibration was used to adjust for within-person variation in CRP and fibrinogen. Regression coefficients and predicted usual values were estimated from linear regression of log-transformed CRP or fibrinogen levels from resurvey samples (taken median 5 years after baseline) on the corresponding baseline values for the same participants. Quintile plots used predicted usual values for CRP and fibrinogen levels; adjusted log ORs per SD of usual log CRP and fibrinogen levels and corresponding standard errors were calculated by dividing those derived using baseline CRP and fibrinogen levels by the corresponding regression coefficient³¹⁶.

Where stated, significance thresholds for *P* values from multiple hypothesis testing were corrected using the Bonferroni method. Analyses were performed using R version 3.5.0.

3.3 Results

3.3.1 Baseline characteristics of participants

Systematic differences in baseline traits between cases and controls can lead to differences in risk of outcomes that are not necessarily due to differences in levels of

risk factor under investigation.

To understand whether there are any systematic differences in key baseline traits (selected based on previous studies of CRP and CVD^{317,318}, see Table 1.1) between MCE/IS/ICH cases and controls, I derived means and proportions of these traits standardised to age, sex, study area and case status, where appropriate. The means and proportions were standardised to ensure that any systematic differences observed between cases and controls would not be due to differences in age, sex, study area or case status (statistical procedure to obtain standardised means and proportions are detailed in Chapter 3.2.5).

The baseline characteristics of the study participants are shown in Table 3.3. Controls (mean age 55.8 years, 49.8% female) were older and had a higher proportion of men than what was previously reported for the overall CKB population²⁹⁹; these age and sex distribution differences of the present study from the overall CKB population reflected the structure of the larger nested case-control cohort from which the present case-control study population was drawn.

Cases of IS (55.4 years) and MCE (56.5 years) were similar in age to controls but ICH cases (59.3 years) were older. With regards to case-control distribution between men and women, the proportion of women was similar between controls (49.6%) and ICH cases (49.9%) but a lower proportion of women had MCE (41.6%) whilst a higher proportion had an IS event (55.6%), compared to controls. The reason for this sex discrepancy between case and controls is unclear.

Baseline systolic blood pressure (SBP, mmHg) was higher in MCE (143.6) and IS (144.4) cases than in controls (134.4), but was highest in ICH cases (152.9). Compared to controls, larger proportions of MCE and stroke cases had reported at baseline, poor health (MCE: 18%; IS: 15%; ICH: 16%; controls: 8%), hypertension (MCE: 22%; IS: 24%; ICH: 26%; controls: 10%), and (screen-detected or self-reported) diabetes (MCE: 14%; IS: 12%; ICH: 9%; controls: 6%). This was expected, given that hypertension and diabetes are known risk factors of MCE and

stroke, and having hypertension and diabetes may lead participants to self-report a state of poor health. There were, however, only small differences between cases and controls with respect to adiposity traits (e.g. BMI kg/m², MCE: 23.9; IS: 24.1; ICH: 23.8; controls: 23.4) or levels of self-reported physical activity (MET h/day, MCE: 18.7; IS: 18.7; ICH: 19.0; controls: 20.1) presumably due to the smaller sample size of the present study (since the present study selected individuals with clinical chemistry measurements).

Baseline levels of markers of inflammation (CRP and fibrinogen), lipids (LDL-C and TG), and subclinical atherosclerosis (proportion of participants with plaque, carotid intima media thickness) were higher in participants with incident MCE and IS (i.e. atherosclerotic CVD) than in controls. By contrast, differences in these traits were smaller (for CRP and plaque) between ICH cases and controls. The systematic differences in biomarker concentrations and markers of subclinical atherosclerosis between MCE/IS and ICH when compared to controls was expected as it likely reflects the predominant atherosclerotic processes involved in the pathogenesis of MCE and IS compared to ICH which can have both atherosclerotic and non-atherosclerotic processes¹⁹².

Table 3.3: Baseline characteristics of study participants.

The table shows the baseline characteristics of study participants who had incident MCE, IS or ICH compared with controls who had no CVD event until 1st Jan 2017.

Trait	Controls (n=5285)	MCE (n=1508)	IS (n=5418)	ICH (n=4476)
Age and Sociodemographic factors				
Age ^a , y	55.8 (10.9)	56.5 (9.72)	55.4 (9.36)	59.3 (10.1)
Female ^b , %	49.6	41.6	55.6	49.9
Rural ^c , %	79.2	66.6	54.6	79.5
Middle School and above, %	38.3	36.8	38.2	36.0
Income >10,000 yuan/year, %	62.2	61.5	65.6	62.5
Lifestyle factors				
Male smokers, %	87.1	90.9	89.5	86.9
Female smokers, %	3.9	7.8	5.4	5.9
Male regular drinkers, %	32.1	26.5	30.9	33.1
Female regular drinkers, %	2.1	1.8	2.2	2.1
Physical activity, MET h/day	20.1 (11.8)	18.7 (11.6)	18.7 (11.5)	19.0 (11.6)
Anthropometric, blood pressure and circulatory measures				
Standing height, m	1.59 (0.06)	1.59 (0.06)	1.59 (0.06)	1.58 (0.06)
BMI, kg/m ²	23.4 (3.1)	23.9 (3.4)	24.1 (3.4)	23.8 (3.4)
Waist-hip ratio	0.89 (0.07)	0.91 (0.07)	0.90 (0.07)	0.90 (0.07)
SBP, mmHg	134.4 (19.7)	143.6 (25.1)	144.4 (23.9)	152.9 (27.3)
DBP, mmHg	78.5 (11.0)	83.0 (13.3)	83.9 (12.5)	88.5 (14.2)
Heart rate, per minute	78.3 (11.7)	80.4 (12.7)	79.2 (12.1)	79.9 (12.6)
Markers of subclinical atherosclerosis				
Plaque, %	39.9	60.5	50.7	44.4
cIMT, mm	0.71 (0.13)	0.78 (0.23)	0.75 (0.16)	0.72 (0.16)
Inflammatory biomarkers				
CRP ^d , mg/L	0.90 (0.86-0.93)	1.31 (1.24-1.39)	1.12 (1.08-1.16)	1.08 (1.04-1.13)
Fibrinogen ^d , g/L	3.00 (2.98-3.03)	3.07 (3.04-3.11)	3.05 (3.02-3.08)	3.02 (2.99-3.05)
Lipids				
Triglycerides ^d , mg/dL	145.8 (143.1-148.5)	163.4 (158.5-168.5)	156.6 (153.9-159.45)	145.5 (142.7-148.5)
LDL cholesterol ^e , mg/dL	91.6 (23.9)	95.9 (29.1)	95.9 (25.8)	90.8 (26.4)
Self-reported health and disease				
Poor self-rated health, %	8.1	18.4	14.6	15.5
Hypertension, %	10.4	21.7	23.7	26.4
Diabetes, %	6.0	14.0	12.4	9.3

CRP: C-reactive protein; cIMT: carotid intima-media thickness; DBP: Diastolic Blood Pressure; ICH: Intracerebral haemorrhage; IS: ischaemic stroke; LDL: Low-density lipoprotein; MCE: Major Coronary Event; MET: metabolic equivalents; SBP: Systolic Blood Pressure;

Values are mean (standard deviation) for continuous traits or proportion (weighted percentage) for binary/categorical traits, standardized to age, sex and study area of case-control cohort unless otherwise stated.

^aStandardized to sex and study area only

^bStandardized to age and study area only

^cStandardized to age and sex only

^dGeometric mean (95% CI)

^eLow-density lipoprotein (LDL) cholesterol was directly measured

3.3.2 Cross-sectional associations of CRP and fibrinogen with selected CVD risk factors

To identify confounders, I assessed the cross-sectional associations of CRP and fibrinogen with CVD risk factors in CKB. I have divided this assessment into three parts illustrated by the three figures in this section:

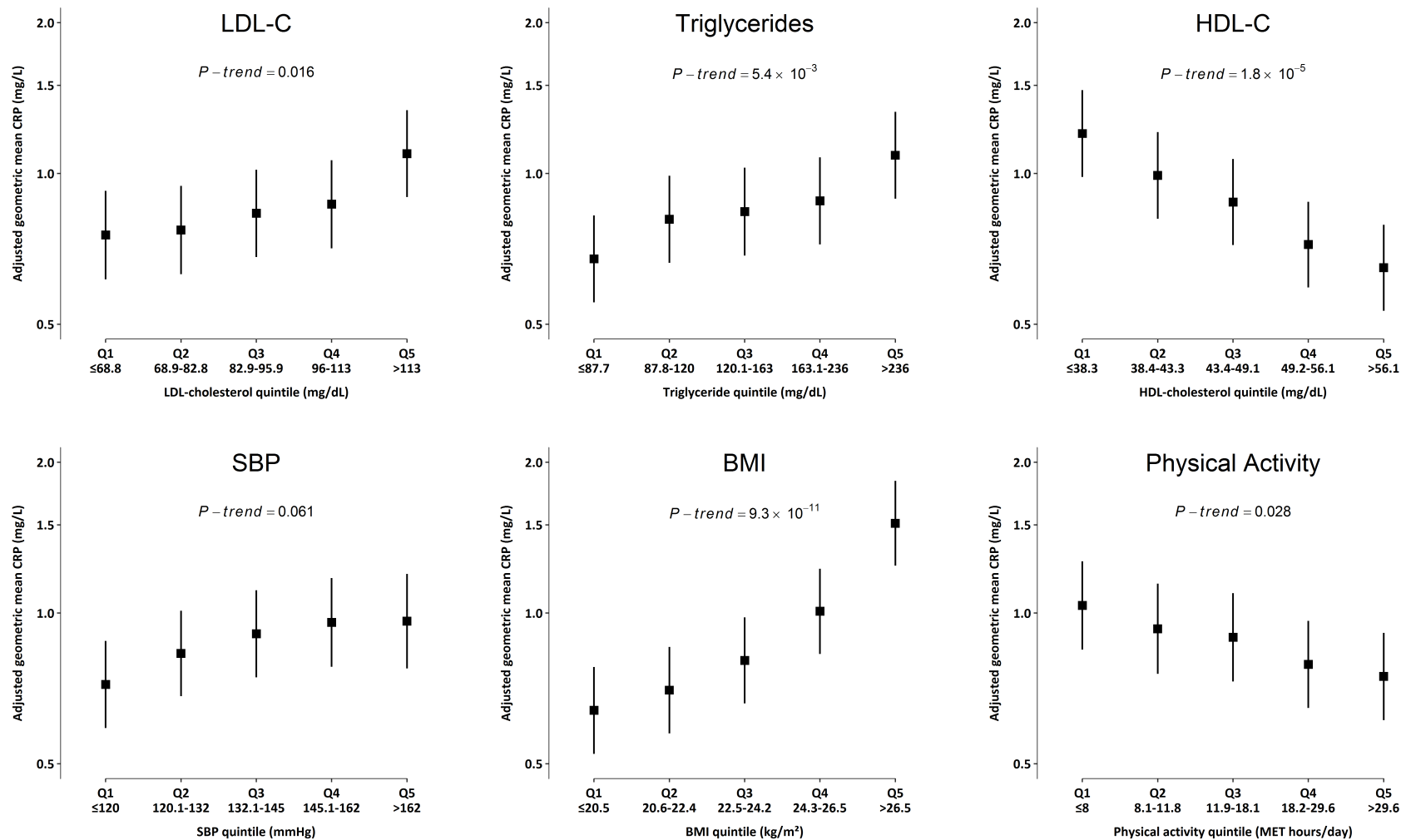
1. Shape of cross-sectional association of CRP with six well-established CVD risk factors.

The six risk factors were LDL-C, TG, HDL-C, SBP, BMI and physical activity. These risk factors were selected based on their previously reported associations with CRP and CVD (see Table 1.1 for a list of these studies). Additional risk factors, e.g. Apolipoprotein A1 (Apo-A1) and B (ApoB) were not included in the present analysis (but included in the next section) due to the correlation of Apo-A1 and ApoB to HDL-C³¹⁹ and LDL-C³²⁰ respectively, and HDL-C and LDL-C have been selected for this assessment.

If this assessment was done in both cases and controls, the observed association between CRP and the risk factors may be due to the higher levels of these risk factors in cases than controls (see Table 3.3). Hence, I conducted this analysis only in controls. Because a larger number of controls had measurements of CRP than fibrinogen (5,285 vs. 3,028) and there was more evidence of association of these risk factors with CRP than fibrinogen from prior studies^{92,321}, the present assessment was limited to CRP. Cut-offs for quintiles of each of the six selected quantitative risk factors were based on their distribution in controls only, and estimates were adjusted only for age, age², sex and region. *P-trend* was estimated using chi-square for linear trend with one degree of freedom and tests the null hypothesis of no linear trend between CRP and quintiles of quantitative traits (Chapter 3.2.5) .

C-reactive protein showed strong dose-response relationships with quintiles of five out of the six CVD risk factors. There were positive associations of CRP with LDL-C (*P-trend*=0.016), TG (*P-trend*=5.3 x 10⁻³), and BMI (*P-trend*=9.3 x 10⁻¹¹).

A positive but weaker association of CRP was observed with SBP (P -trend=0.06). Inverse dose-response associations of CRP were observed with HDL-C (P -trend=1.8 x 10⁻⁵) and physical activity (P -trend=0.028) (Figure 3.2). This suggests that, in the present cohort, higher CRP concentration is linked to higher levels of atherogenic lipids (LDL-C and TG) and adiposity (BMI) but with lower levels of HDL-C and physical activity.



BMI: Body-mass index; CRP: C-reactive protein; HDL-C: High-density lipoprotein cholesterol; LDL-C: Low-density lipoprotein cholesterol; SBP: Systolic Blood Pressure;

Figure 3.2: Cross-sectional associations of C-reactive protein with selected traits.

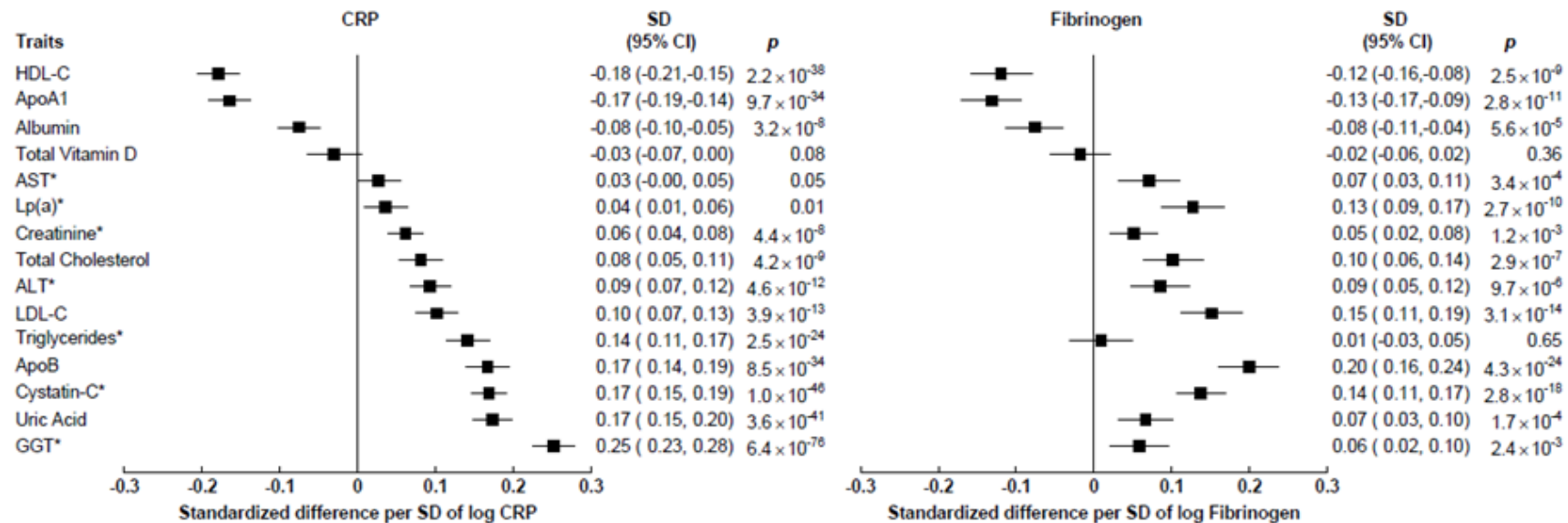
The figure demonstrates dose-response relationships of CRP with CVD risk factors selected on the basis of prior evidence. Analyses were carried out only in controls ($n=5,285$). Cut-offs for biomarker quintiles were based on their distribution in controls only. Geometric mean of CRP was adjusted for age, age², sex and region. $P\text{-trend}$ tests the null hypothesis of no trend and was calculated using chi-square test statistic with 1 degree of freedom. Error bars represent 95% CI for each quintile

2. Cross-sectional associations of CRP/fibrinogen with other clinical chemistry measurements

In this assessment, I included fibrinogen along with CRP as markers of systemic inflammation. The objective here was to gain a broader understanding of the pattern of relationships of CRP and fibrinogen with clinical chemistry measurements; some of these clinical chemistry measurements were used as biomarkers for other chronic diseases (e.g. cystatin-C for chronic kidney disease³²²). Like Figure 3.2, these associations were assessed only in controls to assess associations of CRP and fibrinogen with clinical chemistry measurements independent of the association of these quantitative traits with CVD. To assess association of baseline CRP/fibrinogen with baseline clinical chemistry measurements, I regressed each clinical chemistry measurement on CRP or fibrinogen, adjusting the model for only age, age², and sex. For each biomarker, linear regression models were fitted for each region and estimates per SD increase of log CRP were meta-analysed across regions using a fixed-effect modelling with inverse variance weighting. The corresponding *P*-values were extracted from the meta-analytic procedure and tests the null hypothesis of no association (Chapter 3.2.5).

C-reactive protein showed strong inverse associations with HDL-C, Apolipoprotein-A1, and albumin (Figure 3.3). The association estimates (SD (95% CI)) for HDL-C, Apo-A1, and albumin per SD increase of log CRP were -0.18 (-0.21, -0.15), -0.17 (-0.19, -0.14), and -0.08 (-0.10, -0.05) respectively. Strong positive associations of CRP were observed with LDL-C, triglycerides and apolipoprotein-B with effect estimates (SD (95% CI)) for each of these per SD increase of log CRP were 0.10 (0.07, 0.13), 0.14 (0.11, 0.17), 0.17 (0.14, 0.19) respectively. C-reactive protein also exhibited positive associations with markers of liver (AST, ALT, and GGT) and renal function (creatinine, uric acid, and cystatin-C). I observed a similar pattern of associations (i.e. similar to CRP) of fibrinogen for all clinical chemistry measurements, except triglycerides; there was no evidence of association of fibrinogen with triglycerides.

These analyses showed that CRP and fibrinogen concentrations were broadly similarly associated with a range of biomarkers that included lipids, lipoproteins, markers of liver and renal function. Some of these biomarkers such as LDL-C are also strongly associated with CVD risk and are, thus, adjusted for in models investigating associations of CRP and CVD in the present study.



ApoA1: Apolipoprotein-A1; ApoB: Apolipoprotein B; ALT: Alanine transaminase; AST: Aspartate transaminase; CRP: C-reactive protein; GGT: Gamma-glutamyl transferase; HDL-C: High-density lipoprotein cholesterol; LDL-C: Low-density lipoprotein cholesterol; Lp(a): Lipoprotein (a);

Figure 3.3: Cross-sectional associations of CRP and Fibrinogen with biomarkers.

The figure demonstrates cross-sectional associations of CRP and fibrinogen with a range of clinical chemistry measurements. Analyses were carried out only in controls. SD (95% CI) estimates of CRP and fibrinogen were adjusted for age, age², sex and stratified and meta-analysed (fixed-effects inverse variance method) by region. *P* for each trait was extracted from the regression model. Error bars represent 95% CI.

3. Cross-sectional associations of CRP/fibrinogen with categorical traits

There are several categorical traits ascertained or measured at baseline that can provide insights on the pattern of variation between CRP and fibrinogen across different categories of the same trait. For this assessment, I explored associations of CRP and fibrinogen with a wide range of categorical baseline traits. As before, these traits were selected based on evidence from previous studies that either examined association of CRP or fibrinogen with CVD risk adjusting for these traits or studies that investigated these traits as risk factors for CVD in general.

For these associations, I regressed CRP or fibrinogen on the categorical traits adjusted for age, sex, study area and case status, where appropriate. From the linear regression models constructed, I extracted the adjusted geometric means of CRP and fibrinogen from each category of the trait using the *lsmeans* R package (Chapter 3.2.5). The P value tests the null hypothesis that, for ordered categorical traits, there is no evidence of trend of CRP/fibrinogen across the categories or, for discrete categories, there is no evidence of heterogeneity in the variation of CRP/fibrinogen across categories.

All trait categories included both males and females, except sex, smoking, drinking, and standing height. There was no evidence of variation ($P = 0.3$) of geometric mean [95% CI] CRP levels (mg/L) between males (1.06 [1.03, 1.09]) and females (1.04 [1.01, 1.07]), but a difference between sex was observed for fibrinogen ($P = 8 \times 10^{-56}$); in males, the geometric mean fibrinogen (g/L) concentrations in males was lower (2.93 [2.90, 2.95]) than females (3.15 [3.12, 3.17]). Due to the lower number of women who smoke (3% women vs. 68% men in CKB reported to smoke regularly³²³) and drink (2% women vs. 33% men in CKB reported to drink at least once a week³²⁴) in CKB, I included only men while examining the variation of CRP and fibrinogen concentrations among smokers and drinkers. I observed higher fibrinogen levels in men who consumed alcohol less than weekly (2.98 [2.95, 3.02]) than those who consumed alcohol weekly (2.80 [2.76, 2.84]) ($P = 1.51 \times 10^{-16}$). There was no evidence of variation of CRP or fibrinogen levels across smoking categories in men

($P > 0.05/32$, accounting for the total number of traits examined in this assessment for CRP and fibrinogen). Due to substantial variations of standing height between males and females in CKB²⁹⁹, I examined standing height separately in males and females. Like smoking, there was no evidence of trend across categories of standing height in either sex.

When examining variation of CRP and fibrinogen from youngest to oldest, I observed evidence of a strong trend with age, with differences between the youngest (30 - 49 years) and oldest (60 and over) age categories of 0.50 mg/L ($P=1.16 \times 10^{-126}$) for CRP and 0.61 g/L ($P=5.42 \times 10^{-307}$) for fibrinogen. The trend of higher CRP and higher fibrinogen with increasing age was consistent with variations observed in other traits correlated with age. For example, aging is correlated to lower physical activity, and there was evidence of trend of lower CRP and fibrinogen with higher levels of self-reported physical activity (CRP: $P=1.42 \times 10^{-27}$; fibrinogen: $P=1.45 \times 10^{-7}$). Similarly, CRP and fibrinogen levels (CRP: $P=9.02 \times 10^{-19}$; fibrinogen: $P=2.49 \times 10^{-5}$) were observed to be lower in manual occupation workers (CRP: 0.97 [0.94, 1.00]; fibrinogen: 3.01 [2.98, 3.03]) than non/semi-manual workers (CRP: 1.18 [1.14, 1.23]; fibrinogen: 3.08 [3.05, 3.11]), and this may be due to the higher physical activity in manual work occupations compared to non/semi-manual workers. C-reactive protein and fibrinogen were also observed to be higher in individuals with diabetes (CRP: $P=6.12 \times 10^{-46}$; fibrinogen: $P=1.26 \times 10^{-12}$), higher with increasing BMI categories (CRP: $P=9.45 \times 10^{-218}$; fibrinogen: $P=5.86 \times 10^{-6}$), higher SBP (CRP: $P=9.34 \times 10^{-34}$; fibrinogen: $P=2.09 \times 10^{-7}$) and higher LDL-C tertiles (CRP: $P=3.12 \times 10^{-40}$; fibrinogen: $P=1.60 \times 10^{-51}$).

Table 3.4: Geometric mean and 95% confidence interval (CI) of C-reactive protein (CRP) and Fibrinogen by key baseline characteristics.
The table shows cross-sectional associations of CRP and fibrinogen across categories of potential confounders

Baseline traits	N	CRP, mg/l ^a	P*	N	Fibrinogen, g/l ^a	P*
Age Categories, years^b						
30-49	4024	0.76 (0.73-0.78)	1.16 x 10 ⁻¹²⁶	3279	2.70 (2.68-2.72)	5.42 x 10 ⁻³⁰⁷
50-59	5387	1.01 (0.98-1.04)		2518	3.06 (3.03-3.08)	
60 and over	7275	1.26 (1.23-1.29)		2801	3.31 (3.29-3.34)	
Gender^c						
Male	8463	1.06 (1.03-1.09)	0.3	4426	2.93 (2.90-2.95)	8.00 x 10 ⁻⁵⁶
Female	8224	1.04 (1.01-1.07)		4173	3.15 (3.12-3.17)	
Region^d						
Urban	4987	1.18 (1.14-1.23)	5.32 x 10 ⁻¹⁸	2486	3.01 (2.98-3.04)	5.03 x 10 ⁻⁰²
Rural	11700	0.99 (0.97-1.02)		6113	3.04 (3.02-3.06)	
Education						
Below Middle School	10192	1.06 (1.03-1.09)	0.52	4774	3.04 (3.02-3.06)	0.25
Middle school and above	6495	1.04 (1.00-1.08)		3825	3.02 (2.99-3.05)	
Income (annual)						
<10,000 yuan	6480	1.03 (1.00-1.07)	0.25	3227	3.03 (3.01-3.06)	0.81
>10,000 yuan	10207	1.06 (1.03-1.09)		5372	3.03 (3.01-3.05)	
Occupation						
Non/Semi-manual	6904	1.18 (1.14-1.23)	9.02 x 10 ⁻¹⁹	3198	3.08 (3.05-3.11)	2.49 x 10 ⁻⁰⁵
Manual	9783	0.97 (0.94-1.00)		5401	3.01 (2.98-3.03)	
Drinking (Males)						
Less than weekly	5898	1.02 (0.98-1.06)	0.68	3044	2.98 (2.95-3.02)	1.51 x 10 ⁻¹⁶
Weekly	2565	1.04 (0.98-1.09)		1382	2.80 (2.76-2.84)	
Smoking (Males)						
Never smokers	1100	0.94 (0.87-1.01)	8.77 x 10 ⁻⁰³	560	2.97 (2.92-3.03)	5.33 x 10 ⁻⁰²
Ever smokers	7363	1.04 (1.00-1.08)		3866	2.92 (2.89-2.95)	
Diabetes						
No	15238	1.00 (0.98-1.03)	6.12 x 10 ⁻⁴⁶	7947	3.02 (3.00-3.04)	1.26 x 10 ⁻¹²
Yes	1449	1.58 (1.49-1.68)		652	3.21 (3.15-3.26)	
BMI (kg/m²)						
<22.0	5849	0.78 (0.75-0.80)	9.45 x 10 ⁻²¹⁸	2946	3.01 (2.99-3.04)	5.86 x 10 ⁻⁰⁶
22.0 - 24.9	5313	0.97 (0.94-1.00)		2748	3.00 (2.97-3.03)	
>25.0	5525	1.50 (1.45-1.55)		2905	3.07 (3.05-3.10)	
SBP (mmHg)						
<120.0	3125	0.85 (0.81-0.89)	9.34 x 10 ⁻³⁴	1845	2.99 (2.95-3.02)	2.09 x 10 ⁻⁰⁷
120.0 - 139.9	5496	1.01 (0.98-1.05)		2922	3.02 (2.99-3.05)	
>140.0	8066	1.15 (1.12-1.19)		3832	3.06 (3.04-3.09)	
LDL-cholesterol (mg/dL)						
<80.0	5942	0.92 (0.89-0.96)	3.12 x 10 ⁻⁴⁰	3281	2.92 (2.90-2.95)	1.60 x 10 ⁻⁵¹
80 - 99.9	5020	1.01 (0.97-1.05)		2632	3.02 (3.00-3.05)	
>100	5725	1.22 (1.18-1.26)		2686	3.16 (3.13-3.19)	
TG (mg/dL)^f						
<110.0	5640	0.90 (0.87-0.93)	6.05 x 10 ⁻⁴⁷	2882	3.02 (2.99-3.05)	9.02 x 10 ⁻⁰²
110 - 179.9	5360	1.02 (0.98-1.05)		2753	3.05 (3.02-3.08)	
>180.0	5687	1.24 (1.20-1.28)		2964	3.03 (3.00-3.05)	
Physical activity (MET hours/day)						
<10.0	5218	1.22 (1.17-1.26)	1.42 x 10 ⁻²⁷	2363	3.11 (3.08-3.14)	1.45 x 10 ⁻⁰⁷
10.0 - 14.9	3366	1.12 (1.07-1.17)		1632	3.05 (3.01-3.08)	
>15.0	8103	0.95 (0.92-0.98)		4604	3 (2.97-3.02)	
Standing height of males (m)						
<1.60	2255	1.04 (0.98-1.10)	0.93	1026	2.92 (2.87-2.97)	0.85
1.60 - 1.64	2508	1.02 (0.97-1.07)		1260	2.93 (2.89-2.97)	
1.65 - 1.69	2107	1.02 (0.97-1.08)		1137	2.91 (2.87-2.95)	
>1.70	1593	1.04 (0.97-1.11)		1003	2.93 (2.89-2.98)	
Standing height of females (m)						
<1.50	2495	1.06 (1.01-1.12)	4.41 x 10 ⁻⁰²	1154	3.15 (3.10-3.19)	2.13 x 10 ⁻⁰²
1.50 - 1.54	2656	1.08 (1.03-1.14)		1321	3.17 (3.13-3.21)	
1.55 - 1.59	2044	1.09 (1.03-1.15)		1098	3.13 (3.09-3.18)	
>1.60	1029	0.96 (0.89-1.03)		600	3.09 (3.03-3.14)	

BMI: body-mass index; LDL: Low-density lipoprotein; TG: Triglycerides.

^aGeometric means (95% CI) standardised for age, sex, study area and case status unless otherwise stated

^bStandardised for sex, study area and case status only. Age groups inclusive of the high end of each age group

^cStandardised for age, study area and case status only

^dStandardised for age, sex and case status only

*derived from a multivariable regression model and tests the null hypothesis of no association between CRP/fibrinogen and categorical trait.

3.3.3 Associations of CRP and fibrinogen with CVD

A key objective of my thesis is to understand whether observational associations of CRP and fibrinogen with CVD outcomes observed in European populations are also evident in East Asians. A number of observational studies in East Asians has already suggested that CRP is indeed associated with CHD and IS but not with ICH^{230,231,311}. However, the number of ICH cases in previous studies were limited compared to the number of ICH cases in the present study (383 to 738 in previous studies vs 4,476 in the present study). Furthermore, associations of fibrinogen (which, along with CRP, is also regulated by IL-6 signalling) with CVD outcomes in East Asians are currently lacking. In particular, there are no studies in East Asians or Europeans that compared the shape and strength of associations between CRP and fibrinogen - two key proteins regulated by IL-6 signalling - with MCE, IS and ICH in the same population.

This section investigated the associations of CRP and fibrinogen with CVD outcomes - MCE, IS and ICH in a large cohort of East Asians. To accomplish this objective, I fitted two separate logistic regression models. The first model adjusted for age, age² and sex (**base model**); a quadratic term for age is used to improve model fit. The second model adjusted for the following covariates: income, occupation, education, systolic blood pressure, body-mass index, diabetes, physical activity, standing height, smoking, alcohol, LDL-C and triglycerides (**full model**). I identified these covariates as potential confounders based on my reading of prior observational studies of inflammation and CVD or their associations with CRP and fibrinogen documented in the previous sections of the present Chapter (Chapter 3.3.1 and 3.3.2). Hence, some covariates that were only associated with fibrinogen and not CRP (e.g. gender) or not associated with either CRP or fibrinogen (e.g. education and income) were nevertheless retained in the model, as they have been associated with systemic inflammation or CVD in previous studies (see Chapter 1.5.2, studies listed in Table 1.1). Retaining the same covariates for analysis also enabled comparisons between CRP and fibrinogen.

However, it is possible that the impact of potential confounders on CRP/fibrinogen-CVD associations observed in previous studies may not necessarily be the case for CRP/fibrinogen-CVD associations in the present study. To evaluate the impact of potential confounders, I illustrated CRP/fibrinogen-CVD associations in three different ways in the present section.

The first figure (Figure 3.4) shows the dose-response associations of quintiles of CRP and fibrinogen with the log-odds of MCE, IS and ICH. This is typically done prior to obtaining regression coefficients^{92,93}, to establish the linearity of the relationship between the log-odds of outcome with unit change in independent variables - a key assumption of logistic regression models³²⁵. Both CRP and fibrinogen were log-transformed to normal distributions for ease of interpretation and to be comparable to previous studies which had also presented estimates of CVD association using log-transformed CRP/fibrinogen^{92,93}. The quintiles were derived from the distribution of CRP/fibrinogen in the combined case-control cohort. To derive the usual levels of log CRP/fibrinogen in each quintile, I regressed log CRP/fibrinogen measured at resurvey on log CRP/fibrinogen measured at baseline; this linear regression model was used to predict usual levels of CRP/fibrinogen for participants in each baseline-defined quintiles (see Chapter 3.2.5). Logistic regression models were then fitted for each of the 10 regions separately for each quintile of CRP and fibrinogen and estimates were meta-analysed using a fixed-effect modelling with inverse variance weighting; for each region in each quintile, models were excluded when they failed to converge (9 models out of 600 models in total for CRP and fibrinogen analyses).

As illustrated in Figure 3.4, I observed log-linear positive associations of log CRP with risks of MCE, IS and ICH, in both base and full models. The associations of CRP were stronger with MCE than with IS or ICH. The shape of associations of CRP illustrated in Figure 3.4 were similar between IS and ICH. In contrast to CRP, although fibrinogen showed weaker log-linear associations with MCE and IS,

these associations showed no significant trend in the full model. Furthermore, no associations of fibrinogen were observed with ICH in either base or full model.

Two things are of note in these observations. First, the contrasting shape of associations between CRP and fibrinogen in the present analysis suggested that although both CRP and fibrinogen are under the common control of IL-6 signalling, CRP, but not fibrinogen, is associated with CVD risk in the present study. Furthermore, although CRP remained associated with MCE, IS and ICH in the full models, the slope of the association became less steep after adjustment for potential confounders, suggesting an important role of confounders to potentially distort associations of CRP and fibrinogen with CVD risk.

Given the important role of confounders, the next analyses in this chapter assessed the impact of confounders on the magnitude of the associations of CRP and fibrinogen with MCE, IS and ICH. Confounders were first assessed in groups (grouped based on type of confounder) and then individually to understand if there are specific confounders or groups of similar confounders that disproportionately affected the relationship of CRP and fibrinogen with CVD risk. This is followed by a subgroup analysis to assess whether associations of CRP and fibrinogen vary by pre-specified subgroups.

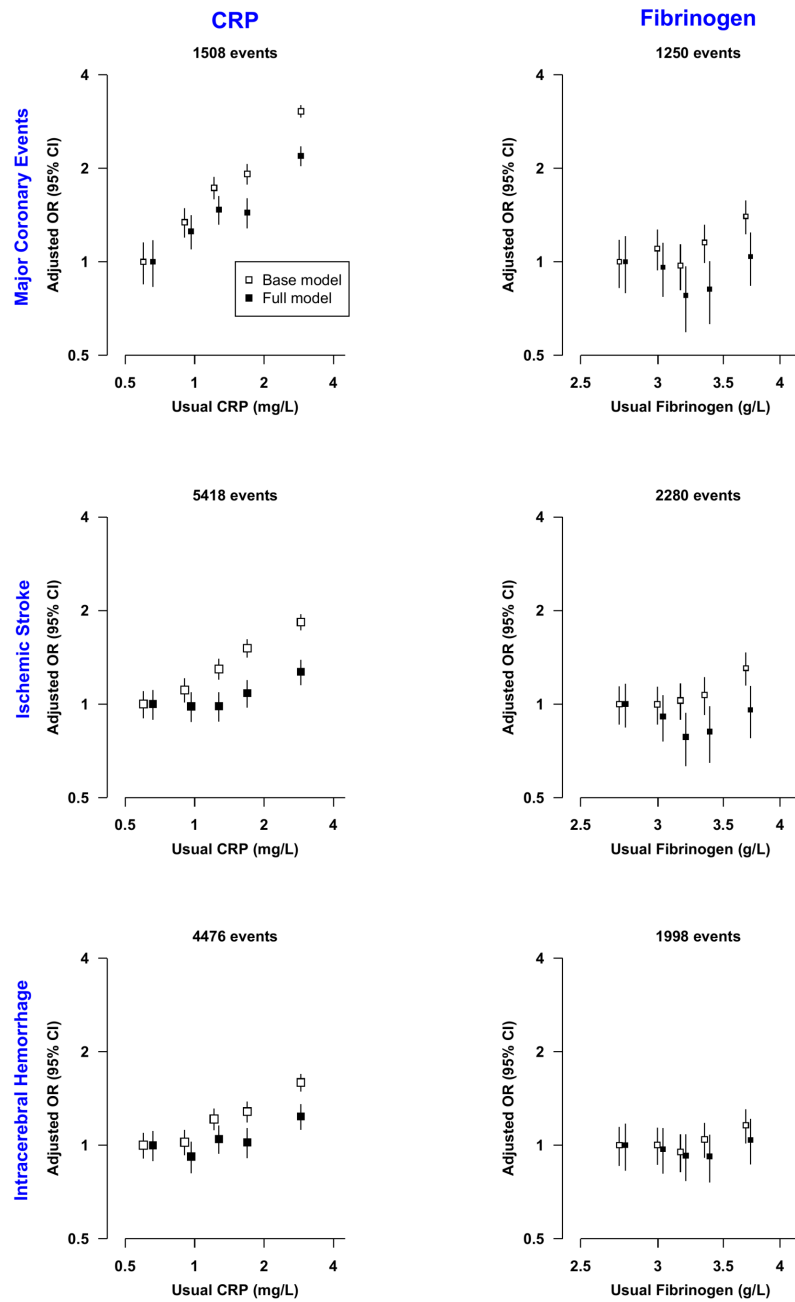


Figure 3.4: Risk of Major Coronary Events, Ischaemic Stroke, and Intracerebral haemorrhage by quintiles of C-reactive protein (CRP) and fibrinogen. The figure illustrates associations of CRP and fibrinogen with MCE, IS and ICH. Base model included age, age² and sex; full model included base model + income, occupation, education, SBP, BMI, diabetes, physical activity, standing height, smoking, alcohol, LDL-C and triglycerides. Models were fitted for each of the 10 regions separately for each quintile of CRP and fibrinogen and estimates were meta-analysed using a fixed-effect with inverse variance weighting. Error bars represent 95% floated confidence intervals (CI). The areas of the boxes are proportional to the inverse of the variance of the log ORs. OR, CRP, and fibrinogen are plotted on log scale

To assess the impact of these confounder groups, I constructed five models similar to what was done in previous studies of Europeans⁹²:

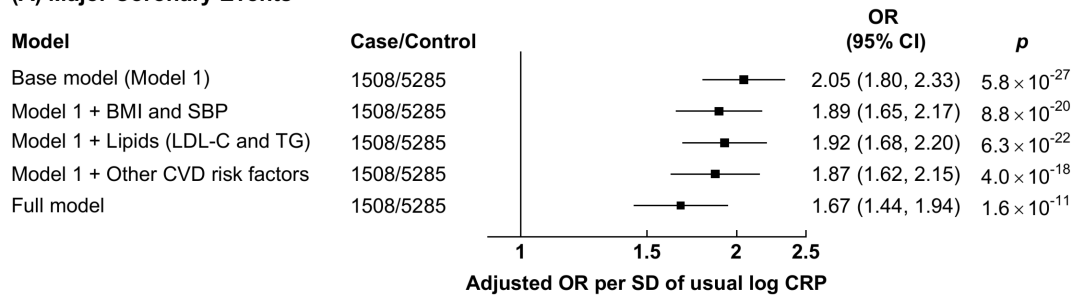
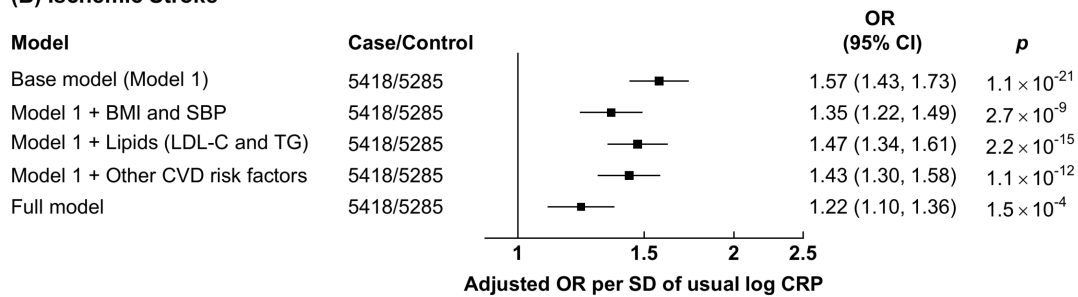
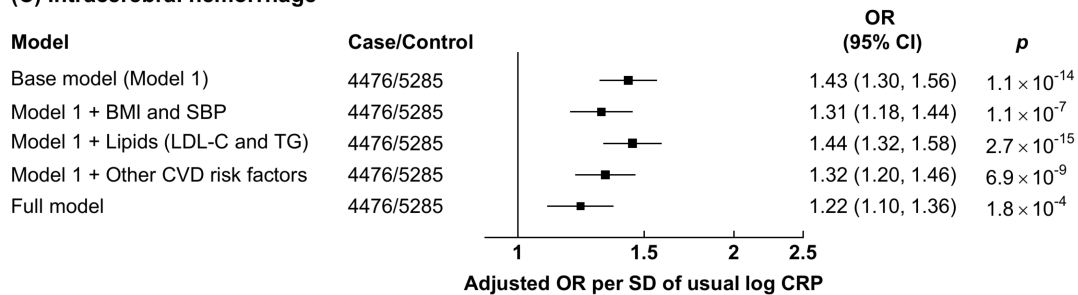
1. base model (adjusted for age, age² and sex),
2. base model adjusted for BMI and SBP,
3. base model adjusted for lipids (LDL-C and TG),
4. base model adjusted for other CVD risk factors (income, occupation, education, physical activity, standing height, smoking and alcohol),
5. full model (base model + all covariates)

This approach is useful because it will help me understand whether there are groups of confounders that have a disproportionately large effect on CRP/fibrinogen - CVD association. As before, the models were stratified by region and meta-analysed using a fixed effect modelling with inverse variance weighting.

In the base model, the ORs [95% CI] per-SD higher usual log CRP was 2.05 [1.80, 2.33] for MCE, 1.57 [1.43, 1.73] for IS, and 1.43 [1.30, 1.56]) for ICH. However, on adjustment for BMI and SBP, there was substantial attenuation of effects for all CVD outcomes: adjusted ORs [95% CI] per-SD higher usual log CRP was 1.89 [1.65, 2.17] for MCE, 1.35 [1.22, 1.49] for IS, and 1.31 [1.18, 1.44]) for ICH. Between the base model and base model adjusted for SBP and BMI, the largest attenuations of effect were observed in the following order: IS (22%), MCE (16%), and ICH (12%). This suggests that SBP and BMI accounts differentially for the associations of CRP exhibited with MCE and IS; despite MCE and IS sharing a similar pathophysiological process (i.e. atherosclerosis), accounting for SBP and BMI attenuate associations of CRP with IS greater than MCE. Compared to adjustment of base model with SBP and BMI, the attenuation of association of CRP with MCE/IS/ICH was less when base model was adjusted for LDL-C and TG or ‘other’ CVD risk factors. However, in each case, for MCE, IS and ICH, there was some attenuation when adjustments were made to their base models; the only exception was ICH. In ICH, there was no observed attenuation of effect size when the base model of ICH (OR [95% CI] per-SD higher usual log CRP: 1.43 [1.30, 1.56]) was

adjusted for lipids [1.44 [1.32, 1.58]]. This suggests that the association of CRP with ICH is largely independent of contributions from LDL-C and TG, unlike the atherosclerotic CVD phenotypes - MCE and IS. In all CVD outcomes, despite substantive attenuation of effects post-adjustment, CRP remained associated in the full model; strong associations were observed with MCE (OR [95% CI]: 1.67 [1.44, 1.94]), and similar associations observed with IS (OR [95% CI]: 1.22 [1.10, 1.36]) and ICH (OR [95% CI]: 1.22 [1.10, 1.36]) in the full model. The full model for stroke types was notable, as identical effect sizes are observed for IS and ICH - two pathophysiologically distinct stroke types (Figure 3.5).

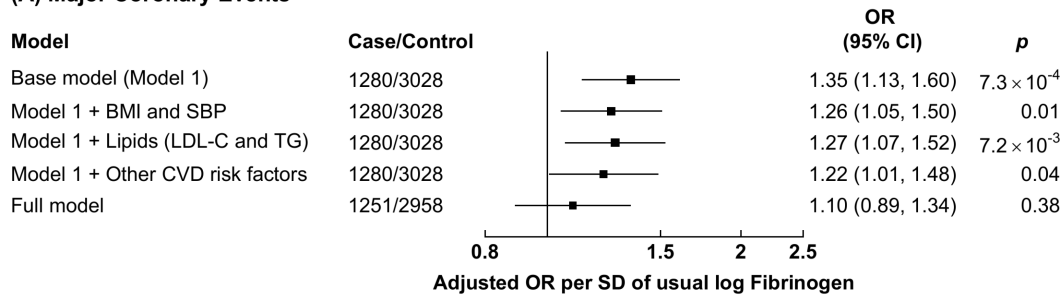
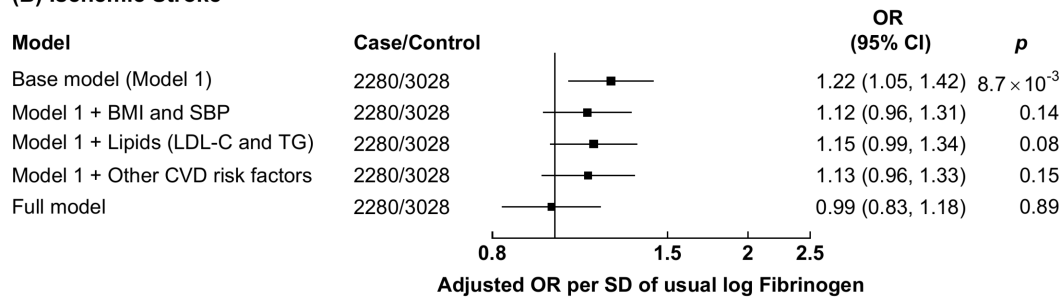
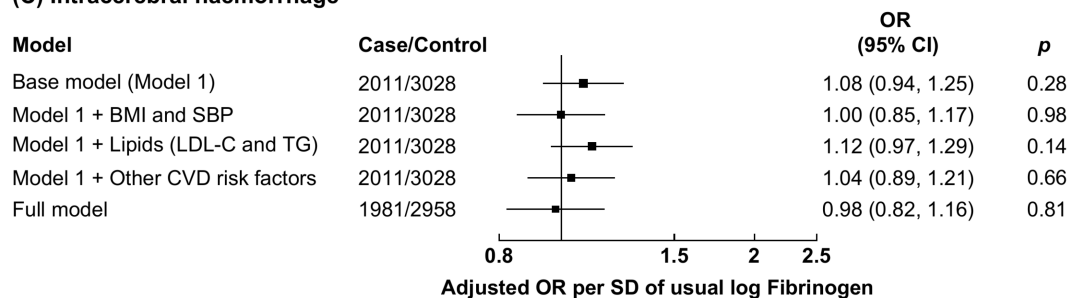
Unlike associations of MCE, IS, and ICH with CRP, associations of fibrinogen with MCE and IS were weaker in the base model compared to associations of these outcomes with CRP, with largely neutral associations observed with ICH. In the base model, the ORs [95% CI] per-SD higher usual log fibrinogen was 1.35 [1.13, 1.60] for MCE, 1.22 [1.05, 1.42] for IS, and 1.08 [0.94, 1.25] for ICH. However, there is no evidence of association of fibrinogen with MCE in the full model (OR [95% CI] per-SD higher usual log fibrinogen: 1.10 [0.89, 1.34]). The respective associations of fibrinogen concentrations with IS and ICH attenuate earlier than MCE. With IS, the association of fibrinogen becomes neutral on adjustment for BMI and SBP (OR [95% CI] per-SD higher usual log fibrinogen: 1.12 [0.96, 1.31]). With ICH, there was no evidence of association of fibrinogen in the base model (OR [95% CI] per-SD higher usual log fibrinogen: 1.08 [0.94, 1.25]) (Figure 3.6).

(A) Major Coronary Events**(B) Ischemic Stroke****(C) Intracerebral hemorrhage**

CVD: Cardiovascular disease; BMI: body-mass index; LDL-C: Low-density lipoprotein cholesterol; TG: triglycerides

Figure 3.5: Risk of (A) Major Coronary Events, (B) Ischaemic Stroke, and (C) Intracerebral haemorrhage per standard deviation (SD) of usual log-transformed high-sensitivity C-reactive protein (CRP).

The figure illustrates associations of CRP with MCE, IS and ICH. Base model included age, age² and sex; full model included base model + income, occupation, education, SBP, BMI, diabetes, physical activity, standing height, smoking, alcohol, LDL-C and triglycerides. Models were fitted for each of the 10 regions separately and estimates were meta-analysed using a fixed-effect with inverse variance weighting. Error bars represent 95% floated confidence intervals (CI). The areas of the boxes are proportional to the inverse of the variance of the log ORs. OR, CRP, and fibrinogen are plotted on log scales.

(A) Major Coronary Events**(B) Ischemic Stroke****(C) Intracerebral haemorrhage**

CVD: Cardiovascular disease; BMI: body-mass index; LDL-C: Low-density lipoprotein cholesterol; TG: triglycerides

Figure 3.6: Risk of (A) Major Coronary Events, (B) Ischaemic Stroke, and (C) Intracerebral haemorrhage per standard deviation (SD) of usual log-transformed Fibrinogen.

The figure illustrates associations of fibrinogen with MCE, IS and ICH. Base model included age, age² and sex; full model included base model + income, occupation, education, SBP, BMI, diabetes, physical activity, standing height, smoking, alcohol, LDL-C and triglycerides. Models were fitted for each of the 10 regions separately and estimates were meta-analysed using a fixed-effect with inverse variance weighting. Error bars represent 95% floated confidence intervals (CI). The areas of the boxes are proportional to the inverse of the variance of the log ORs. OR, CRP, and fibrinogen are plotted on log scales.

To identify specific confounders that had maximal impact on the association of CRP and fibrinogen with CVD, I additionally fitted logistic regression models to assess the progressive impact of individual risk factors on the associations of CRP and fibrinogen with MCE, IS and ICH (Figure 3.7, arranged in descending order of Wald χ^2). As before, models were fitted separately for each of the 10 study areas for each covariate and estimates were meta-analysed; for each analysis, study areas were excluded when models failed to converge (2 models out of a total of 780 models). The equivalent figure that illustrates the effect of adjustment of individual risk factors is provided in the Appendix Figure A.2.

As shown in Figure 3.7, addition of SBP to the base model had the largest impact on the associations of CRP with MCE (OR [95% CI] changed from 2.05 [1.80, 2.33] in the base model to 1.88 [1.64, 2.15] on adjustment for SBP, with corresponding change of χ^2 from 115.6 to 85.3), and with ICH (OR [95% CI] in base model to SBP adjustment changed from 1.43 [1.30, 1.56] to 1.26 [1.14, 1.38], with corresponding change of χ^2 from 59.7 to 21.5). Similar to CRP, associations of fibrinogen with MCE and IS were similarly affected on adjustment for SBP. From base model to adjustment for SBP, associations of fibrinogen with MCE changed from 1.35 [1.13, 1.60] to 1.27 [1.06, 1.352] (with corresponding change of χ^2 from 11.4 to 6.9) and association of fibrinogen with IS changed from 1.22 [1.05, 1.42] to 1.14 [0.97, 1.33] (with corresponding change of χ^2 from 6.9 to 2.6). This highlights the important role of SBP that acts as the strongest confounder to the association of CRP with MCE and ICH, and association of fibrinogen with MCE and IS.

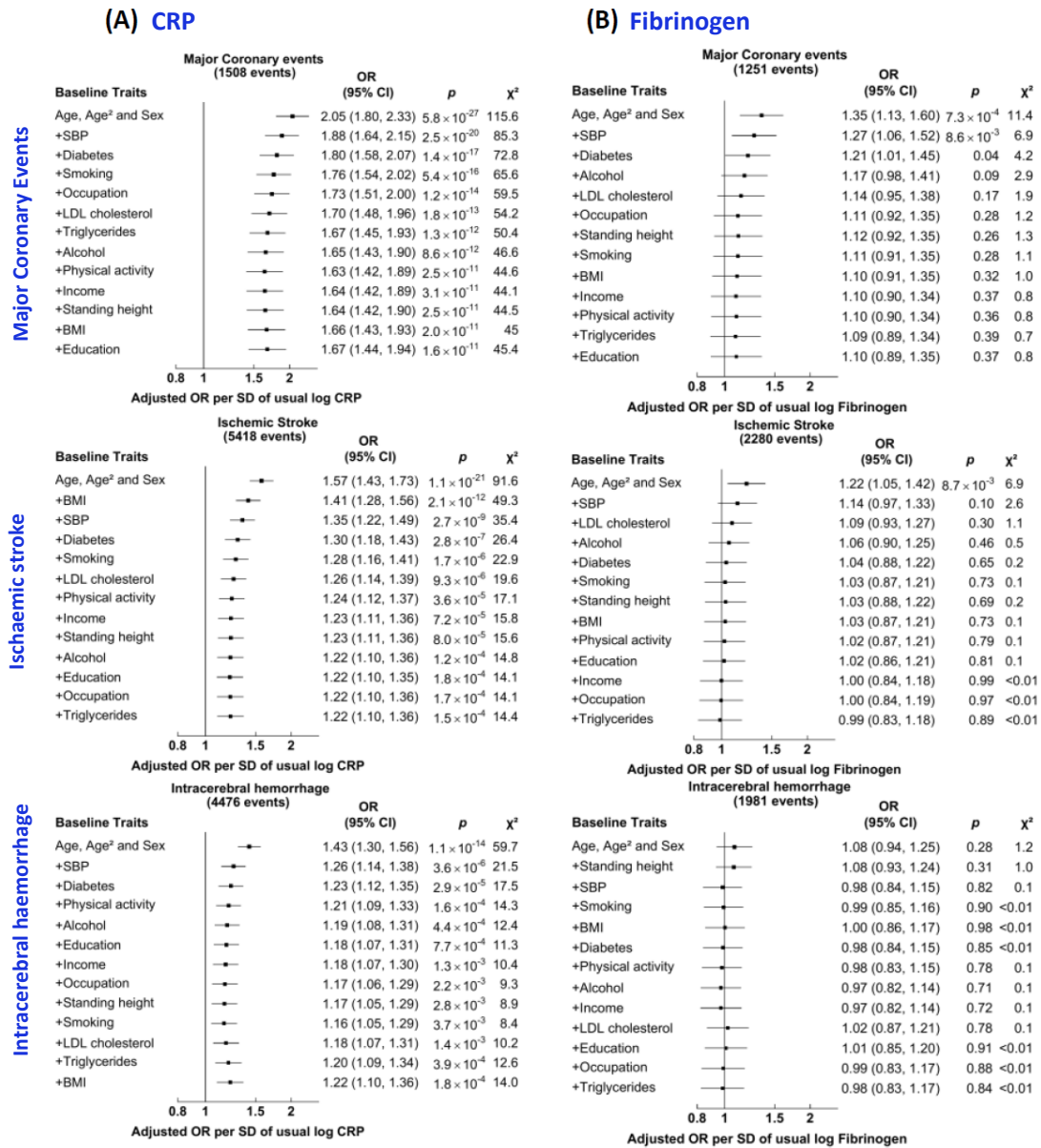
For the association of CRP with IS, BMI had a larger impact than SBP (OR [95% CI] in base model to BMI adjustment changed from 1.57 [1.43, 1.73] to 1.41 [1.28, 1.56], with corresponding change of χ^2 from 91.6 to 49.3), suggesting an important role of adiposity to confound associations between CRP and IS. As in Figure 3.5, Figure 3.7 showed that CRP remained strongly associated with MCE, and associated in similar magnitudes with IS and ICH in the fully adjusted models; this further supports the relevance of inflammation for MCE and highlights the

similar association of CRP with pathophysiological distinct stroke types - IS and ICH.

Although fibrinogen was associated with MCE and IS in the base model, there was no significant association after adjustments. For the association of fibrinogen with MCE, associations attenuated to include the null (OR [95% CI]: 1.17 [0.98, 1.41]) after adjustment for the first three covariates - SBP, diabetes and alcohol - added on the basis of decreasing Wald χ^2 change. Similarly, for the association of fibrinogen with IS, associations attenuated to include the null after adjustment for SBP. For the association of fibrinogen with ICH, no associations were observed even in the base model.

The present analyses of CRP and fibrinogen that investigated progressive associations with MCE, IS and ICH generated two important insights. The first insight is related to the importance of SBP for CVD associations of both CRP and fibrinogen (SBP was consistently one of the strongest confounders of the associations of CRP and fibrinogen with MCE, IS or ICH). The second insight is related to the differential impact of specific confounders for associations of CRP and fibrinogen with the same outcomes. For example, both SBP and diabetes had maximal impact on the associations of CRP and fibrinogen with MCE; but while SBP and BMI are most important for explaining associations of CRP with IS, SBP and LDL-C appear to be the strongest confounders for associations of fibrinogen with IS.

The next analyses examined whether associations of CRP and fibrinogen with MCE, IS and ICH vary by pre-specified potential effect modifier subgroups.



CVD: Cardiovascular disease; BMI: body-mass index; LDL-C: Low-density lipoprotein cholesterol; TG: triglycerides

Figure 3.7: Progressive adjustment of (A) CRP and (B) Fibrinogen for Major Coronary Events, Ischaemic Stroke and Intracerebral haemorrhage.

The figure illustrates associations of CRP and fibrinogen with MCE, IS and ICH in progressively adjusted models. Base model included age, age² and sex; for each adjustment, models were fitted for each of the 10 regions separately and estimates were meta-analysed using a fixed-effect with inverse variance weighting. Progressive adjustment used at each step the covariate giving the largest change in χ^2 . Error bars represent 95% floated confidence intervals (CI). The areas of the boxes are proportional to the inverse of the variance of the log ORs. OR, CRP, and fibrinogen are plotted on log scales.

3.3.4 Prespecified subgroup analyses of C-reactive protein and fibrinogen

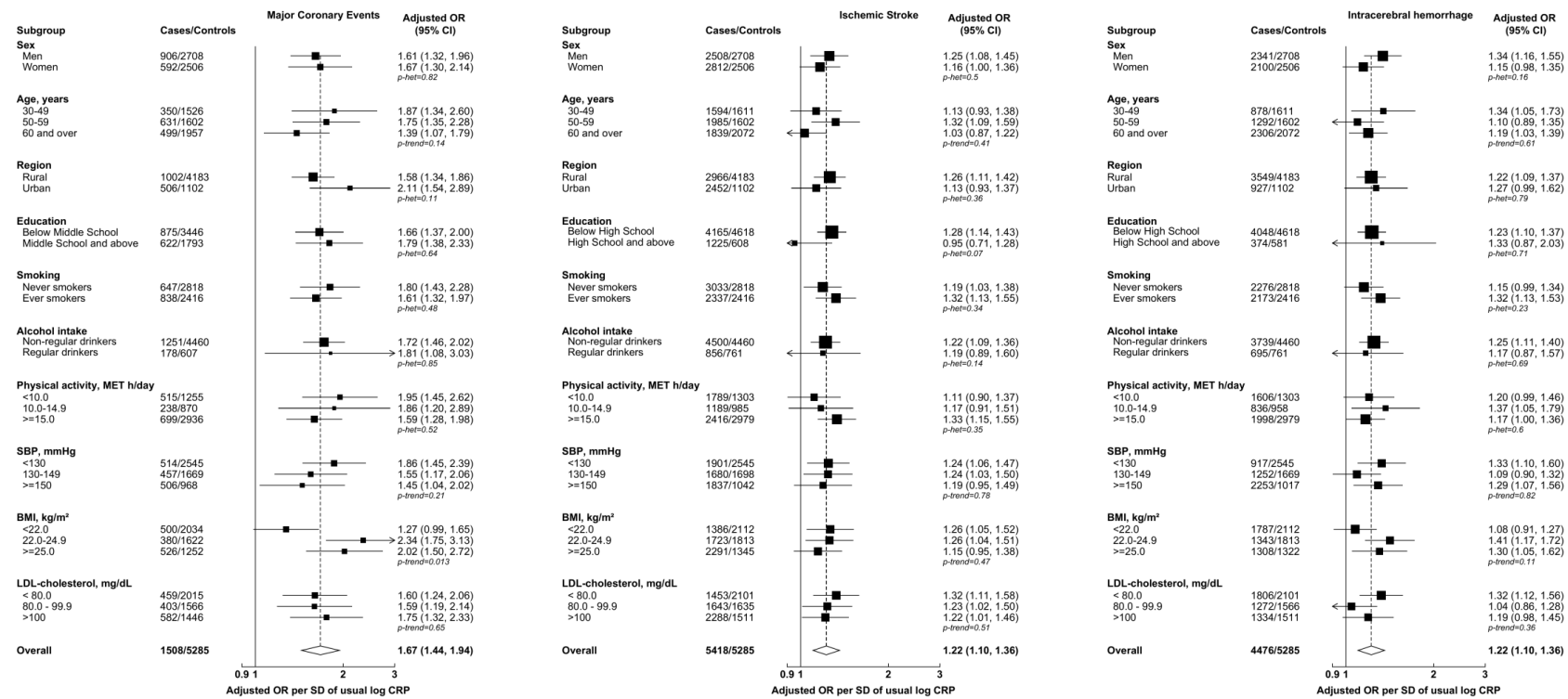
There are suggestions from previous studies that associations of inflammatory markers with CVD may vary by specific subgroups, like sex^{22,326} and levels of LDL-C^{247,327}. The specific mechanisms by which these variations occur are unknown. I sought to understand whether associations of CRP and fibrinogen with CVD outcomes in the present study would vary by specific subgroups. These potential effect modifier traits were pre-specified and measured or ascertained at baseline and included age, gender, region, education, smoking, alcohol consumption, physical activity, SBP, BMI, and LDL-C. The potential effect modifier traits selected for subgroup analysis were based on evidence of these traits being risk factors for CVD from previous studies^{92,93}.

For each subgroup, analyses used full model omitting the relevant covariates under investigation (e.g. the subgroup analysis for sex included all covariates used in the full model except sex). The full model was fitted for each of 10 regions and estimates were meta-analysed by fixed-effect inverse-variance weighted method. For each subgroup in each region, models were excluded if they failed to converge (62 models out of 750 models for CRP and 141 models out of 750 models for fibrinogen failed to converge and were excluded).

For associations of CRP, there was no evidence (after Bonferroni adjustment, $P < 0.05/20$) for effect modification by any of these traits (Figure 3.8). This suggests that associations of CRP with CVD were likely robust to effect modification by specific subgroups. However, the subgroup analyses (that is, again, stratified by region in the present analysis) are also likely to be under-powered to detect an effect.

By contrast, although associations of fibrinogen with MCE, IS, and ICH risk did not exhibit differences between most of the pre-specified subgroups, there was evidence of heterogeneity (heterogeneity and trend tests were conducted by χ^2 tests applied to the log ORs and their SEs, Chapter 3.2.5) by levels of physical activity

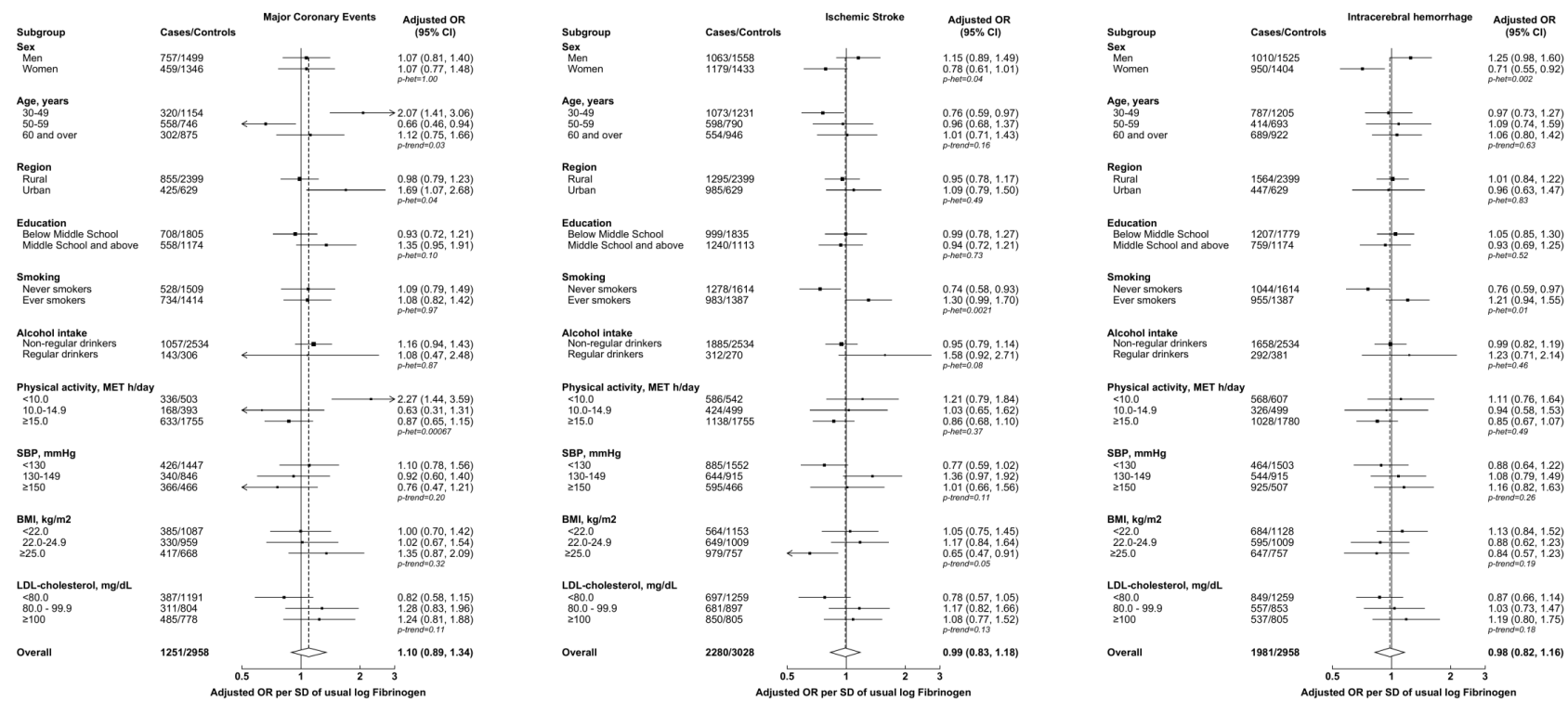
($P=6.7 \times 10^{-4}$) for MCE, and suggestive evidence of an interaction with smoking ($P=0.0021$) for IS (Figure 3.9). Although the subgroup analysis for fibrinogen-MCE associations, like CRP, were corrected for multiple testing by the Bonferroni method, it is still possible that the interactions observed with physical activity and smoking may be false positives and require replication in other independent East Asian datasets.



BMI: Body-mass index; LDL: low-density lipoprotein cholesterol; MET: metabolic equivalent of task; SBP: systolic blood pressure; *p-het*: p-value for heterogeneity; *p-trend*: p-value for trend

Figure 3.8: Risk of Major Coronary Events, Ischaemic Stroke, and Intracerebral haemorrhage per standard deviation (SD) of usual log-transformed high-sensitivity C-reactive protein (CRP) by pre-specified subgroups.

The figure illustrates associations of CRP with MCE, IS and ICH in pre-specified subgroups. Analyses for each subgroup used the full model (age, age², sex, income, occupation, education, SBP, BMI, diabetes, physical activity, standing height, smoking, alcohol, LDL-C and triglycerides), omitting covariates under investigation. Models were fitted for each of the 10 regions separately and estimates were meta-analysed using a fixed-effect method with inverse variance weighting. The sizes of the boxes are proportional to the inverse of the variance of the log ORs. The dotted line represents the overall adjusted OR. The open diamond represents the overall adjusted OR and its 95% CI



BMI: Body-mass index; LDL: low-density lipoprotein cholesterol; MET: metabolic equivalent of task; SBP: systolic blood pressure; *p-het*: p-value for heterogeneity; *p-trend*: p-value for trend

Figure 3.9: Risk of Major Coronary Events, Ischaemic Stroke, and Intracerebral haemorrhage per standard deviation (SD) of usual log-transformed fibrinogen by pre-specified subgroups.

The figure illustrates associations of fibrinogen with MCE, IS and ICH in pre-specified subgroups. Analyses for each subgroup used the full model (age, age², sex, income, occupation, education, SBP, BMI, diabetes, physical activity, standing height, smoking, alcohol, LDL-C and triglycerides), omitting covariates under investigation. Models were fitted for each of the 10 regions separately and estimates were meta-analysed using a fixed-effect method with inverse variance weighting. The sizes of the boxes are proportional to the inverse of the variance of the log ORs. The dotted line represents the overall adjusted OR. The open diamond represents the overall adjusted OR and its 95% CI

3.4 Discussion

In this Chapter, I examined patterns of associations of baseline CRP and fibrinogen concentrations, two well-established markers of systemic inflammation, with CVD risk factors and incident CVD outcomes in a large cohort of East Asians. The CVD outcomes assessed were incident MCE, IS, and ICH, three CVD outcomes that are of particular relevance to the East Asian population given their rising incidence and prevalence in China³²⁸. Although observational associations of CRP and fibrinogen with CVD outcomes are well-established in Europeans, studies in larger East Asian populations are scarce. Furthermore, there is a dearth of studies comparing the *pattern* of association between CRP and fibrinogen with CVD outcomes in an East Asian cohort. Given that both CRP and fibrinogen synthesis are regulated by IL-6 signalling, comparing their patterns of association with CVD would provide useful insights with respect to their distinct biological roles (see Chapter 1.4.3). Additionally, studies that evaluated association of CRP with ICH had conflicting reports; most of these studies had lower number of ICH cases (n=383 to 738), and the present study sought to resolve the conflicting reports of association of CRP with ICH by examining the association in the present cohort that had a higher number of ICH cases (n=4,476).

3.4.1 Associations of CRP and fibrinogen with baseline traits

The aim of these correlative analyses was to assess whether baseline traits, that may or may not be established CVD risk factors, were associated with markers of systemic inflammation. In this regard, I have used CRP and fibrinogen as markers of systemic inflammation and assessed effect on baseline traits for every 1 SD increase in CRP or fibrinogen corresponding to a 3.3 fold and 1.3 fold differences in CRP and fibrinogen respectively (see Chapter 3.2.5). This has, therefore, also provided an opportunity to examine patterns of association between CRP and fibrinogen with these traits.

To start with, I examined the variation of CRP and fibrinogen with higher age and between sex. When examining variation of CRP and fibrinogen from youngest to oldest, I observed evidence of a strong trend with age, with differences between the youngest (30 - 49 years) and oldest (60 and over) age categories of 0.50 mg/L for CRP and 0.61 g/L for fibrinogen. This association with age is consistent with the observed associations of CRP and fibrinogen with age in Europeans^{92,329}. Furthermore, it is consistent with the notion that ageing and age-related diseases converge on inflammation as a mechanism of age-related adverse events³³⁰. This phenomenon has been termed as ‘inflammageing’ and both CRP and fibrinogen are well-established markers of inflammageing³³¹. By contrast, I observed a discordant pattern of association with sex between CRP and fibrinogen; while fibrinogen was observed to be higher in females than males in the present study, there was no evidence of a difference in CRP levels between the sexes. The observation of sex-discrepancy with fibrinogen is consistent with a number of previous studies^{302,329,332,333}; however, the precise biological reasons why levels of fibrinogen are higher in women than men are unclear. Similar to fibrinogen, prior studies have demonstrated higher CRP concentrations in women⁹² that is not observed in the present study. The discrepancy may be due to the larger number of men and women in the European study⁹² (~ 10 times more than the present study) offering greater power to detect a difference or simply reflect an intrinsic biological difference between Europeans and East Asians. East Asians also have substantially lower CRP levels than Europeans (median CRP in Europeans is ~ 2.5 times higher than in East Asians^{334,335}), and there was no evidence of substantial differences of CRP being noted between males and females in East Asian studies^{335–337}. Indeed, one study noted large trans-ethnic variations of mean CRP concentrations that cannot be readily explained by study design, CRP assay, age or BMI³³⁸.

Like sex, there were other traits that exhibited a discordant pattern of association between CRP and fibrinogen. For example, the discordant association between CRP and fibrinogen with TG. For context, I observed a positive association of higher CRP with atherogenic lipids (LDL-C and TG) and inverse association with

HDL-C. The largest study of CRP that meta-analysed 54 prospective Europeans studies⁹² also observed a similar trend of associations with HDL-C (i.e. inverse association with CRP) and non-HDL-C lipids (i.e. a positive association with CRP); non-HDL-C lipids comprise of LDL-C and TG³³⁹. The observed associations of CRP with lipids were directionally consistent with the associations of fibrinogen with lipids, except for TG; whereas CRP showed a positive association with TG, associations with fibrinogen were neutral. The neutral association of fibrinogen with TG was consistent with observations in the European population where association of fibrinogen with TG were either ‘weak’³⁴⁰, ‘curvilinear’³²⁹, or null³⁴¹. By contrast, consistent patterns of association of CRP and fibrinogen were observed with LDL-C and HDL-C; with LDL-C, higher CRP and fibrinogen were associated with higher LDL-C and lower HDL-C and this pattern is consistent with observations in Europeans³²⁹. The precise biological mechanisms underlying these associations remain elusive, although one study has suggested a central role of IL-6, the cytokine that regulates hepatic synthesis of CRP and fibrinogen to also regulate HDL-C concentrations³⁴² via inhibition of lipoprotein lipase activity, and stimulation of lipolytic activity of endothelial lipase³⁴³; lipoprotein lipase, although it metabolises triglycerides, is correlated with HDL-C levels and endothelial lipase is associated with metabolism of HDL particles in animal models³⁴⁴. C-reactive protein and fibrinogen also exhibited a similar inverse relationship with Apo-A1 with similar magnitudes of associations. This association with Apo-A1 was expected given the well-established biological relationship of Apo-A1 with HDL-C³⁴⁵; Apo-A1 is the major protein component of HDL particles and enables the transport of HDL-C in plasma. Similar to HDL-C and Apo-A1, I observed an inverse association of CRP and fibrinogen with albumin concentrations; this association was consistent with prior studies in Europeans^{92,329} and aligns with the suggestion that lower albumin levels is one of the markers of systemic inflammation³⁴⁶.

Beyond lipids, lipoproteins, and albumin, associations of CRP and fibrinogen were directionally consistent to each other with respect to their associations with markers of liver function (AST, ALT, and GGT) and renal function (creatinine, uric acid,

and cystatin-C). Although, studies examining epidemiological associations of CRP and fibrinogen with liver and renal function traits are limited, the few studies that exist do support positive associations of markers of liver and renal function with markers of systemic inflammation³⁴⁷⁻³⁴⁹.

In addition to biological markers, I also assessed associations of CRP and fibrinogen with CVD risk factors like SBP and BMI. Consistent with previous studies^{92,329}, there was some evidence of increasing geometric mean CRP and fibrinogen concentrations with progressively higher SBP tertiles. However, when this analysis was limited to controls, and a dose-response curve was constructed to assess the shape of the association, there was no evidence of a trend for CRP across higher SBP quintiles. Indeed, the shape of CRP-SBP association curve is similar to the shape observed in Europeans⁹² and also showed weak correlations in Europeans between CRP and SBP ($r = 0.13$). With BMI, which is a well-established measure of adiposity, as observed in Europeans^{92,120,329} - higher CRP and fibrinogen was associated with higher BMI. Physical activity, which is correlated with measures of adiposity, showed inverse associations with higher CRP and fibrinogen in the present study, consistent with observations in Europeans^{92,329} and consistent with the suggestion that regular exercise may have ‘anti-inflammatory’ roles^{149,321,350}.

With regards to social habits such as smoking and alcohol consumptions, the evidence of their associations with CRP and fibrinogen were not as strong as evidence for other risk factors. For example, although I observed higher CRP levels in smokers than non-smokers, this was not statistically significant after correction for multiple testing, and associations with fibrinogen were neutral. This was inconsistent with previous studies that showed higher CRP and fibrinogen levels in smokers than non-smokers with statistical significance^{92,329,332,351}. Reasons for this inconsistency may be inadequate power to examine differences in CRP between smokers and non-smokers in the present study due to lower sample size or reflect valid trans-ethnic differences. I had a lower sample size because I included only men for the analysis of smoking and CRP given the fact that a very small

number of women (3%) reported smoking in CKB³²³. Similar to smoking, there was no evidence of differences in CRP between weekly drinkers and non-drinkers in the present study, in contrast to previous reports in Europeans which observed lower CRP in drinkers than non-drinkers⁹². Like smoking, this could reflect limited statistical power in my analysis (analysis of alcohol, like smoking, was limited to men only given that only 2% women reported drinking in CKB³²⁴) or real trans-ethnic differences or have occurred due to trans-ethnic differences in distributions of effect measure modifiers. By contrast, despite limiting analysis to men, I observed lower fibrinogen concentrations in drinkers than non-drinkers, and this was consistent with observations in Europeans³²⁹. The observed effect of alcohol on levels of CRP and fibrinogen was previously used to explain the apparent cardio-protective benefits of moderate alcohol consumption. But as the most recent Mendelian randomisation studies (see Chapter 1.5.4 for a primer on Mendelian randomisation) have shown^{75,352}, even light consumption of alcohol is deleterious and the observed associations are likely the result of residual confounding and/or reverse causation.

3.4.2 Associations of CRP and fibrinogen with CVD outcomes

In Chapter 1.5.2, I tabulated different studies that have examined associations of CRP and fibrinogen with CVD markers (e.g. atherosclerosis) and outcomes (i.e. CVD events) using different study designs. In this section, I tabulated the effect estimates (with confidence intervals) reported by the different studies that investigated associations of CRP/fibrinogen on CVD risk in different populations for comparison of effect estimates in the present study (Table 3.5). For comparability of effect estimates with previous studies and ease of interpretation, I examined association with CVD outcomes per 1 SD increase of log CRP or fibrinogen that corresponded to a 3.3 fold and 1.3 fold difference of CRP and fibrinogen, respectively (the fold change is similar to the largest European studies of CRP and fibrinogen^{92,93}). The CVD outcomes assessed are MCE and stroke; the types of stroke examined were IS and ICH. Below I will discuss the patterns of associations of CRP and

Table 3.5: Observational Evidence

Tabulating observational studies of markers of inflammation showing the association CRP and fibrinogen with ischaemic stroke, intracerebral haemorrhage and coronary heart disease

Protein	Study reference	Study design	Population	N	IS	N	ICH	N	CHD
					RR (95% CI) ^a		RR (95% CI) ^a		RR (95% CI) ^a
CRP	[92]	Meta-analysis	Europeans	1931	1.27 (1.15-1.40)	469	1.07 (0.86-1.32)	5373	1.36 (1.29-1.45)
	[230]	Meta-analysis	East Asians	1105	1.40 (1.08 - 1.81)	738	1.04 (0.66 - 1.65)	529	1.75 (0.96 - 3.19)
	[231]	Prospective study	Chinese	1049	1.15 (1.06 - 1.24)	383	0.97 (0.86 - 1.08)	-	-
	[235]	Prospective study	European women	517	1.69 (1.26 - 2.29)	111	0.83 (0.44 - 1.57)	-	-
	Present study	Nested case-control	Chinese	5418	1.22 (1.10 - 1.36)	4476	1.22 (1.10 - 1.36)	1508	1.67 (1.44 - 1.94)
Fibrinogen	[93]	Meta-analysis	Mostly Europeans	596	1.93 (1.47 - 2.52)	215	1.45 (0.91 - 2.29)	2534	1.82 (1.60 - 2.06)
	[235]	Prospective study	European women	517	1.01 (0.77 - 1.34)	111	1.14 (0.64 - 2.00)	-	-
	Present study	Nested case-control	Chinese	2280	0.99 (0.83 - 1.18)	1981	0.98 (0.82 - 1.16)	1251	1.10 (0.89 - 1.34)

N: Number of cases; IS: Ischaemic stroke; ICH: Intracerebral haemorrhage; CHD: Coronary heart disease' CRP: C-reactive protein; RR: Relative risk; CI: Confidence Interval

^aAll studies reported RR per SD increase of natural log-transformed biomarker except [235] - top vs bottom quartile.

fibrinogen observed with each CVD outcome and compare them to observations in the European and other East Asian study cohorts.

Associations of CRP and fibrinogen with MCE

The present study used a composite outcome defined as major coronary events or MCE (composed of fatal IHD, AMI, and patients who have undergone an stent placement procedure or bypass surgery, see Chapter 3.2.3) to represent CHD and this is largely equivalent to CHD definitions used in European studies⁹².

In the present study, I observed log-linear positive associations of log-transformed CRP with risk of MCE that remained after adjustment for potential confounders, with no evidence of effect modifications observed in pre-specified subgroups. The shape of the observed associations were consistent with several studies that examined associations of CRP with CHD, including the largest study of CRP-CHD in Europeans that meta-analysed 54 prospective studies⁹² and is consistent with associations observed in studies of East Asians^{230,231,311}. The study in Europeans had 5,373 CHD cases (first-ever myocardial infarction or fatal CHD) and observed that per SD higher usual log CRP concentration (equivalent to approximately a

3-fold difference of CRP in the original scale, similar to the 3.3 fold difference in the present study) was associated with higher CHD risk. Comparing the estimates observed in the meta-analysis with estimates in the present study, I noted that the MCE estimate in the present study (OR [95% CI] per 1 SD higher usual log CRP: 1.67 [1.44, 1.94]) was appreciably greater than which was observed in the meta-analysis (OR [95% CI] per 1 SD higher usual log CRP: 1.36 [1.29, 1.45]) higher risk of CHD, with very little overlap of confidence intervals (Table 3.5). The difference in the estimates likely reflects differences in the spectrum of disease severity and/or aetiology, since there were a much higher proportion of fatal MCE cases in the present study (65% in CKB vs 29% in the European meta-analysis⁹²). Indeed, when the European study was restricted to fatal events, the risk estimate was closer to the estimate in the present study (OR [95% CI] per 1 SD higher usual log CRP: 1.57 [1.44, 1.71]) (eTable 10 of⁹²). This suggests that concentrations of CRP were a stronger predictor of risk of fatal CHD events than risk of non-fatal CHD events.

By contrast, there was no evidence of association for fibrinogen with MCE in the present East Asian cohort. This was unexpected and inconsistent with previous studies (Table 3.5). In the largest study of fibrinogen-CHD association in Europeans, the authors meta-analysed 31 prospective studies, involving a total of 154,211 participants (including 7% from Japan) and 7118 CHD cases⁹³. This fibrinogen-CHD study in Europeans demonstrated positive log-linear associations of concentration of fibrinogen with CHD; although the association substantially attenuated on adjustment for classical CVD risk factors, smoking, alcohol, BMI, LDL-C, TG, and SBP, the association remained in the full model. Although I observed similar positive associations (e.g. with age, LDL-C, BMI) and inverse associations (e.g. males, alcohol consumption, HDL-C) of fibrinogen with CVD risk factors as the European study, the association of fibrinogen with MCE in the present study was largely neutral on adjustment for age and sex. There are several possible reasons for this discrepancy; first, with approximately 9,000 individuals (mean fibrinogen [g/L] = 3.00) in the present cohort having fibrinogen measurements (compared to ~18,000 individuals

with CRP measurements), the present study had only 1250 MCE events compared to 7118 CHD events in the European study (mean fibrinogen [g/L] = 3.02)³⁵³. Hence, a lack of power in the present study can potentially explain the neutral association observed for fibrinogen with MCE. Second, unlike CRP, almost a third of the variation in plasma fibrinogen in the European cohort (~12% Asian origin) was explained by non-modifiable risk factors that included ethnic group, cohort, age, sex, and season³⁵³. Therefore, differences in the distribution of potential confounders and potential effect measure modifiers between different population groups can potentially account for these trans-ethnic differences in associations^{232,234}. This can happen if background CVD risk is, for example, higher in one ethnic group compared to the other due to differences in anthropometric traits that are associated with CVD risk (e.g. tall stature³⁵⁴, higher BMI³⁵⁵, higher lean body mass³⁵⁶, among others) or well-established CVD risk factors (e.g. higher mean SBP³⁵⁷ or elevated mean plasma LDL-C levels⁵⁴). Compared to Europeans in UK Biobank for example, study participants in CKB, on average, were shorter (Men: 165.2 cm in CKB vs. 175.7 cm in UKB), leaner (lean body mass in Men: 49.5 kg in CKB vs. 63.7 kg in UKB), had lower BMI (Men: 23.4 kg/m² in CKB vs. 27.8 kg/m² in UKB), lower plasma LDL-C concentrations (2.4 mmol/L in CKB vs. 3.6 mmol/L in UKB), and lower mean SBP (130.6 mmHg in CKB vs. 137.9 mmHg in UKB)³⁵⁶.

This emphasizes the need to investigate the same epidemiological associations in different population groups in order to have a broader understanding of exposure-outcome relationships in different epidemiological contexts.

Associations of CRP and fibrinogen with stroke types (IS and ICH)

Similar to what was observed for MCE, the associations of CRP and fibrinogen were discrepant with respect to their associations with stroke types. While higher CRP was associated with higher risk of IS, consistent with previous studies^{92,230,231,311}, there was no evidence of association of higher fibrinogen concentration with risk of IS, discrepant with prior studies⁹³. The reasons for these discrepant associations of fibrinogen with IS are the same as described for MCE (see above); either low power

in the present study to detect associations and/or different effect measure modifier distributions in East Asians than Europeans or simply represent true trans-ethnic differences.

Beyond differences in association patterns between CRP and fibrinogen, the present study also provided an opportunity to examine differences of association between two pathophysiologically distinct stroke types: IS and ICH³⁵⁸. When comparing associations of CRP with IS and ICH in the present study, I observed consistency in both direction and strength (Figure 3.5). These results differ from previous studies which reported null associations of CRP with ICH^{230,231,311}, although these studies (Table 3.5) included relatively fewer ICH cases (383 to 738 in previous studies vs 4,476 in the present study), typically leading to effect estimates with wider confidence intervals (e.g.²³⁰, OR [95% CI] per 1 SD higher usual log CRP: 1.04 [0.66, 1.65]) that overlap present study estimates of a 22% increase in risk (OR [95% CI] per 1 SD higher usual log CRP: 1.22 [1.10, 1.36]). Furthermore, in a separate study in CKB, other markers of inflammation that are only moderately correlated with CRP, such as glycoprotein acetyls³⁴⁷, also exhibited positive associations with both IS and ICH, providing additional support for the role of inflammation in both stroke types¹⁰¹.

3.4.3 Strength and limitations

A key strength of the present observational study was the large number of well-characterized stroke cases in CKB. The majority (90%) had CT or MRI scans at the time of hospital admission⁴¹; further adjudication by clinician-led review of medical records refuted the diagnosis in only 7% of cases with IS and 8% with ICH. In addition, measurements of CRP and fibrinogen using plasma samples of study participants were collected at baseline, prior to the occurrence of incident vascular events, thus somewhat limiting reverse causality bias in the present observational study. Furthermore, I was able to correct for measurement error and within-person variation in CRP and fibrinogen levels (i.e. regression dilution bias) by using

measurements from repeat plasma samples collected 6-9 years later, so that the results reflect the risk associated with usual levels of these biomarkers.

The present observational study had several limitations as well. The observed associations of CRP with risk of incident MCE, IS and ICH were attenuated by inclusion of additional risk factors, and it is possible that there remains residual confounding due to unknown confounders or inaccurate measurements of potential confounders (e.g. physical activity), or biases from other sources such as reverse causality. Although the prospective study design likely limits reverse causation bias, the bias from residual confounding remains difficult to account for in observational studies. Conversely, the lack of association of fibrinogen with MCE and stroke types in the present study may reflect limited statistical power (number of samples with fibrinogen measurements was half the number of samples with CRP measurements in the present study, i.e. ~9,000 vs. ~18,000). While, the present study included twice the number of IS cases compared to the largest fibrinogen meta-analysis to date (our study vs meta-analysis, 2,280 vs 1,370 IS cases in fully adjusted models)⁹³, I am currently unable to sub-classify the proportion of CKB individuals classified as an IS or ICH case to different stroke sub-types. Associations of CRP and fibrinogen may not be consistent across subtypes of IS (e.g. large artery, small vessel, and cardio-embolic) and ICH (e.g. lobar and non-lobar), and should be investigated once ongoing characterization of stroke subtypes in CKB is complete.

It should be emphasized that CRP does not play a causal role in stroke, as indicated by large-scale evidence from Mendelian randomisation studies in the past^{94,359,360}. Rather, the consistently observed associations of CRP with MCE, IS and ICH likely reflect roles for biomarkers upstream in the inflammatory signaling pathway (e.g. IL-6 or IL-1 β), as has been shown previously for IL-6 signalling (using the functional genetic variant rs2228145 in *IL6R* locus) and CHD in Europeans^{96,97}. However, the contributions of different components of IL-6 signalling to different CVD outcomes beyond CHD remain unclear, and was therefore the subject of research for Chapters 4 and 5 of my DPhil thesis.

3.5 Conclusion

In conclusion, the present study examined the observational associations of CRP and fibrinogen with MCE, IS and ICH in a large cohort of East Asian individuals. The associations observed in the present study suggested that higher concentrations of CRP was associated with higher risks of MCE, IS and ICH, consistent with associations observed in Europeans in previous studies. However, there was no evidence of associations of fibrinogen with MCE or stroke types and this was inconsistent with what was observed in Europeans where fibrinogen was strongly associated with CVD outcomes; possible reasons that account for this inconsistency include lower sample sizes for fibrinogen analysis in the present study resulting in lower statistical power to detect associations, or different distributions of (unknown) effect measure modifiers in East Asian vs. European populations relevant for the association of fibrinogen with CVD outcomes, or represent true trans-ethnic differences of association of fibrinogen concentrations with CVD risk.

4

Interleukin-6 signalling and Cardiovascular Disease

4.1 Introduction

Interleukin-6 (IL-6) is a protein composed of 184 amino acids and is synthesized by multiple cell types such as monocytes, T-cells, fibroblasts and endothelial cells³⁶¹. Interleukin-6 signals via the interleukin-6 receptor (IL-6R)³⁶². The mode of IL-6 signalling depends on the state of IL-6R, whether it is membrane-bound (*classic signalling*, mIL-6R) or in soluble form (*trans-signalling*, sIL-6R). Whether circulating IL-6 binds to mIL-6R or sIL-6R, both modes of signalling require the interaction of IL-6:IL-6R complex with another protein, glycoprotein 130 (gp130)³⁶³, a transmembrane protein which serves as the signal transducer for all members of the IL-6 cytokine family³⁶⁴. The binding of IL-6:IL-6R complex to gp130 activates the Janus kinase (JAK)-signal transducer and activator of transcription factor 3 (STAT3) pathway that stimulate transcription of IL-6 responsive genes (as illustrated in Figure 1.9). Among the IL-6 responsive genes are genes that encode the acute-phase proteins: C-reactive protein (CRP), fibrinogen, and serum amyloid A (SAA) protein. Hence, elevated levels of CRP, fibrinogen, or SAA is a strong indicator of

the magnitude of IL-6 signalling²⁷⁵.

While gp130 expression is ubiquitous across human cells, expression of mIL-6R is limited to specific cell types (liver cells and some immune cells)^{215,365}. Hence, the spectrum of cells responsive to IL-6 signalling is increased by sIL-6R. Although some sIL-6R are alternatively spliced, the majority of circulating sIL-6R is proteolytically cleaved from mIL-6R by ‘A disintegrin and metalloprotease 17’ (ADAM17) enzyme^{281,366}. A non-synonymous genetic variant (rs2228145-C, *Asp358Ala*, that changes aspartic acid, at the 358th position of the IL-6R polypeptide chain, for alanine) in the *IL6R* locus that encodes the cleavage site of mIL-6R, increases the cleavage of the extracellular portion of mIL-6R to sIL-6R. The *Asp358Ala* variant, therefore, results in higher sIL-6R plasma concentrations²⁸¹ and alters the ratio of sIL-6R to mIL-6R. This altered sIL-6R:mIL-6R ratio is leveraged in the present study (similar to previous studies^{96,97}) to represent the effects of ‘altered IL-6R signalling’. Among the associations reported with the *Asp358Ala* variant include lower plasma CRP and fibrinogen concentrations and higher circulating sIL-6R and IL-6 concentrations^{96,97}. Reduced CRP and fibrinogen, and elevated sIL-6R and IL-6 levels were also observed in clinical trials of tocilizumab⁹⁷ - a monoclonal antibody that competes with circulating IL-6 to occupy both sIL-6R and mIL-6R, broadly antagonising both modes of IL-6 signalling. The directional consistency of associations of the *Asp358Ala* variant and the effect of tocilizumab on plasma concentrations of CRP, fibrinogen, sIL-6R and IL-6 prompted the possibility that the *Asp358Ala* variant can be used to mimic the effects of tocilizumab. Furthermore, the association of *Asp358Ala* with lower risk of CHD^{96,97} raised the question of whether tocilizumab, typically indicated for chronic inflammatory conditions like rheumatoid arthritis, can be repurposed for CHD.

There are, however, two queries that remain unresolved. First, whether despite the higher concentrations of sIL-6R and IL-6 protein observed with tocilizumab and *Asp358Ala* variant, use of *Asp358Ala* instrument still represents effects of a global reduction of IL-6 signalling. Homeostatic feedback mechanisms can explain

the seemingly paradoxical effect of tocilizumab on plasma levels of sIL-6R and IL-6, i.e. lower IL-6R signalling with tocilizumab disrupts the negative feedback that would typically occur with steady-state IL-6R signalling and lower sIL-6R and IL-6, resulting in higher levels of plasma sIL-6R and IL-6. Tocilizumab was also shown to form tocilizumab:sIL-6R complexes that can prolong the half-life of sIL-6R²⁸⁴. In either case, the higher plasma levels of sIL-6R and IL-6 observed with tocilizumab does not represent an increase in IL-6 signalling. By contrast, the biological mechanism of higher plasma sIL-6R and IL-6 with the *Asp358Ala* variant remains unclear. The pro-inflammatory effects of higher sIL-6R and IL-6 are reportedly buffered by higher levels of soluble gp130 (sgp130)³⁶⁷; this was used to explain why *Asp358Ala*, in physiological states, was associated with lower levels of markers of inflammation (CRP, fibrinogen) despite being associated with higher levels of sIL-6R and IL-6. But, to date, there is no evidence to suggest that *Asp358Ala* variation also associated with higher levels of sgp130. If levels of sgp130 are unaffected by natural *Asp358Ala* variation, it stands to reason that the postulated buffering phenomenon is insufficient to explain the paradoxical association of *Asp358Ala*, where well-established markers of inflammation - CRP and fibrinogen - are in opposing directions to proteins - sIL-6R and IL-6 - that govern IL-6R trans-signalling, considered the ‘pro-inflammatory’ arm of IL-6R signalling³⁶⁷. Second, if we assume *Asp358Ala* indicates effects of IL-6 receptor-mediated reduction of global IL-6 signalling, would the use of *Asp358Ala* instrument be representative of broader effects of reduced plasma IL-6 protein concentrations?

With the present study, I aimed to address the knowledge and evidence gaps by carrying out Mendelian randomisation (MR) analysis (Figure 4.1) to investigate the relevance of the individual IL-6R and IL-6 protein components to CVD risk factors and outcomes, for equivalent differences in CRP concentrations (the second objective of my thesis, see Chapter 1.7).

As described in Chapter 1.5.4, I used locus-specific genetic instruments, specifically trans-pQTL genetic instruments of CRP - that is, genetic variants associated with

CRP located near the IL6R and IL6 loci (the trans-pQTL instrument selection was based on prior knowledge of IL-6 signalling influencing levels of circulating CRP). The aim was to understand whether CVD associations observed with the IL-6R instrument were concordant with CVD associations of the IL-6 instrument, for equivalent differences in CRP; the assessment of consistency between the two instruments would allow us to predict the likely pharmacological consequences for CVD with IL-6 signalling inhibition. The use of locus-specific instruments is, therefore, a more appropriate approach than selecting genetic instruments from across the genome - which would trade-off specificity for power. Although the limited number of genetic variants that make up the IL-6R instrument (one) and the IL-6 instrument (three) preclude use of the extra sensitivity analysis methods like MR-Egger for assessment of horizontal pleiotropy, the possibility of horizontal pleiotropy is minimised by use of the locus-specific approach.

To represent altered IL-6 signalling, I constructed IL-6 genetic instrument by testing the association of genetic variants near *IL6* locus in a GWAS of CRP in CKB (i.e. CRP was used as a marker of IL-6 protein-mediated signalling). To represent altered IL-6R signalling, I used *Asp358Ala* as the IL-6R instrument. For the present work, I have used CKB data as the primary data source on East Asians. However, where possible, I have presented and meta-analysed with estimates of similar traits using equivalent genetic instruments from the European population.

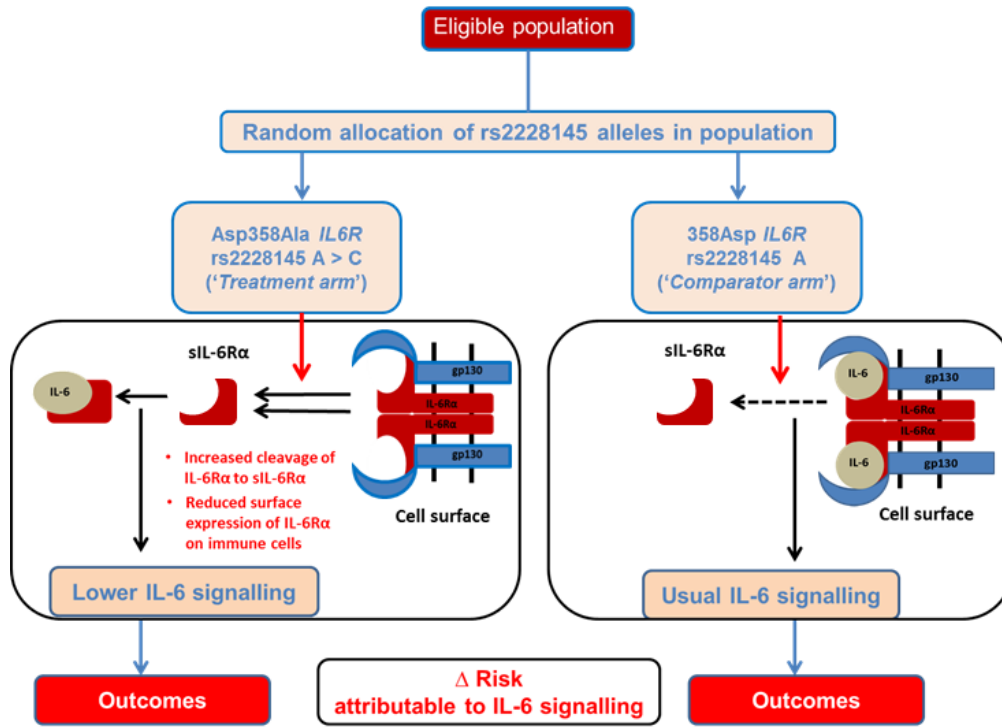


Figure 4.1: Drug-target Mendelian randomisation.

Illustrates the Mendelian randomisation study design (as analogous to an RCT design) investigating effects of rs2228145-C allele that increases cleavage of interleukin-6 receptor alpha (IL-6R α) to soluble interleukin-6 receptor alpha (sIL-6R α).

4.2 Methods

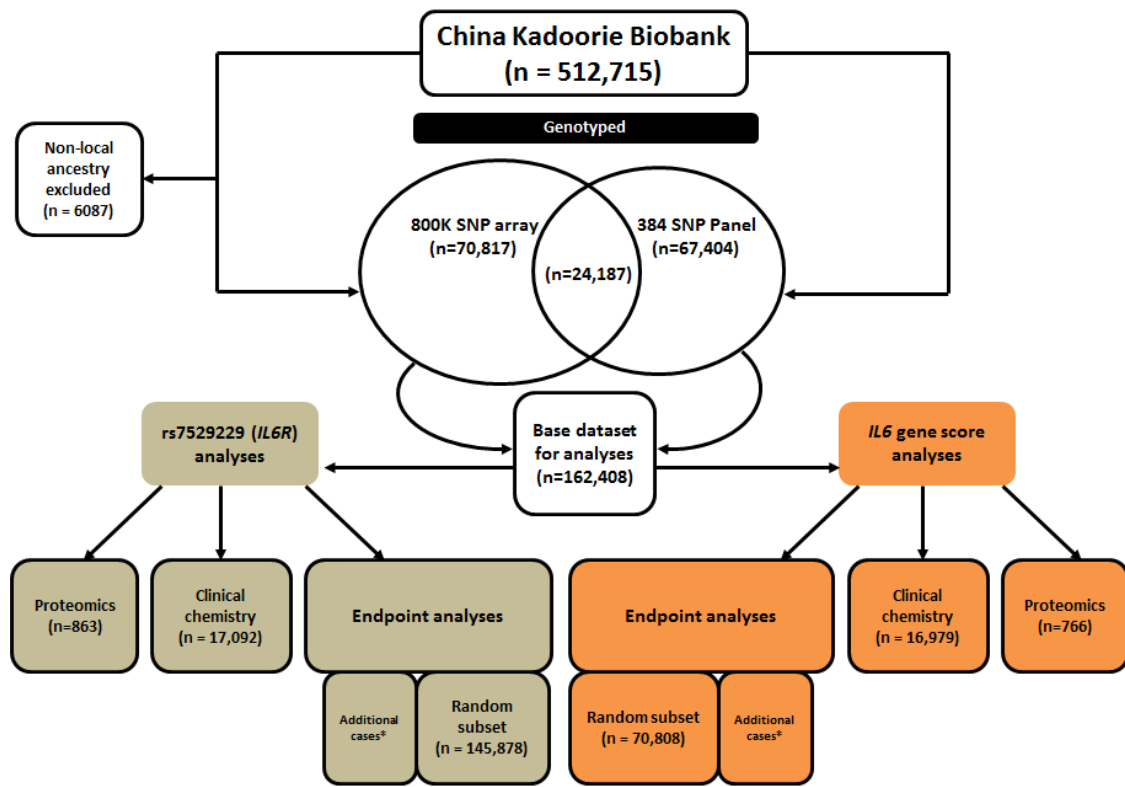
4.2.1 Study population

For the present study, I used genetic and phenotype data from both individual-level CKB data and summary data from external sources for Europeans (Chapter 2.1.1). This approach will enable the assessment of associations between East Asians and Europeans to check for consistency and provide more precise estimates on trans-ethnic meta-analysis.

In CKB, out of 512,715 East Asian study participants, genetic data were available for 162,408 study participants. The 162,408 study participants were genotyped using a 384-single-nucleotide polymorphism (SNP) panel (GoldenGate; Illumina) or a custom-designed Affymetrix Axiom (~800,000 [800K] SNPs) array at BGI (Shenzhen, China) (for all variants, call rates were more than 98% and allele frequency difference

less than 20% with 1000 genomes project [1KGP] East Asian population) (Figure 4.2). For individuals genotyped using the 800K-SNP array, 6,087 were identified as non-locals (i.e. they hail from a region that is not the reported region they were based in when included in the CKB study) using principal components (PCs) of genetic ancestry and excluded from all downstream analyses. To construct PCs, genetic data from CKB and East Asian 1000 Genomes Phase 3 population were used in the PLINK implementation of the GCTA (Genome-wide complex trait analysis) tool to obtain 142,165 SNPs with r^2 less than 0.2 (i.e. near-independent SNPs). The first 12 national PCs were used as a covariate in subsequent association studies which adjust for the region. A subset of participants with genetic data also had clinical chemistry/metabolomics data (these were participants part of a larger nested case-control study of stroke and CHD), and proteomics measurements (these were participants part of a larger nested case-control study of pancreatic cancer). Details of these measurements and assays have been provided previously in Chapter 2.2.4. Details of the study population from external sources are provided elsewhere^{291–296} and tabulated in the Appendix Table 2.1.

Case ascertainments of vascular and non-vascular (COPD) outcomes were the result of them being selected on the basis of case status at the time of case/control selection for genotyping for a genome-wide association study (GWAS) of CVD and COPD in CKB. Hence for MR analyses of non-CVD and non-COPD endpoints, using this GWAS of cardiovascular disease (CVD) or chronic obstructive pulmonary disease (COPD) may result in associations arising due to correlations with CVD or COPD, unless these ascertainments are accounted for. For example, if IL-6R or IL-6 is associated with COPD, any association of the IL-6(R) instruments with lung cancer may be due to the correlation of lung cancer with COPD. Therefore, for both IL-6R and IL-6 MR analyses in CKB, I obtained cases and controls from a study population-representative subset (or ‘random’ subset, see Figure 4.2) within the GWAS (800K SNP array) in CKB. Additional cases and controls were available among individuals genotyped by the 384-SNP panel for IL-6R analyses (rs7529229 was one of 384 SNPs in the panel); individuals in the 384-SNP panel were randomly



*Number of additional cases will vary by endpoint

Figure 4.2: Summary of genotyped participants in CKB, genotyped with 384-SNP panel or 800k SNP array.

Illustrates how genotyped participants with endpoint data are aggregated to a base dataset for assessment of genetic associations using *IL6* and *IL6R* instruments with proteomic traits, clinical chemistry traits and outcome endpoints

selected (and, therefore, considered population-representative by default) and so can be used in all analyses of IL-6R. Additional cases were included from ascertained cases myocardial infarction (MI), ischaemic stroke (IS), intracerebral haemorrhage (ICH), subarachnoid haemorrhage (SAH), fatal ischaemic heart disease (IHD) or any chronic obstructive pulmonary disease (COPD) during follow-up, and included only in analyses of the respective disease endpoints. Application of this selection protocol for IL-6R [IL-6] analyses, yielded 145,878 [70,808] cases and controls from the random subsets, with varying numbers of additional cases for each endpoint (Table 4.1). A subset of those with IL-6R [IL-6] genetic data also had clinical chemistry measurements (n= 17,092 [16,979], part of the nested case-control study of stroke and CHD - the same dataset used to obtain individuals for the observational analysis

of CRP and fibrinogen with CVD in Chapter [3](#)), and proteomics data (n=863 [766], part of a nested case-cohort study of pancreatic cancer). Details of clinical chemistry and proteomics measurements have been described in Methods Chapter [2.2.4](#).

Table 4.1: Case-control selection for *IL6R* Mendelian randomisation analyses

Group	Endpoint (ICD-10)	Random subset (case/control where appropriate)	Ascertainment group							Total ² (case/control where appropriate)
			IS	ICH	SAH	MI	Fatal IHD	COPD	Others	
Total		145878	2456	4105	353	1308	634	2993	4681	16530
Total_cases			2261	3957	319	1265	621	2624	X	11047
Primary endpoints	Major Coronary Events ¹	4952/140926	X/X	X/X	X/X	1265/X	621/X	X/X	203/X	7041/140926
	Acute Myocardial Infarction (I21)	2220/143658	X/X	X/X	X/X	1265/X	X/X	X/X	92/X	3577/143658
	Ischaemic stroke (I63, I69.3)	13601/132277	2261/X	X/X	X/X	X/X	X/X	X/X	357/X	16219/132277
	Intracerebral haemorrhage (I61, I69.1)	2761/143117	X/X	3957/X	X/X	X/X	X/X	X/X	80/X	6798/143117
Secondary endpoints	Oesophageal cancer (C15)	730/145148	X/X	X/X	X/X	X/X	X/X	X/X	32/X	762/145148
	Stomach cancer (C16)	985/144893	X/X	X/X	X/X	X/X	X/X	X/X	44/X	1029/144893
	Colorectal cancer (C18-C20)	997/144881	X/X	X/X	X/X	X/X	X/X	X/X	26/X	1023/144881
	Liver cancer (C22)	887/144991	X/X	X/X	X/X	X/X	X/X	X/X	38/X	925/144991
	Lung cancer (C33-C34)	1697/144181	X/X	X/X	X/X	X/X	X/X	X/X	103/X	1800/144181
	Breast cancer (females only) (C50)	780/86188	X/X	X/X	X/X	X/X	X/X	X/X	17/X	797/86188
	Diabetes (Incident [E10-E14] and prevalent at baseline)	13622/132256	X/X	X/X	X/X	X/X	X/X	X/X	433/X	14055/132256
	COPD (incident [J41-44] and GLI-defined at baseline)	13665/132210	X/X	X/X	X/X	X/X	X/X	2624/X	554/X	16843/132210
	Asthma (J45-46)	755/145123	X/X	X/X	X/X	X/X	X/X	X/X	38/X	793/145123
	Tuberculosis (A15-A19, J65, K23.0, K67.3, M01.1, M49.0, M90.0, N33.0, N74.0, N74.1, O98.0, P37.0)	793/145085	X/X	X/X	X/X	X/X	X/X	X/X	41/X	834/145085
	Pneumonia (J12-J18)	7854/138024	X/X	X/X	X/X	X/X	X/X	X/X	338/X	8192/138024
	Depression (CIDI A and F32)	1228/3617	X/X	X/X	X/X	X/X	X/X	X/X	34/X	1262/3617

IS: Ischaemic stroke; ICH: Intracerebral haemorrhage; SAH: Subarachnoid haemorrhage; MI: Myocardial infarction; IHD: Ischaemic Heart Disease; COPD: Chronic Obstructive Pulmonary Disease

¹ Major coronary events includes myocardial infarction [I21], coronary revascularization, or death from IHD [I20-I25]

² Total column represents the sum of case/controls from the random subset and cases from the ascertainment groups as appropriate for a given endpoint; X denotes ascertainment groups not selected as a source of cases or controls for a given endpoint.

4.2.2 Selection of cases and controls

In CKB, the definitions of the main pre-specified vascular outcomes [ICD-10 codes] are outlined in Table 3.2 for the observational study and included major coronary events (myocardial infarction [I21], coronary revascularization, or death from IHD [I20-I25]), ischaemic stroke [I63, I69.3], intracerebral haemorrhage [I61, I69.1] or presence of carotid plaque in either internal carotid artery (identified on ultrasound study during the second resurvey). Controls for vascular outcomes excluded participants reporting a history of IHD, stroke, or transient ischaemic attack at baseline or during follow-up.

4.2.3 Statistical analysis

To create a genetic instrument for IL-6 protein, a multivariable linear regression model regressed residuals of log-transformed hsCRP (adjusted for age, age², sex, ascertainment status, region, and 12 principal components of genetic ancestry) on the three *IL6* SNPs to obtain independent per-allele estimates with hsCRP; these estimates were used to construct a weighted *IL6* gene score using the following formula,

$$wGS = \sum_{i=1}^n W_i X_i$$

where i is each *IL6* SNP, n is the number of *IL6* SNPs (3 in our case), W_i is the independent per-allele estimates of each *IL6* SNP with hsCRP, and X_i is the effect allele dosage. However, derivation of weights using data under analysis (in this case, CKB participants with hsCRP measurements) can lead to substantial bias, typically in the direction of the confounded association³⁶⁸. Therefore, for participants with hsCRP measurements in CKB, unbiased internal weights were derived by 100-fold cross-validation (i.e. weights for each 1% of the sample were derived iteratively from 99% of the data). For participants without hsCRP measurements, the per allele estimates for each SNP from the multivariable model was used to derive weights.

For each SNP in the score, the allele associated with lower CRP was selected as the effect allele.

In CKB, associations of *IL6R* (rs7529229-C) and the *IL6* instruments (gene score and rs13306436) with quantitative and binary traits (including disease outcomes) were scaled to correspond to the effect of the halving of genetically predicted hsCRP levels. The scaling factor in CKB was derived separately for each genetic instrument by regressing log2-transformed hsCRP on the genetic instrument adjusted for age, age², sex, ascertainment and 9 regional PCs (regional PCs account for the local heterogeneity between individuals in each CKB region); this model was fitted separately for each region and estimates were meta-analysed by fixed effect inverse-variance weighted method.

For Europeans, I obtained estimates of CRP association from UK Biobank (UKB) (Neale Lab²⁹⁶) using the same (e.g. *IL6R* variant in UKB is the same *IL6R* variant used in CKB: rs7529229) or near-equivalent (e.g. *IL6* SNP in UKB is rs1880241 vs. *IL6* gene score composed of 3 SNPs in CKB) genetic instrument. To convert standardized effect estimates of genetic variants with CRP obtained through Neale Lab (UKB biomarkers were rank-inverse normally transformed by Neale Lab) to log-transformed CRP, I log-transformed the values for the deciles of the CRP distribution as provided in UKB data showcase of CRP (<http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=30710>). The log-transformed decile values of CRP were plotted against the Z-scores corresponding to those deciles; the slope of that line was used to derive the standard deviation (SD) of log-transformed deciles (using the *qnorm* function in R). I used the derived SD of logCRP to convert beta from the SD scale, as processed by Neale lab for UKB CRP GWAS, to the logCRP scale for use in my thesis. To calculate the variance of CRP explained by each genetic variant or gene score as well as their F-statistic, I derived the variance directly from a univariate regression model (containing the SNP with age, age², sex, region and principal components of genetic ancestry) in CKB. In UKB, the variance (R^2) was estimated using the formula:

$$R^2 = 2f(1 - f)\beta^2$$

where f is the minor allele frequency and β is the per-allele effect on standardized CRP; the R^2 obtained was used to estimate F-statistic (F) or instrument strength of UKB SNPs by the formula

$$F = \frac{R^2(n - 1 - K)}{(1 - R^2)K}$$

where R^2 is the variance, n is the sample size, and K is the number of genetic predictors (however, when used to estimate the F-statistic of a genetic score using multiple SNPs, k is 1).

Similar to Chapter 3.2.5, means and proportions of baseline characteristics across *IL6R* genotype groups or tertiles of *IL6* gene scores were standardized (as appropriate) to age, sex, study area or case status using R package *lsmeans*. Natural log transformations were used to correct skewed biomarkers distributions. Where comparisons of effects between different biomarkers on a trait are the objective, I standardized distributions of the biomarkers (using *scale* function in R with default settings) to enable comparisons to be made on the same scale (i.e. in SD units). All quantitative and binary traits (which include disease outcomes) were assessed by fitting linear (*lm* function in R) and logistic regression (*glm* function in R) models respectively in each region adjusted for age, age² and sex, with estimates meta-analysed by fixed effects inverse-variance weighted method (using the *rma* function in metafor R package³¹⁴), unless otherwise stated. A per-allele (additive) model was used for analyses of individual variants. Alleles or gene scores were oriented to represent genetically predicted lower CRP concentrations.

P values presented are unadjusted for multiple testing unless otherwise stated. Many of the tests planned for the analyses in my thesis were not completely independent. Hence, for assessment of significance, I used a 5% false-positive discovery rate

(Benajmini-Hochberg method)³⁶⁹ rather than a more stringent familywise error rate (FWER) controlling procedures (e.g. Bonferroni correction). Power analyses used the program Quanto³⁷⁰. Analyses were performed using R version 3.5.0.

4.3 Results

4.3.1 Selection of SNP to construct the IL-6R genetic instrument

The SNP with the strongest association with sIL-6R in Sun et al's protein GWAS dataset (which was the largest protein GWAS at the time of writing this thesis) was rs12126142 ($P = 2.24 \times 10^{-1099}$)³⁷¹; rs12126142 was also strongly associated with CRP in CKB ($P = 7.90 \times 10^{-28}$). Using r-squared (r^2) to assess for evidence of linkage disequilibrium (LD), I examined the LD between rs12126142 and the functional SNP rs2228145 in CKB. Both SNPs were in high LD at an r^2 of 0.99. However, in CKB, both rs2228145 and rs12126142 were only present in the 800k SNP array. To maximise sample size with genetic data, I explored SNPs in LD rs12126142/rs2228145 that were present in both the 384-SNP panel and 800k SNP array. Three SNPs qualified this criterion, i.e. they were present in both the panel and array: rs12740969, rs4845625, and rs7529229. Of these 3 SNPs, rs7529229 was the SNP with the strongest association with CRP and is also in high LD with rs2228145 ($r^2 = 0.99$). Furthermore, when adjusted for rs7529229, associations of rs12126142, rs12740969, and rs4845625 with CRP in CKB become null (Table 4.2). Therefore, as the IL-6R instrument in CKB, I selected rs7529229 as this SNP was genotyped both in the 800K SNP array and the 384-SNP panel and, therefore, had a larger sample size ($n = \sim 150,000$) compared to rs12126142 or the functional SNP rs2228145 - both of which were genotyped only in the 800K SNP array ($n = \sim 100,000$). Furthermore, rs7529229 was previously used to represent altered IL-6R signalling in Europeans^{96,97} and, thus, it was also used to represent altered IL-6R signalling for Europeans in the present study.

Table 4.2: Candidate SNPs in *IL6R* locus associated with CRP in CKB

SNP_EA	N	beta ³	SE ³	P ³	r ² with rs2228145	P ³ (adjusted for rs7529229)
rs12126142_A ¹	101157 ²	-0.116	0.011	7.17E-27	0.991	0.46
rs7529229_C ¹	168561	-0.116	0.011	8.80E-27	0.987	-
rs12740969_G	168569	-0.042	0.038	0.00348	0.124	0.84
rs4845625_C	168549	-0.092	0.012	1.71E-18	0.636	0.44

EA: Effect Allele; SE: Standard error; SNP: single-nucleotide polymorphism; P: P-value

¹ Note: rs12126142-A and rs7529229-C are in high LD with each other ($r^2 > 0.9$)

² The lower sample size for rs12126142-A is because it was present only in the 800k array (and not part of the 384-SNP panel)

³ Beta, SE, and P are extracted from a regression model where rank-inverse normally transformed (logCRP residuals) were regressed on the respective SNPs. (logCRP residuals derived by regressing logCRP against age, age², sex, 4 levels ascertainment and region)

4.3.2 Selection of SNPs to construct the IL-6 genetic instrument

To represent IL-6 protein activity (using CRP concentrations as proxy), I used separate genetic instruments in CKB and Europeans. For CKB, the following approach was used (Figure 4.3):

1. Summary statistics were available for a GWAS of hsCRP in CKB in 8,981 males and 8,705 females of Han Chinese ancestry (mean age [SD]: 57 [10.5]): rank-inverse normally transformed residuals of log-transformed hsCRP were regressed on age, age-squared, sex, region, 12 PCs representing East Ancestry and genotype arrays. This was implemented in BOLT-LMM v.2.3.2 using a linear mixed model.
2. FINEMAP³⁷² – a tool that employs a Bayesian approach to prioritize credible sets of potentially causal SNPs. Fine-mapping was conducted using summary statistics of CKB CRP GWAS and a SNP correlation matrix using CKB genetic data by a CKB bioinformatician (Kuang Lin) to obtain possible SNP configurations at the *IL6* locus (chr7:22.5 Mbp – 23.0 Mbp, approximately 300 Kb from around the *IL6* gene).
3. Among the possible causal configurations, the first 6 configuration models were selected (ranked by decreasing posterior probability).

4. For the SNPs in each of the first 6 configuration models, I fitted 6 multivariable regression models (where SNPs in each causal configuration model were adjusted for each other) using as the dependent variable residuals from the regression of log-transformed hsCRP on age, age², sex, region, ascertainment, and the first 12 PCs of genetic ancestry.
5. From the regression models above, the SNP set that explained the greatest variance of hsCRP was selected to construct the *IL6* gene score. F-statistic of each SNP and the *IL6* gene score was extracted from the multivariable model and used as a measure of instrument strength.

Using the CRP GWAS in CKB, a search for additional genetic instruments representing different but related components of systemic inflammation (e.g. *IL1 β* , *JAK1*) did not yield any additional signals (Appendix Figure A.3).

To represent IL-6 protein activity (via CRP concentrations) in Europeans, I used the most recent GWAS of CRP in Europeans²⁸⁸ to identify the genetic variant in the *IL6* locus reported to most strongly associate with CRP; the selected rs1880241 in the *IL6* locus has been reported to affect the expression of *IL6* gene in monocytes (Table A.1). The association of SNP rs1880241 with CRP was checked for replication in the UK Biobank (data processed by Neale Lab²⁹⁶). Estimates of variance and strength of IL-6R instrument (rs7529229) and IL-6 instrument (rs1880241) with CRP for Europeans were also extracted from UK Biobank.

Using the above strategy, three SNPs associated with CRP in CKB in the *IL6* locus were selected to construct the IL-6 gene score as an IL-6 genetic instrument. The logic of selection of these three SNPs in CKB is provided in Figure 4.3. A GWAS of CRP in CKB showed a signal near the *IL6* locus as shown in Figure 4.3. This set of SNPs was one of the top six possible configurations in the set obtained by fine-mapping the region containing the SNPs in the *IL6* locus associated with CRP. To identify which of the six sets explained the most variance in CRP, I fitted 6 regression models, regressing on residuals of log-transformed hsCRP (adjusted for

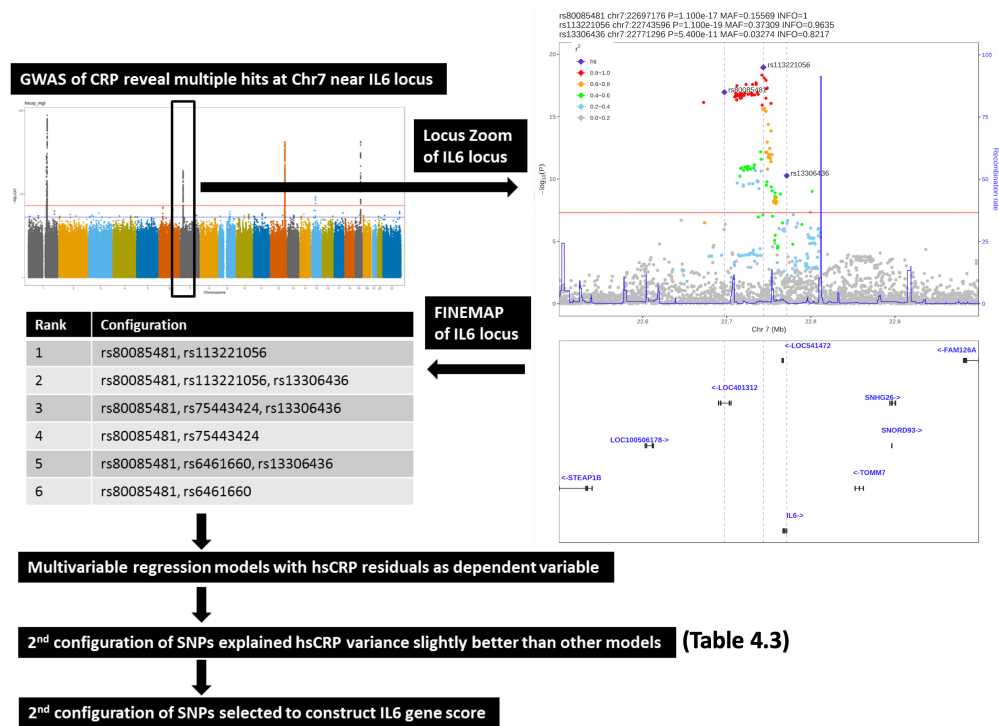


Figure 4.3: Selection of *IL6* SNPs.
The figure illustrates the logic of selection of genetic variants for construction of *IL6* gene score. Associations of genetic variants with CRP in the *IL6* locus in China Kadoorie Biobank was FINE-MAPPED [372] and credible sets of genetic variants were generated for use in the *IL6* gene score.

age, age², sex, region, ascertainment and the first 12 PCs of genetic ancestry) on the combination of SNPs in each set. As shown in Table 4.3, the SNPs in the second model (rs80085481, rs113221056, rs13306436) that explained the most variance ($r^2=0.00727$) in CRP were selected to construct the IL-6 gene score. These variances (r^2) were extracted directly from the regression model.

Table 4.3: Multivariable regression models of first 6 fine-mapped *IL6* SNP sets from FINEMAP.

Estimates of SNP-CRP association were derived from joint regression models of the first 6 sets of SNPs generated using FINEMAP. The regression models used CRP residuals as the dependent variable generated by regressing log CRP against age, age², sex, region, case ascertainment and 12 national principal components of ancestry. P-values and r^2 were extracted directly from the joint regression models

Model	SNP_EA			Effect per allele			SE			P			r^2
	SNP1	SNP2	SNP3	BETA1	BETA2	BETA3	SE1	SE2	SE3	P1	P2	P3	
1	rs80085481_G	rs113221056_A		0.111	0.0802		0.0173	0.0132		1.71×10^{-10}	1.15×10^{-09}		0.00600
2	rs80085481_G	rs113221056_A	rs13306436_G	0.109	0.0648	-0.186	0.0173	0.0135	0.0392	2.88×10^{-10}	1.75×10^{-06}	2.04×10^{-06}	0.00727
3	rs80085481_G	rs75443424_T	rs13306436_G	0.109	0.0646	-0.186	0.0173	0.0136	0.0392	2.93×10^{-10}	1.88×10^{-06}	2.00×10^{-06}	0.00726
4	rs80085481_G	rs75443424_T		0.111	0.0800		0.0173	0.0132		1.75×10^{-10}	1.26×10^{-09}		0.00599
5	rs80085481_G	rs6461660_T	rs13306436_G	0.109	0.0645	-0.186	0.0173	0.0136	0.0392	2.94×10^{-10}	2.00×10^{-06}	1.98×10^{-06}	0.00726
6	rs80085481_G	rs6461660_T		0.111	0.0799		0.0173	0.0132		1.76×10^{-10}	1.35×10^{-09}		0.00599

EA: Effect Allele; SE: Standard error; SNP: single-nucleotide polymorphism; P: P-value;

To compare how the IL-6R (rs7529229-C) and IL-6 instruments differ in performance between CKB and UKB, I have tabulated the variance and F-statistics for the effect of each SNP on logCRP (Table 4.4). As the table shows, the variance of CRP explained by each instrument are small (less than 1%); this is expected when small numbers of *cis*-acting variants are used as instruments. Although the use of genome-wide genetic variants can be used to increase the variance explained²⁸⁷, this would be at the cost of specificity and increase risk of horizontal pleiotropy (Chapter 1.5.4). However, despite the low variance explained, the IL-6(R) instruments from each of CKB and UKB are strong, with F-statistics of more than 100 for single-SNP IL-6R in CKB and UKB and for the IL-6 gene score in CKB, and of 88 for the single-SNP IL-6 instrument in UKB. The high F-statistic also suggests that both the IL-6(R) instruments have limited vulnerability to weak instrument bias, which can potentially bias causal estimates towards observational associations³⁷³.

Table 4.4: Comparison of IL6R and IL6 Instruments between CKB and UKB. The table shows effect size, variance explained, and strength of *IL6* and *IL6R* instruments in CKB and UKB with relation to CRP. Estimates of effect size, variance, and F-statistic for UKB were obtained or derived from Neale Lab data [296].

Instrument-EA	N	Chr:Pos ^a	AA	Effect ^b per allele on log CRP (95% CI)	Variance ^c explained (%)	F-statistic ^d
<i>IL6R</i>						
CKB (rs7529229-T)	17085	chr1:154420778	C	0.127 (0.103 to 0.152)	0.59	105.52
UKB (rs7529229-T)	343524	chr1:154420778	C	0.010 (0.006, 0.015)	0.42	1432.98
<i>IL6</i>						
CKB (rs13306436-A)	16972	chr7:22771296	G	0.186 (0.110 to 0.263)	0.25	45.29
CKB (rs113221056-A)	16972	chr7:22743596	C	0.063 (0.037 to 0.090)	0.36	64.23
CKB (rs80085481-G)	16972	chr7:22697176	A	0.110 (0.076 to 0.143)	0.39	69.77
CKB - IL6 gene score (jack-knifed) ^e	-	-	-	-	0.68	123.59
UKB (rs1880241-A)	343524	chr7:22759469	G	0.043 (0.038, 0.047)	0.026	88.02

AA: Alternate Allele; CKB: China Kadoorie Biobank; EA: Effect Allele; UKB: UK Biobank

^aBuild 37 (GRCh37.p13)

^bFor CKB IL6 SNPs, effects are derived from a joint regression model

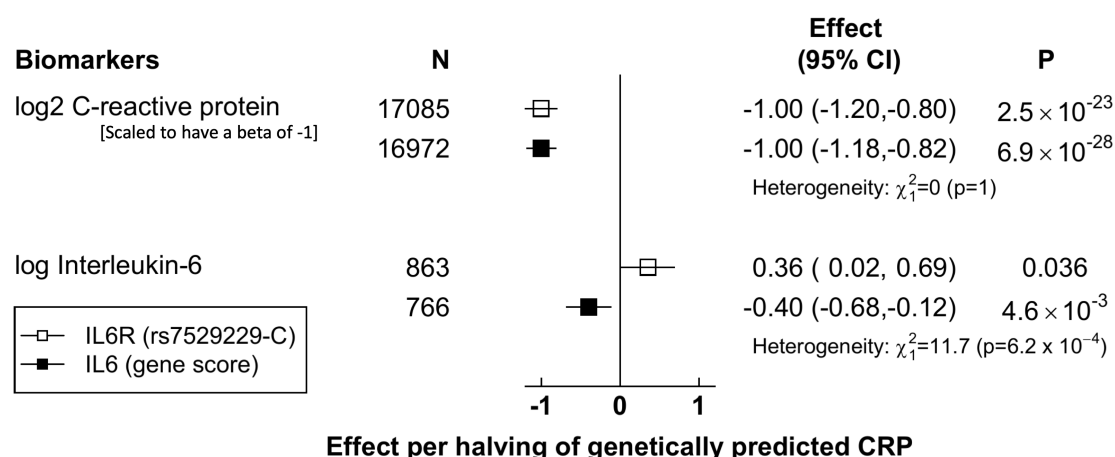
^cVariance for CKB SNPs and IL6 gene score are extracted from a univariate regression model; variance for UKB SNPs derived using the formula $2f(1-f)\beta^2$ where f is minor allele frequency and β is the per allele effect on CRP (expressed in SD units).

^dF-statistic of CKB SNPs and IL6 gene score are extracted from a univariate regression model; F-statistic of UKB SNPs are derived using the formula $R^2 \times (n-1-K)/[(1-R^2) \times K]$, where R^2 is the variance explained, n = sample size, and K is the number of SNPs.

^e Although the *IL6* gene score combines information from the preceding three SNPs (rs13306436, rs113221056, rs80085481), the score explains less variance than the sum of the variance of the three SNPs. This is because the weights for *IL6* gene score were derived by an iterative process of estimating effect of each SNP on CRP for each 10% of study population from the 90% (jack-knifing procedure) of study population. The sum of the product of the number of effect alleles of each SNP and the jack-knifed weights created the IL-6 gene score.

4.3.3 Association of the IL-6(R) instruments with CRP and IL-6 protein

Consistent with how the genetic instruments were selected in both East Asians and Europeans, the IL-6(R) instruments were associated with both CRP and IL-6 protein in CKB and UKB (Figure 4.4 and 4.5). Furthermore, the association of the multi-SNP IL-6 instrument with CRP was not driven by any one particular SNP (Figure 4.5). Nevertheless, it remains possible that SNPs selected from a GWAS of CRP in CKB to create the IL-6 gene score may have been associated with CRP due to the selection being conditioned on cardiovascular disease (collider bias). However,



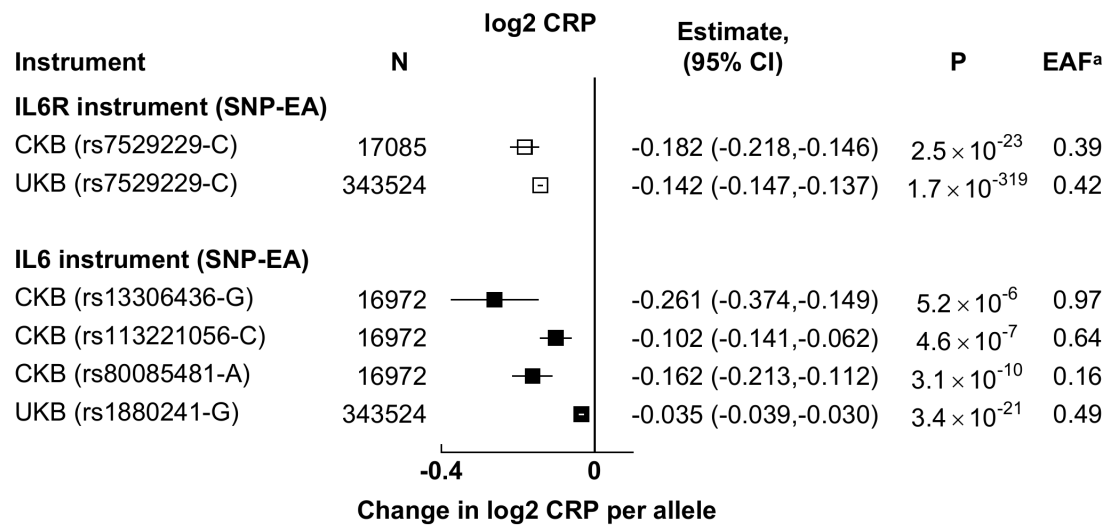
CKB: China Kadoorie Biobank;

Figure 4.4: Associations of the IL-6(R) instruments, scaled to halving of genetically predicted CRP concentrations, with markers of systemic inflammation in CKB. Error bars represent 95% CI. Area of boxes are not proportional to inverse of the variance of effect.

I observed consistent associations between a ‘random’ subset of study participants and the complete dataset (that includes both random and non-random population subsets) of the selected SNPs used in the IL-6(R) instruments with CRP and IL-6 protein (Figure A.4).

Associations of IL-6R instrument with CRP and IL-6 protein

Consistent with selection of variants for the IL-6R instrument, the IL-6R instrument was associated at similar magnitudes in both CKB (change in log2 CRP [95% CI] per allele: -0.182 [-0.218, -0.146]) and UKB (change in log2 CRP [95% CI] per allele: -0.142 [-0.147, -0.137]) study populations (Figure 4.5). A subset of the CKB study population with genetic data also had IL-6 proteomics data that I used to check whether the IL-6R instrument was associated with IL-6 protein (Figure 4.4). With effects scaled to a halving of genetically predicted CRP, IL-6R was associated with a 0.36 (95% CI: 0.02, 0.69) unit increase of log-transformed IL-6 protein in the CKB study population. The association of the IL-6R instrument (orientated to genetically predicted lower CRP concentrations) with IL-6 protein is directionally consistent with what was observed in RCTs of tocilizumab, i.e. tocilizumab that



CKB: China Kadoorie Biobank; UKB: UK Biobank; EA: Effect Allele; EAF: Effect Allele Frequency

^aEAF is the EAF of the relevant SNP in the respective dataset and not from a reference panel.

Figure 4.5: Per-allele associations of the IL-6(R) instruments with C-reactive protein in East Asians and Europeans. Error bars represent 95% CI. Area of boxes are not proportional to inverse of the variance of effect.

blocks both sIL-6R and mIL-6R, is associated with lower CRP and higher IL-6 protein concentrations⁹⁷.

Associations of IL-6 instrument with CRP and IL-6 protein

Because CRP concentrations were used to construct the IL-6 instrument in CKB and for Europeans, I observed a similar magnitude of effects (similar to IL-6R) with SNPs representing the IL-6 instrument in CKB (Figure 4.4). Like IL-6R, I was also able to separately assess the association of the IL-6 instrument with IL-6 protein in CKB. With effects scaled to a halving of genetically predicted CRP, IL-6 instrument was associated with a 0.40 unit (95% CI: -0.68, -0.12) decrease of log-transformed IL-6 protein in the CKB study population. The association of the IL-6 instrument with IL-6 protein was in the opposite direction to the effect of the IL-6R instrument on IL-6 protein for equivalent differences in CRP, validating the method used to select SNPs for the IL-6 instrument for CKB, and provides confidence that the SNP similarly selected for UKB is also instrumenting IL-6 protein levels

4.3.4 Summary and interpretation of genetic instruments selected

A summary of the SNPs used to create the IL-6(R) instruments for analysis in the present thesis for CKB and Europeans are provided in Table 4.5.

As explained in the introduction to this Chapter (Section 4.1), the SNP rs2228145-C (that is the IL-6R instrument in my thesis) represents lower classic signalling (indicated by lower plasma CRP concentrations); however, given the positive association of the IL-6R instrument with plasma IL-6 protein levels (Section 4.3.3), it remains unclear whether it represents higher or lower trans-signalling, making attribution of genetic epidemiological associations of the IL-6R instrument to either lower classic and lower trans- IL-6 signalling or lower classic and higher trans- IL-6 signalling challenging. Hence, to refer to the IL-6 signalling of the IL-6R instrument, I used the term ‘altered IL-6R signalling’.

The interpretation, however, is relatively straightforward for the IL-6 instrument. The IL-6 instrument, composed of three near-independent SNPs in *IL6* locus, is associated with lower plasma CRP and lower plasma IL-6 protein concentrations (Section 4.3.3), and hence likely represents, both lower classic and lower trans-IL-6 signalling. However, like the IL-6R instrument, whether the IL-6 instruments truly represent lower trans-signalling is pending experimental verification. Until then, like the IL-6R instrument, the IL-6 instrument also represents ‘altered IL-6 signalling’.

Table 4.5: Summary of the IL-6(R) instruments for CKB and Europeans

	IL-6R instrument	IL-6 instrument
CKB (East Asians)	rs7529229-C	IL-6 gene score ¹
Europeans	rs7529229-C	rs1880241-C

¹ IL-6 gene score composed of rs80085481, rs113221056, and rs13306436

4.3.5 Baseline characteristics of study participants

To characterise baseline traits for the present study population, I used a similar approach as used in Chapter 3 (Table 3.3). The objective of characterising baseline

traits of study participants was to understand whether there were any systematic differences of relevant traits by IL-6R genotype groups and IL-6 gene score tertiles. Assessment of this table can provide insights on whether any observed differences in risks of disease outcomes with the IL-6R or IL-6 instruments are due to differences in the distribution of baseline traits across genetic subgroups of the IL-6(R) instruments. In addition, this exercise can help assess the assumption of Mendelian randomisation, that randomisation of alleles at the population level is analogous to the process of randomisation in RCTs, ensuring genetically-defined groups are comparable to each other with respect to the distribution of known and unknown confounders. To minimise variations of descriptive estimates of baseline traits due to age, sex, study area or case status, means and proportions were standardized (as appropriate) to age, sex, study area or case status (Chapter 4.2.3).

Among population-representative individuals with genotype data on IL-6R ($n = 145,878$) and IL-6 SNPs ($n = 70,808$), mean (SD) age was 52 (10.5) and 52.1 (10.5) years, respectively (Table 4.6). The proportion of women was approximately equal in IL-6R and IL-6 groups (61%). A slightly higher proportion of individuals were from rural areas in the IL-6R group than IL-6 group (57.6 vs 54.0 %). In a subset of individuals with baseline data on inflammation markers and lipids, selected for a nested case-control study of CVD, the adjusted (adjusted for case status) geometric mean (95% CI) CRP level for participants with genotype data on IL-6R [IL-6 gene score] was 1.03 (1.00, 1.06) mg/L [0.97 (0.93, 1.02) mg/L]; fibrinogen concentration, 3.07 (3.04, 3.10) g/L [3.03 (2.99, 3.07) g/L]; triglyceride concentration, 152 (150, 155) mg/dL [147 (144, 151) mg/dL]; mean (SD) of LDL-cholesterol was 93 (25) mg/dL [92 (25) mg/dL]. Hence, the distribution of inflammation markers and lipids appears to be approximately equal in IL-6R and IL-6 groups. This suggests that any differences in risk of disease observed using the IL-6(R) instruments will unlikely be due to differences in the distribution of baseline traits between IL-6R and IL-6 groups.

A comparison of traits across genetically-defined groups was also undertaken, using genotype groups for IL-6R and tertiles of IL-6 gene score (Table 4.7). Across the genetic subgroups, no major systematic differences were observed except for region; lower proportions of individuals were from rural areas per T allele of rs7529229 (IL6R) or higher IL6 gene score (P for trend = 1.8×10^{-5} for IL-6R, and 4.1×10^{-9} for IL-6 gene score). This suggests that any differences in risk of disease observed using the IL-6(R) instruments will unlikely be due to systematic differences in the distribution of baseline traits (except region) across IL-6R genotype subgroups or IL-6 gene score. To correct for the effect of regional variation on associations of IL6R and IL6 instruments with disease (in subsequent sections) in CKB, where possible I have stratified and meta-analysed estimates across regions or adjusted in regression models.

In summary, both Tables 4.6 and 4.7 showed no major systematic differences with respect to their baseline traits between individuals genotyped for IL-6R and IL-6 SNPs, or across their respective genetic subgroups. Hence, any differences in disease risk observed in MR analyses using the IL-6(R) genetic instruments in subsequent sections would likely be due to differences in IL-6R and IL-6 led signalling, and not due to differences in baseline characteristics.

Table 4.6: Baseline traits of the study population.

Selected baseline traits for the individuals genotyped for *IL6R* and *IL6*. Mean (standard deviation) for continuous traits or proportion (weighted percentage) for binary/categorical traits are standardized to age, sex and study area or case-control cohort unless otherwise stated.

Trait	<i>IL6R</i> (rs7529229)	<i>IL6</i> (gene score)
No. of participants (random subset)	145,878	70,808
Age and Sociodemographic factors		
Age, y ^a	52 (10.5)	52.1 (10.5)
Female ^b , %	60.6	60.9
Rural ^c , %	57.6	54.0
Income >10,000 yuan/year, %	72.7	73.0
Lifestyle factors		
Male smokers, %	87.8	87.4
Male drinkers, %	36.6	37.5
Physical activity, MET h/day	22.1 (12.0)	21.5 (12.0)
Anthropometric and physiological factors		
BMI, kg/m ²	24.0 (3.25)	24.1 (3.26)
Waist-hip ratio	0.89 (0.07)	0.88 (0.07)
Mean SBP, mmHg	131.6 (19.7)	131.6 (19.7)
Mean DBP, mmHg	79.3 (10.9)	79.4 (10.9)
Markers of subclinical atherosclerosis		
Plaque, %	40.7	40.9
cIMT, mm	0.71 (0.14)	0.71 (0.14)
Inflammatory markers		
CRP, mg/L ^d	1.03 (1.00-1.06)	0.97 (0.93-1.02)
Fibrinogen, g/L ⁴	3.07 (3.04-3.10)	3.03 (2.99-3.07)
Lipids		
TG ^d , mg/dL	152.0 (149.5-154.6)	147.0 (143.6-150.6)
LDL-C ^e , mg/dL	93.3 (25.0)	92.4 (24.6)
Self-reported health and disease		
Hypertension, %	11.1	11.2
Diabetes, %	6.1	6.2

BMI: body-mass index; CRP: C-reactive protein; cIMT: carotid intima-media thickness; DBP: diastolic blood pressure; HR: heart rate ;LDL-C: low-density lipoprotein cholesterol; MET: metabolic equivalents; SBP: systolic blood pressure; SRH: self-reported health; TG: triglycerides;

^aStandardized to sex and study area only

^bStandardized to age and study area only

^cStandardized to age and sex only

^dGeometric mean (95% CI)

^eLow-density lipoprotein (LDL) cholesterol was directly measured

Table 4.7: Baseline traits of the study population according to genotype subgroups of IL6R and tertiles of IL6 gene score. Selected baseline traits for subgroups according to *IL6R* and *IL6* gene score are shown. Values are mean (standard deviation) for continuous traits or proportion (weighted percentage) for binary/categorical traits, standardized to age, sex and study area of case-control cohort unless otherwise stated.

Traits	<i>IL6R</i> (rs7529229)				<i>IL6</i> (gene score) tertiles			
	CC	CT	TT	P*	Lower	Middle	Upper	P*
No. of participants	22813	69042	54023		29341	22917	18550	
Age and Sociodemographic factors								
Age, y ^a	52 (10.6)	52 (10.5)	52 (10.6)	0.429	52.2 (10.6)	52.2 (10.5)	52.1 (10.6)	0.146
Female ^b , %	60.4	60.6	60.7	0.506	60.5	61.2	60.9	0.462
Rural ^c , %	58.3	58	56.9	1.8x10 ⁻⁵	55.7	54.1	52.9	4.1x10 ⁻⁹
Income>10,000 yuan/year, %	72.4	72.7	72.7	0.442	73.1	73	72.8	0.422
Lifestyle factors								
Male smokers, %	87.8	88.1	87.6	0.305	87.2	87.7	87.3	0.964
Male drinkers, %	36.1	36.6	36.7	0.35	37.7	37.5	37.5	0.793
Physical activity, MET h/day	22.11 (12.16)	22.15 (12.01)	22.07 (11.79)	0.496	21.45 (12.09)	21.51 (11.9)	21.5 (11.84)	0.677
Anthropometric and physiological factors								
BMI, kg/m ²	23.96 (3.23)	23.99 (3.25)	23.96 (3.25)	0.504	24.05 (3.25)	24.05 (3.27)	24.08 (3.25)	0.295
Waist-hip ratio	131.43 (19.71)	131.62 (19.78)	131.51 (19.56)	0.926	131.48 (19.55)	131.6 (19.78)	131.75 (19.75)	0.151
Mean SBP, mmHg	79.24 (10.95)	79.35 (10.94)	79.26 (10.78)	0.791	79.26 (10.83)	79.38 (10.96)	79.48 (10.85)	0.042
Mean DBP, mmHg	78.4 (11.68)	78.56 (11.78)	78.6 (11.63)	0.0473	78.53 (11.62)	78.45 (11.64)	78.33 (11.69)	0.073
Markers of subclinical atherosclerosis								
Plaque, %	40	40.4	41.4	0.0758	41.2	41.1	40.5	0.295
cIMT, mm	0.7 (0.14)	0.71 (0.14)	0.71 (0.14)	0.144	0.71 (0.14)	0.71 (0.14)	0.7 (0.14)	0.196
Lipids								
TG ^d , mg/dL	151.1 (146.2-156.2)	151.3 (148.1-154.5)	153.4 (149.9-157.1)	0.335	147.2 (142.3-152.3)	146.9 (141.6-152.5)	147.0 (142.1-152.0)	0.935
LDL-C ^e , mg/dL	92.3 (25.0)	93.2 (25.0)	93.8 (24.9)	0.0656	91.8 (24.3)	93.6 (24.6)	92.2 (24.8)	0.655
Self-reported health and disease								
Hypertension, %	11.2	11.2	11	0.411	10.9	11.5	11.3	0.262
Diabetes, %	5.98	6.09	6.23	0.15	6.05	6.41	6.12	0.916

BMI: body-mass index; CRP: C-reactive protein; cIMT: carotid intima-media thickness; DBP: Diastolic Blood Pressure; HR: heart rate ;LDL-C: Low-density lipoprotein cholesterol; MET: metabolic equivalents; SBP: Systolic Blood Pressure; SRH: self-reported health; TG: Triglycerides;

^aStandardized to sex and study area only

^bStandardized to age and study area only

^cStandardized to age and sex only

^dGeometric mean (95% CI)

^eLow-density lipoprotein (LDL) cholesterol was directly measured

*P-values test the null hypothesis that there is no trend across the genetic subgroups

4.3.6 Associations of interleukin-6 signalling with cardiovascular diseases

To investigate the causal relevance of IL-6 signalling to cardiovascular disease, I first used IL-6R (rs7529229) and IL-6 (IL-6 gene score) instruments to investigate associations of IL-6 signalling with CVD phenotypes in East Asians using individual-level data from CKB. Second, I examined associations of the IL-6(R) instruments with CVD outcomes in Europeans using summary data from the largest available GWAS using the same IL-6R instrument (rs7529229) and similarly derived IL-6 (rs1880241) instruments. The effects of the IL-6(R) instruments in Europeans and East Asians were orientated and scaled to correspond to a halving of genetically predicted CRP (Section 4.3.3).

Associations of IL-6R instrument with CVD outcomes

As shown in Figure 4.6, in CKB, the IL-6R instrument was associated with a reduced risk of both CHD and the more specific and better-defined MI endpoint. For equivalent differences in genetically predicted CRP concentrations, the effects in East Asians and Europeans were similar. On trans-ethnic meta-analysis of estimates, associations of the IL-6R instrument were similar between MI (OR [95% CI]: 0.67 [0.58, 0.77], $p = 2.6 \times 10^{-8}$) and CHD (0.74 [0.66, 0.82], $p = 2.5 \times 10^{-8}$). The IL-6R instrument was also associated with a reduced risk of carotid plaque in a subset of participants in CKB who had sonographic measurements of carotid atherosclerosis. At the time of submission, summary statistics for GWAS of carotid plaque in Europeans³⁷⁴ were not publicly available.

I also examined associations of the IL-6R instrument with two pathophysiologically distinct stroke types (as done in Chapter 3 for observational associations of plasma CRP and fibrinogen concentrations) - IS and ICH. The IL-6R instrument was associated with a lower risk of IS (OR_{meta} across CKB and Europeans [95% CI]: 0.89 [0.81, 0.99]) with no evidence of heterogeneity observed between CKB and European effect estimates ($P_{het} = 0.13$). With ICH, as for IS, there were consistent associations of the IL-6R instrument between CKB and Europeans. When the IL-6R

instrument - ICH associations were examined separately in CKB and Europeans, effect estimates in CKB (OR_{CKB} [95% CI]: 1.20 [0.98, 1.47]) and Europeans (OR_{EUR} [95% CI]: 1.64 [0.74, 3.66]) had CIs that contained the null value. The CIs in Europeans were wider than CIs in CKB, owing to the fact that CKB had almost twice the number of ICH cases than MEGASTROKE (nCases: 3226 Europeans vs. 6757 East Asians). However, there was no evidence of heterogeneity of effect estimates between CKB and Europeans ($P_{het} = 0.46$) and, on meta-analysis, the IL-6R instrument was, by contrast with the effect on IS, associated with a higher risk of ICH (OR_{meta} [95% CI]: 1.22 [1.01, 1.49]).

Associations of IL-6 instrument with CVD outcomes

In contrast to the IL-6R instrument, largely null associations were identified with CHD phenotypes (i.e. CHD, MI) using the IL-6 instrument, individually in both CKB and Europeans and also on trans-ethnic meta-analyses (Figure 4.6) with no evidence of heterogeneity between CKB and Europeans (P_{het} in CHD = 0.44; P_{het} in MI = 0.31).

The discrepancy between the IL-6(R) instruments with respect to their associations with CHD phenotypes in CKB may be due to several factors. Firstly, it may be due to a lack of sufficient power of the IL-6 instrument. In CKB, I had more than 80% power to detect association of IL-6 instrument with CHD and MI at an OR of 0.90 (Appendix Chapter A.3); so if IL-6 instrument was truly associated with a lower risk of CHD or MI, it is likely that this relationship is weaker than that observed for the association CHD/MI with IL-6R and may therefore be of limited clinical relevance. Second, this null association may represent a real discordant receptor-ligand relationship with respect to CHD/MI. This is supported by the replication of the null association of the IL-6 instrument with CHD/MI in Europeans (which had larger numbers of CHD/MI cases than CKB). These results suggest independent roles of the IL-6 receptor and ligand with respect to their contributions to CHD/MI risk. Additionally, histopathological studies showed that similar discordant IL-6 receptor-ligand relationships exist for other phenotypes such as prognosis for ovarian

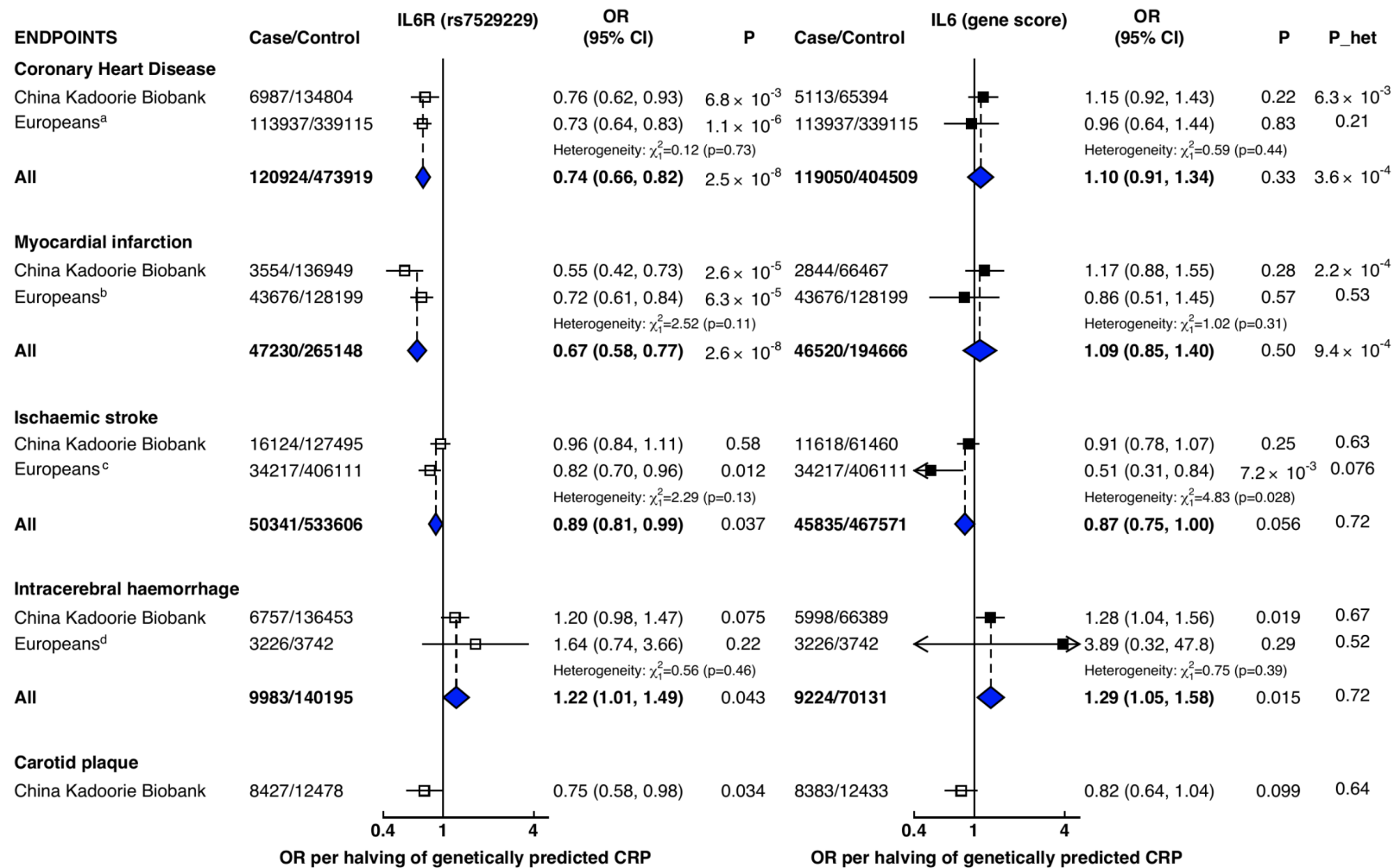
cancer^{375,376}. Overall, these observations suggest that the effects of drugs that target IL-6R may not be equivalent to the blockade of the IL-6 ligand directly, especially in reducing the risk of CHD or MI.

With IS, the association of the IL-6 instrument with IS was directionally consistent to IS associations of the IL-6R instrument, but the CIs for the point estimate for the IL-6 instrument included the null (OR_{meta} [95% CI]: 0.87 [0.75, 1.00]). When the IS association of IL-6 instrument was examined separately in CKB and Europeans, the CI of effect estimates in Europeans did not contain the null value (OR_{EUR} [95% CI]: 0.51 [0.31, 0.84]), suggesting the null association in CKB (OR_{CKB} [95% CI]: 0.91 [0.78, 1.07]) may be due to lack of power (nCases: 34,217 Europeans vs. 11,618 East Asians), although the CKB study had 80% power to detect an OR as low as 0.95 for IS in CKB (Appendix Chapter A.3). Alternatively, the null association of IL-6 with IS observed in CKB may represent the real association as opposed to the reduced risk of IS observed in Europeans. This suggestion is supported by better characterization of IS in CKB (90% of stroke cases were diagnosed for stroke type with an expert examination of case notes, including radiologist's reports on brain imaging data⁴¹) compared to MEGASTROKE (data source for European IS cases/controls) aggregated cases from multiple studies, and it was unclear to what extent a similar quality control protocol was applied for adjudication of IS in Europeans. A third possibility is related to differences in the spectrum of IS subtypes between CKB and MEGASTROKE that could potentially explain the heterogeneous effect estimates between CKB and MEGASTROKE. At the time of writing this chapter and performing analyses, I was not able to access CKB IS subtype data and hence unable to compare the composition of IS subtypes between CKB with MEGASTROKE to verify differences in case spectrum.

With the IL-6 instrument, I observed a higher risk of ICH in CKB (OR_{CKB} [95% CI]: 1.28 [1.04, 1.56]). The corresponding estimate in the Europeans was directionally consistent but overlapped the null with wide CIs (OR_{EUR} [95% CI]: 3.89 [0.32, 47.8]). The wide CIs in Europeans vs. CKB may be due to low statistical power

given the relatively lower number of Europeans with ICH examined vs. numbers of CKB participants as stated previously. Additionally, there was no evidence of heterogeneity of estimates between CKB and Europeans for IL-6 instrument - ICH associations ($P_{het} = 0.39$) and trans-ethnic meta-analysis was consistent with estimates observed in CKB (OR_{meta} [95% CI]: 1.29 [1.05, 1.58]).

The IL-6 instrument in CKB was composed of 3 SNPs and it is possible that pleiotropic effects across the 3 SNPs on CVD phenotypes can potentially dilute associations of the IL-6 instruments with CVD risk. To check for heterogeneity, I investigated vascular associations with the individual IL-6 SNPs (Appendix Figure A.5), including using rs13306436 (the IL-6 SNP with regulatory function) as the IL-6 genetic instrument, and observed a similar pattern of cardiovascular associations as the IL-6 instrument (Appendix Figure A.6). A similar pattern of associations suggests that there are no single SNPs in the IL-6 gene score that would differentially affect CVD association magnitude or direction.



^aCARDIoGRAMplusC4D [292]

^bCARDIoGRAMplusC4D [297]

^cMEGASTROKE [291]

^dInternational Stroke Genetics Consortium [294]

Figure 4.6: Trans-ethnic association of the IL-6(R) instruments with cardiovascular disease traits and outcomes.

The figure illustrates trans-ethnic meta-analysis (where possible) of associations of the IL-6(R) instruments with CVD outcomes. Models for MR analysis in CKB adjusted for age, age², sex and principal components of genetic ancestry in each of the 10 regions, and meta-analysed across the regions using fixed-effect method with inverse variance weighting. Estimates for Europeans used summary statistics from the largest genome-wide association studies of the respective diseases. Error bars represent 95% CI.

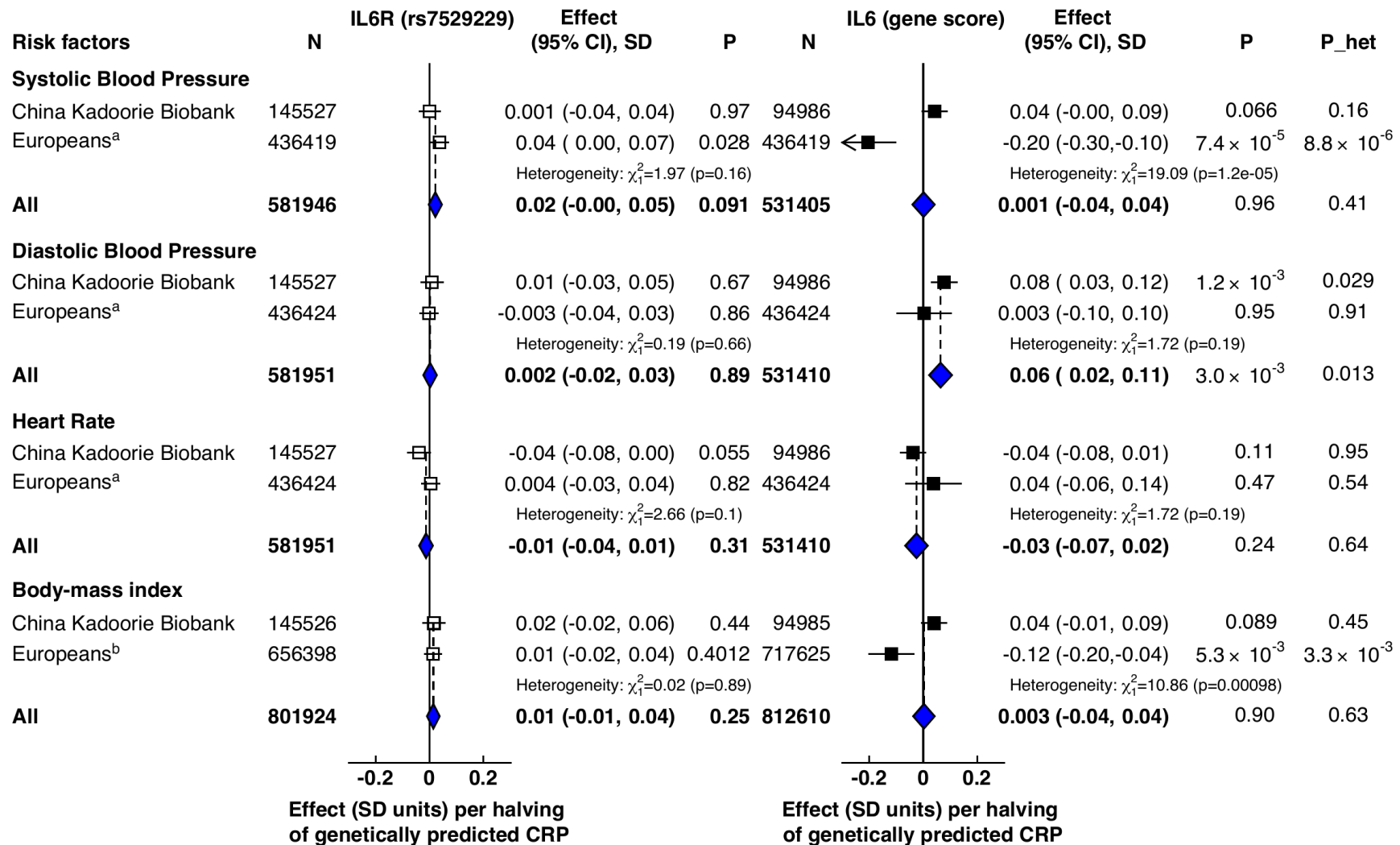
4.3.7 Associations of interleukin-6 signalling with CVD risk factors

Several risk factors are associated with CVD risk (Box 1.1). In the previous section, I have shown that the IL-6R instrument consistently associated with CVD risk in both East Asians and Europeans. However, whether this risk is potentially mediated by any known CVD risk factors is unclear.

In this section, I investigated associations of the IL-6(R) instruments with selected CVD risk factors in CKB. The selected CVD risk factors comprise SBP, DBP, BMI, physical activity, and heart rate. The selection of CVD risk factors was based on the evidence of association of these risk factors with CRP (Figure 3.7), prior evidence of association with CVD risk (Chapter 1.3), and what was available to investigate in CKB. Because of the relatively larger number of lipids in CKB, the association of the instruments with lipids is presented separately in the next section. As before, associations of the IL-6(R) instruments with CVD risk factors were investigated in Europeans to assess trans-ethnic consistency of direction and estimates as well as to get a better-powered analysis through meta-analysis.

With the IL-6R instrument, there were largely null associations of IL-6R with SBP, DBP, BMI, HR, and physical activity in both East Asians and Europeans (Figure 4.7). However, with the IL-6 instrument, there was some evidence of association of IL-6 instrument with only DBP both in CKB and after meta-analysis ($Effect_{meta}$ in SD [95% CI]: 0.06 [0.02, 0.11], $p = 3 \times 10^{-3}$, significant at 5% FDR) with no evidence of heterogeneity between CKB and Europeans ($P_{het} = 0.19$). This association of the IL-6 instrument with DBP may potentially explain the higher risk of ICH observed with the IL-6 instrument (Figure 4.6); higher DBP is a risk factor for ICH⁴¹. However, as noted earlier, there was no evidence of association of IL-6R with DBP or SBP, suggesting the higher risk of ICH with IL-6R may be independent of blood pressure. Between CKB and Europeans, although many of the CVD risk factor associations of IL-6 were directionally consistent, some were directionally discrepant. The directionally discrepant traits include systolic blood pressure ($P_{het} = 1.2 \times$

10^{-5}) and body mass index ($P\text{-het} = 9.8 \times 10^{-4}$); the source of these trans-ethnic discrepancies is presently unclear.



^aUKB [293] via MRC-IEU (<http://gwas.mrcieu.ac.uk/>) - IL6 genetic variant used: rs1880241 [288]

^bGIANT + UKB [295]

Figure 4.7: Trans-ethnic association of the IL-6(R) instruments with selected cardiovascular disease risk factors.

The figure illustrates trans-ethnic meta-analysis of associations of the the IL-6(R) instruments with CVD traits. Models for MR analysis in CKB adjusted for age, age², sex and principal components of genetic ancestry in each of the 10 regions, and meta-analysed across the regions using fixed-effect method with inverse variance weighting. Error bars represent 95% CI.

4.3.8 Associations of interleukin-6 signalling with lipids

Differences between the IL-6(R) instruments for their association with CHD/MI (Figure 4.6) may be due to differences in associations of the IL-6(R) instruments with lipids. For example, LDL-C is a known causal risk factor for CVD⁵¹.

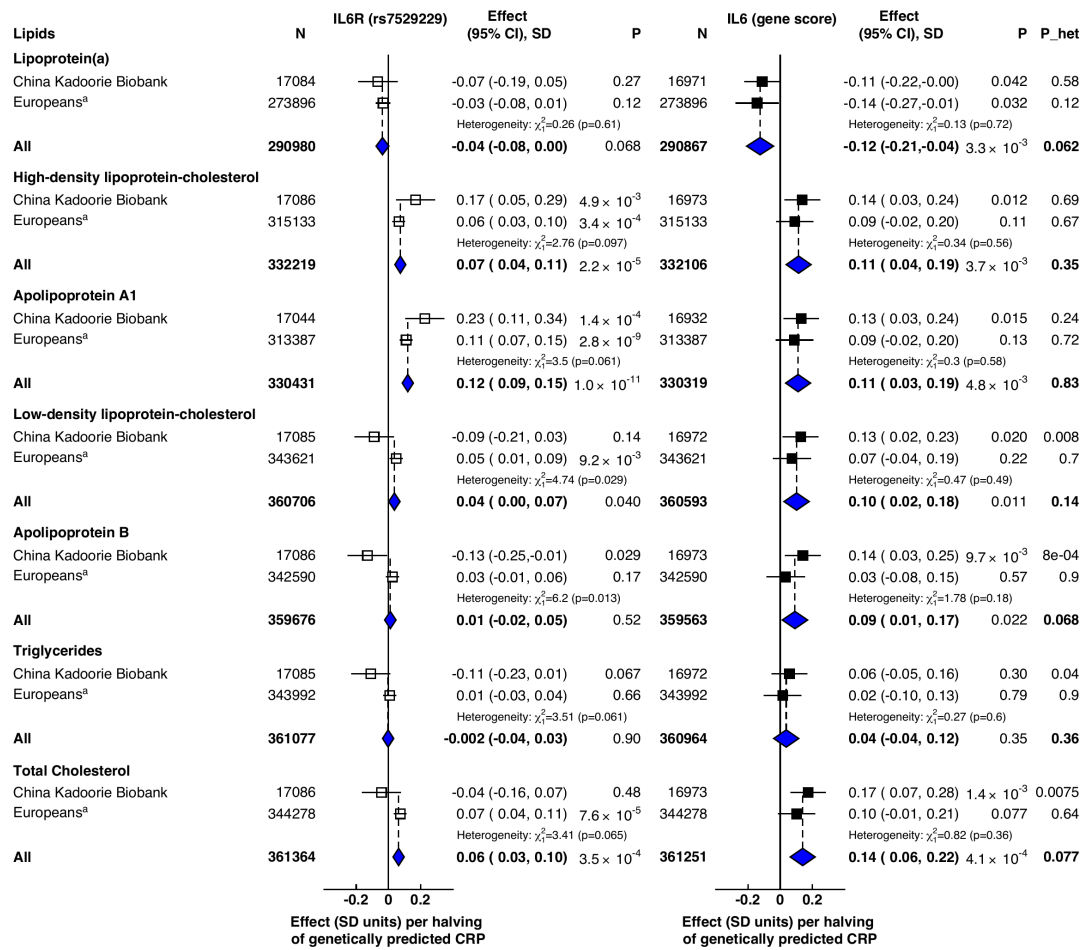
To examine for associations with lipids, I investigated associations of the IL-6(R) instruments in CKB and UKB with lipoprotein (a) (Lp(a)), high density lipoprotein-cholesterol (HDL-C) and its carrier protein apolipoprotein A1 (Apo-A1), low density lipoprotein-cholesterol (LDL-C) and its carrier protein apolipoprotein B (Apo-B), triglycerides (TG) and total cholesterol (TC) (Figure 4.8). The selection of these lipids was based on prior MR and observational studies that have examined the association of these lipids with CVD risk^{51,101} and what was available in CKB and UKB²⁹⁶.

In Europeans (UKB), associations with selected lipid traits between the IL-6(R) instruments were generally consistent (all $P_{het} > 0.05$). In CKB, only associations with Lp(a), HDL-C, and Apo-A1 were consistent between the IL-6(R) instruments; inconsistent associations between the IL-6(R) instruments were observed with LDL-C ($P_{het} = 0.008$), Apo-B ($P_{het} = 8.0 \times 10^{-4}$), TG ($P_{het} = 0.04$), and TC ($P_{het} = 0.0075$) in CKB and the source of inter-instrument discrepancy for these traits is presently unclear. Nevertheless, there is no evidence of heterogeneity between the IL-6(R) instruments when heterogeneity was assessed with the CKB-EUR meta-analysed estimates for all selected lipid traits. Similarly, when comparing estimates between UKB and CKB for each IL-6(R) instruments, there was no evidence of heterogeneity (at 5% FDR) between CKB and UKB for associations with all selected lipid traits (Figure 4.8).

On further assessment of CKB-EUR meta-analysed estimates at 5% FDR, three lipid traits (HDL-C, Apo-A1, and TC) exhibited significant associations with both the IL6(R) instruments; for equivalent differences in genetically predicted CRP, the IL-6(R) instruments were associated with higher HDL-C (IL-6R, $Effect_{meta}$ in

SD units [95% CI] per halving of genetically predicted CRP: 0.07 [0.04, 0.11], $P = 2.2 \times 10^{-5}$; IL-6, $Effect_{meta}$: 0.11 [0.04, 0.19], $P = 3.7 \times 10^{-3}$), Apo-A1, and TC (Figure 4.8). The meta-analysed association of only one lipid trait (Lp(a)) was significant using the IL-6 instrument ($Effect_{meta}$: -0.12 [-0.21, -0.04], $P = 3.3 \times 10^{-3}$) with estimates from the IL-6R instrument including the null, but no evidence of heterogeneity between the IL-6R and IL-6 estimates ($P_{het} = 0.06$).

To understand if the lipid associations observed with the IL-6(R) instruments were driven by the association of the instruments with specific lipid metabolites, I investigated the association of the IL-6(R) instruments in a small subset of CKB study participants (part of a nested CVD case-control study) with lipidomic measurements (Appendix Figure A.7). These measurements were carried out using a high-throughput proton nuclear magnetic resonance (1H NMR) spectroscopy platform (Chapter 2.2.4). Associations of the IL-6(R) instruments with lipidomic traits were largely null at 5% FDR; the lack of associations is potentially due to low statistical power owing to the small sample size ($n = \sim 4500$) for these lipidomic trait measurements.



^aUK Biobank [296] - IL6 genetic variant used: rs1880241 [288]

Figure 4.8: Trans-ethnic association of the IL-6(R) instruments with lipids. The figure illustrates associations of the IL-6(R) instruments with lipid traits in East Asians (CKB). Models for MR analysis in CKB adjusted for age, age², sex and principal components of genetic ancestry in each of the 10 regions, and meta-analysed across the regions using the fixed-effect method with inverse variance weighting. Error bars represent 95% CI. Estimates for Europeans used summary statistics from the UK Biobank for the respective lipid traits. Area of boxes are not proportional to the inverse of the variance of effect

4.3.9 Pre-specified subgroup analysis of the IL-6(R) instruments with CVD

Associations of the IL-6(R) instruments observed in Figure 4.6 may be subject to potential effect modification by other factors, potentially confounding or biasing the observed associations. To identify whether the IL-6(R) instrument associations in CKB differ by specific subgroups, I examined their associations with CVD endpoints in 10 pre-specified subgroups of 10 potential confounding variables (Figure 4.9 and Figure 4.10). The methods used for this subgroup analysis were similar to the subgroup analysis in Chapter 3.3.4: associations in each stratum of each subgroup trait were stratified by 10 regions in CKB and meta-analysed using fixed-effect inverse variance method (Chapter 4.2.3). As with previous analyses, I used a 5% FDR threshold rather than the conservative Bonferroni adjustment to assess for significant trends or heterogeneity (where appropriate).

Using the IL-6R instrument, there was no evidence of effect modification by any of the 10 subgrouped traits examined for AMI, IS or ICH. Although there was a nominally significant trend observed for IS with SBP (i.e. risk of IS was observed to be lowest in those with $SBP \geq 150$ mmHg, $P\text{-trend} = 0.013$), this was not significant at 5% FDR (Figure 4.9). Thus associations of the IL-6R instrument with CVD were not subject to appreciable effect modification by any subgroup strata.

Similar to subgroup analysis using the IL-6R instrument, I observed no evidence of effect modification using the IL-6 instrument, across categories of any of the 10 subgrouped traits for any of AMI, IS or ICH outcomes (Figure 4.10). Thus the null associations of IL-6 with CVD were not subject to appreciable effect modification by any subgroup strata.

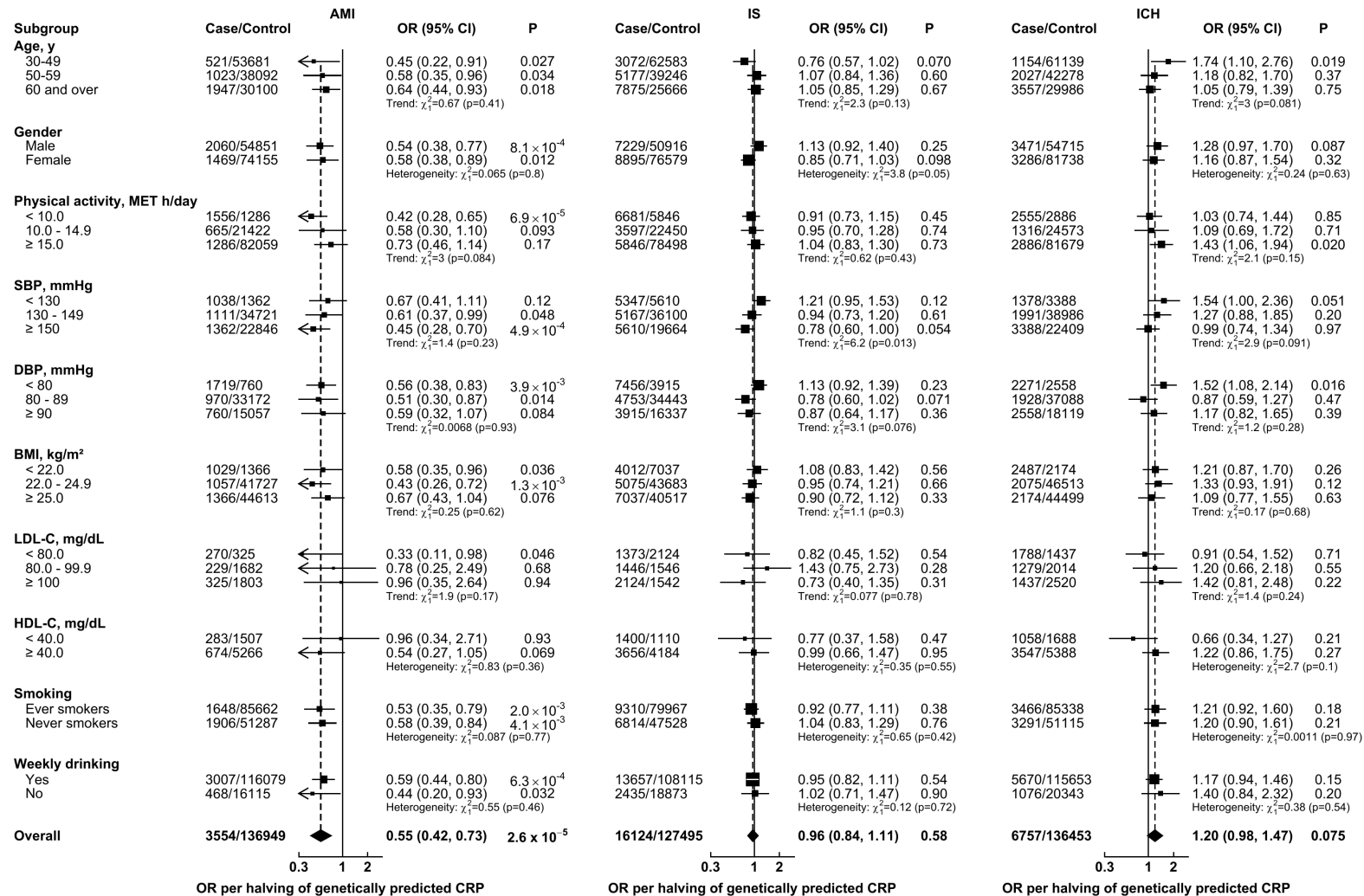


Figure 4.9: Subgroup analysis of the IL-6R instrument with cardiovascular disease outcomes.

The figure illustrates associations of the IL-6R instrument with pre-specified subgroups in East Asians (CKB). Models for MR analysis in CKB adjusted for age, age², sex and principal components of genetic ancestry in each of the 10 regions in each subgroup strata, and meta-analysed across the regions using fixed-effect method with inverse variance weighting. Error bars represent 95% CI. Area of boxes are proportional to inverse of the variance of effect

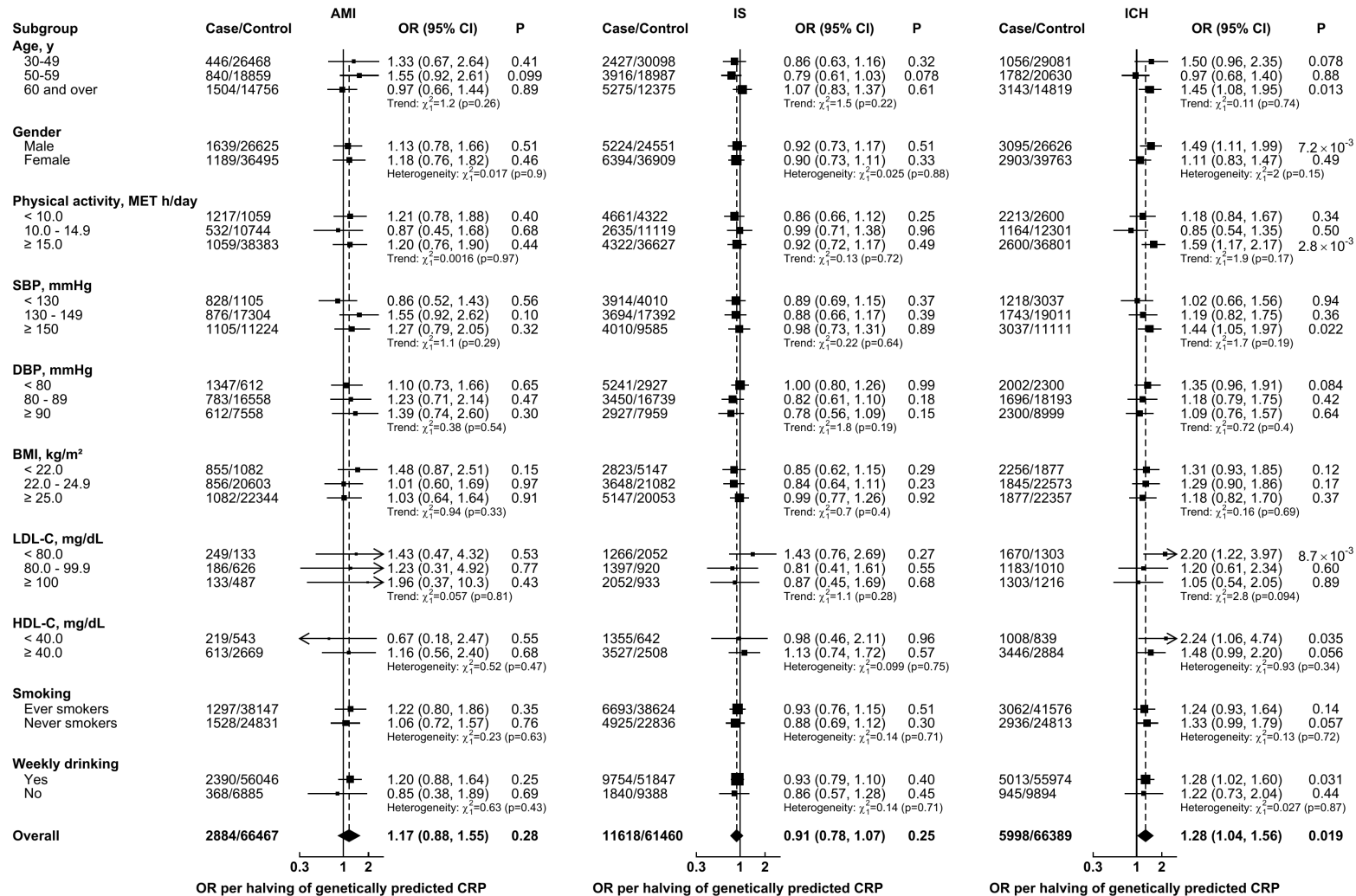


Figure 4.10: Subgroup analysis of the IL-6 instrument with cardiovascular disease outcomes.

The figure illustrates associations of the IL-6 instrument with pre-specified subgroups in East Asians (CKB). Models for MR analysis in CKB adjusted for age, age², sex and principal components of genetic ancestry in each of the 10 regions in each subgroup strata, and meta-analysed across the regions using the fixed-effect method with inverse variance weighting. Error bars represent 95% CI. Area of boxes are proportional to inverse of the variance of effect

4.4 Discussion

The present study investigated on-target associations of the IL-6(R) instruments, for equivalent differences in CRP (a marker of IL-6 signalling), with risk of CVD outcomes and CVD risk factors in CKB and Europeans. I found that while the IL-6R instrument was associated with a lower risk of CHD, IS, and carotid plaque, only null associations were observed with the IL-6 instrument (constructed to be a better instrument of IL-6 signalling inhibition than the IL-6R instrument) with the same CVD phenotypes (except IS). Furthermore, both IL-6(R) instruments were associated with a higher risk of ICH.

4.4.1 Genetic instruments representing interleukin-6 signalling

As discussed previously, a major reason for drug-target MR studies is to examine putative therapeutic or adverse effects of drugs using genetic instruments that can reasonably represent the effects of drugs in humans. Drugs that interfere with IL-6 signalling pathway comprise drugs that block the IL-6 receptor (both mIL-6R and sIL-6R, e.g. tocilizumab) or bind to the IL-6 ligand (e.g. siltuximab) and prevent binding of the IL-6 ligand to IL-6R. To represent the effects of drugs that act on the IL-6 signalling pathway, IL-6 drug-target MR studies to date used the non-synonymous genetic variant rs2228145-C to represent effects of IL-6R blockers such as tocilizumab. However, IL-6R blockers block both mIL-6R and sIL-6R while the rs2228145-C simply alters the balance between mIL-6R and sIL-6R (reducing mIL-6R and increasing sIL-6R via increased proteolytic cleavage of mIL-6R); although many of the downstream effects of both rs2228145-C and IL-6R blockers are consistent with each other (e.g. lower CRP concentrations, higher IL-6 protein concentrations, among many others discussed in Swerdlow et al⁹⁷), the alteration of the balance of mIL-6R and sIL-6R associated with rs2228145-C limits comparisons with drugs like tocilizumab that blocks both forms of IL-6R. Indeed, the higher sIL-6R concentrations should increase IL-6 trans-signalling (the two forms of IL-6 signalling are described in Chapter 1.5.5). Interleukin-6 trans-signalling is reported

to contribute to the pathogenesis and progression of chronic inflammatory diseases (e.g. rheumatoid arthritis³⁷⁷). These chronic diseases are typically characterised by higher CRP concentrations¹⁹¹. But, paradoxically, rs2228145-C associated with a physiological doubling of sIL-6R²⁸¹ (and associated with lower risk of the chronic inflammatory diseases³⁷⁷) is associated with lower CRP concentrations^{96,97}. Garbers et al offers two explanations for this seemingly paradoxical association of rs2228145-C and lower CRP levels²⁸¹. Firstly, experimental evidence suggested higher levels of sIL-6R potentiates the activity of soluble glycoprotein 130 (sgp130)²⁷⁷; sgp130 binds to the IL-6:sIL-6R complex and prevents it from binding to cells that express gp130 receptor (which is present in almost all cells in humans). This buffering effect by sgp130 makes it less likely for IL-6 signalling to promote the hepatic synthesis of CRP. Furthermore, sgp130 is at higher concentrations in plasma than sIL-6R¹⁹¹ in physiological states, and the higher sIL-6R is speculated to increase the buffer capacity of sgp130²⁸¹. A second explanation is that the rs2228145-C allele, via increased proteolytic cleavage, lowers the expression of mIL-6R on the surface of liver cells, which can reduce IL-6 signalling in the liver and lower amounts of CRP. Nevertheless, it remains challenging to use just the rs2228145-C variant (IL-6R instrument) to distinguish whether associations of altered IL-6R signalling observed in the present study is due to lower mIL-6R signalling (classic signalling) or higher sIL-6R signalling (trans-signalling).

To partially address this limitation, the present study also constructed a genetic instrument that represents altered IL-6 protein signalling, as the IL-6 instrument. However, the IL-6 instrument oriented to genetically predicted lower CRP concentrations (as an indicator of altered IL-6 signalling) would represent, to an extent, effects equivalent to lowering of both classic and trans-signalling, and a robust assessment ideally requires separate analysis of classic and trans-signalling pathways and lies beyond the scope of the present thesis.

In the present study, both IL-6(R) instruments were associated with CRP (this being the basis for their construction) and IL-6 protein. With effects scaled to

halving of genetically predicted CRP, the IL-6R instrument was associated with higher IL-6 protein concentrations (Effect in SD units [95% CI]: 0.36 [0.02, 0.69], $p = 0.036$). This was directionally consistent with observations in previous MR studies that used the same IL-6R instrument (Effect [95% CI] in pg/ml of IL-6 concentrations per minor allele of rs7529229: 0.09 [0.08, 0.10], $p = 8.41 \times 10^{-68}$ ⁹⁷, Effect [95% CI] in percentage mean difference of IL-6 concentrations per minor allele of rs7529229: 14.6% [10.7, 18.4], $p = 8.3 \times 10^{-14}$ ⁹⁶). The higher IL-6 protein concentrations observed with rs7529229-C (IL-6R instrument) is consistent with effects of tocilizumab observed in RCTs (from comparisons of tocilizumab and RCTs as documented by Swerdlow et al⁹⁷), i.e. tocilizumab administration (that blocks both mIL-6R and sIL-6R) also results in higher IL-6 protein concentrations.

There are three possible explanations of why tocilizumab and the IL-6R instrument result in higher IL-6 protein concentrations. One hypothesis suggests that the higher sIL-6R levels (observed with both IL-6R instrument and tocilizumab therapy⁹⁷) may result in sIL-6R:IL-6 complexes that may prolong the half-life of IL-6 protein³⁷⁸ (i.e. the IL-6 protein detected by assays may not be free IL-6 but IL-6 complexed with sIL-6R). However, while this may be the case for the IL-6R instrument, it does not explain IL-6 elevations in which tocilizumab competes with free IL-6 to form sIL-6R-tocilizumab immune complexes²⁸⁴. An alternative hypothesis suggests that higher IL-6 protein concentrations observed with blockade of IL-6R by tocilizumab or the IL-6R instrument may be due to the disruption of a negative feedback loop⁹⁷ that would typically occur after IL-6 signalling has occurred and reduced levels of IL-6 in an autocrine or paracrine manner (similar to the regulation of hormones³⁷⁹). A third possibility is that receptor-mediated IL-6 clearance is a major IL-6 clearance pathway and tocilizumab blockade of IL-6 receptors results in delayed clearance and elevations of free IL-6 levels²⁸⁴. While there is limited experimental evidence to support any of these hypotheses, it suggests the likely biological mechanism of higher plasma concentrations of IL-6 protein with tocilizumab may not be the same as the biological mechanism of observed higher plasma concentrations of higher IL-6 with the IL-6R genetic instrument. Genetic instruments representing altered

IL-6 protein-mediated signalling could, therefore, aid assessment of consistency, if any, of associations with altered IL-6R signalling that is associated with higher IL-6 protein concentrations, with the two genetic instruments orientated to genetically predicted lower CRP concentrations.

To assess altered IL-6 signalling, a weighted IL-6 gene score was constructed using three SNPs in the *IL6* locus that were independently associated with CRP in CKB. The alleles of the SNPs were orientated to represent lower CRP concentrations as a marker of altered IL-6 signalling. The IL-6 instrument was validated by assessing its association with IL-6 protein (similar to the assessment of the IL-6R instrument) in a smaller but independent subset of individuals in CKB with both genetic and proteomic data. The equivalent IL-6 instrument in Europeans was the SNP rs1880241, selected based on its association with CRP in *IL6* locus in Europeans²⁸⁸ and linked to IL-6 gene expression³⁸⁰. The lack of studies using similar or equivalent IL-6 genetic instruments limits comparisons of estimates of the association of IL-6 genetic instruments constructed for the present study.

4.4.2 Interleukin-6 signalling and coronary heart disease

The present study showed associations with CHD outcomes that were directionally discrepant between IL-6(R) instruments but directionally consistent for both instruments between East Asians and Europeans. The directions of association of IL-6R instrument with CHD were consistent with previous MR studies using the same IL-6R instrument^{96,97,285}). Only one of the three studies²⁸⁵ provided separate estimates for MI (OR [95% CI] per rs2228145-C: 0.94 [0.92, 0.97] which was also directionally consistent with meta-analysed estimates of IL-6R instrument with MI in the present study (OR [95% CI] per halving of genetically CRP concentration: 0.67 [0.58, 0.77])). The stronger association of the IL-6R instrument with MI compared to the association of IL-6R instrument with CHD in the present study may be due to the less likelihood of outcome misclassification for MI (which is a more specific CVD phenotype with clear definitions) compared to CHD. By contrast, using the IL-6 instrument, for equivalent differences in CRP concentrations as

the IL-6R instrument, was not associated with risk of CHD or MI in either East Asians or Europeans. As before, the lack of studies using similar or equivalent IL-6 instruments prevents comparisons of effect estimates in the present study.

The precise reasons for the IL-6 receptor-ligand inconsistency concerning CHD/MI outcomes are unclear. Nevertheless, the null association of the IL-6 instrument with CHD suggests two follow-up actions. First, the need for further pre-clinical research to investigate the purported cardioprotective benefits of targeting systemic IL-6 signalling. Second, pending sufficient in-depth investigations, to reconsider large investments in clinical trials testing IL-6 blockers to treat CHD³⁸¹. However, the present study is a proof-of-concept demonstrating the need for dual receptor-ligand MR studies to clarify the effects of drugs blocking receptors of a ligand that are not necessarily equivalent to the effects of blocking the ligand directly. The discrepant receptor-ligand associations speak to the complexity of biological pathways where receptors and their ligands may have biological roles independent of each other, as is the case for many orphan G-protein-coupled receptors³⁸².

4.4.3 Interleukin-6 signalling and stroke types

In contrast to the discordant associations of the IL-6(R) instruments with CHD/MI, the association with stroke types - IS and ICH - were consistent with respect to direction and effect size between instruments - IL-6R and IL-6, and between ethnic groups - East Asians and Europeans. On trans-ethnic meta-analysis, the associations of IL-6R instrument with IS in the present study (OR [95% CI] per halving of genetically CRP concentration: 0.89 [0.81, 0.99]) was directionally consistent with the most recent IL-6R signalling MR study in Europeans (OR per rs2228145-C for any stroke = 0.98 [95% CI: 0.97,0.99])²⁸⁶. This consistency is, in part, due to Rosa et al²⁸⁶ using the same stroke dataset (MEGASTROKE²⁹¹) as the present study. It should be noted, however, that IS is a composite phenotype, composed of five IS subtypes (based on TOAST classification scheme³⁸³) with three major subtypes - large artery, small vessel, and cardio-embolic subtypes of IS. At the time of analyses, I was not able to access data on specific IS subtypes in CKB, and

therefore could not assess associations of IL-6R with IS stroke subtypes in East Asians. However, in Europeans, Rosa et al found no difference (in direction or magnitude) between associations of the IL-6R instrument with IS and any of its three major subtypes²⁸⁶.

In addition to IS, the present study also examined associations of the IL-6(R) instruments with ICH in East Asians and Europeans. With the IL-6R instrument, there was no evidence of association with ICH in either East Asians or Europeans. However, when trans-ethnic estimates (which were directionally consistent, $P_{het} = 0.46$) were meta-analysed, there was a nominally significant association of the IL-6R instrument with a higher risk of ICH (OR [95% CI] per halving of genetically CRP concentration: 1.22 [1.01, 1.49]). With the IL-6 instrument, there was weak evidence of a similar association (i.e. similar to IL-6R instrument) with ICH in East Asians ($p = 0.02$), but a similar association was not observed in Europeans ($p = 0.29$). However, as with the IL-6R instrument, the meta-analysis showed there was a nominally significant association of the IL-6 instrument with ICH (mainly driven by the association in East Asians). Although weak, the association, of the IL-6(R) instruments with ICH is novel and had not been reported in prior IL-6R MR studies^{97,286,377,384,385}, presumably due to the lower number of ICH cases.

The mechanism linking lower systemic inflammation (as represented by the IL-6(R) instruments) to a higher risk of ICH is unclear at present. It is possible that lower systemic inflammation that is observed to be associated with reduced risks of CHD and IS may be associated via known CVD risk factors such as blood pressure (discussed in the next section) or alterations of components of the coagulation cascade that promotes an anti-thrombotic state³⁸⁶ which can potentially result in the higher risks of ICH. This anti-thrombotic state as a result of lower systemic inflammation is supported by the association of the IL-6R instrument with a reduced risk of carotid plaque in CKB (OR [95% CI] per halving of genetically CRP concentration: 0.75 [0.58, 0.98]), as carotid plaque build-up is a marker of systemic atherothrombosis³⁸⁷. The association of IL-6R instrument with risk of

carotid plaque in East Asians is directionally consistent with what's observed in the Million Veterans Study using the same IL-6R instrument with atherosclerosis (OR [95% CI] per rs2228145-C: 0.93 [0.91-0.96])²⁸⁵.

4.4.4 Interleukin-6 signalling and cardiovascular disease risk factors

Trans-ethnic assessments of the IL-6(R) instruments with well-established CVD risk factors showed largely consistent associations in the present study between East Asians and Europeans, except for associations with systolic blood pressure (SBP) ($P_{het} = 1.2 \times 10^{-5}$) and body-mass index (BMI) ($P_{het} = 9.8 \times 10^{-4}$) with the IL-6 instrument. Although the associations of SBP and BMI were heterogeneous between East Asians and Europeans with the IL-6 instrument, the associations were null on trans-ethnic meta-analysis. The reasons for this trans-ethnic discrepancy in these two cases (SBP, BMI) with IL-6 instruments is unclear. Some of these trans-ethnic discrepancies may be related to the pleiotropic effects of different SNPs that make up the IL-6 gene score in CKB compared to the single-SNP IL-6 instrument in Europeans. Alternatively, the discrepant trans-ethnic estimates for SBP and BMI with the IL-6 instrument may represent environmental or familial influences that can only be resolved using alternative study designs such as within-family Mendelian randomisation analyses³⁸⁸.

With IL-6R instrument, on trans-ethnic meta-analysis, there were largely null associations of IL-6R instrument with the selected CVD risk factors, including blood pressure (SBP/DBP, SD units [95% CI] per halving of genetically CRP concentration: 0.02 [-0.001, 0.05]/0.002 [-0.02, 0.03]) and BMI phenotypes (SD units [95% CI] per halving of genetically CRP concentration: 0.01 [-0.01, 0.04]). The null associations of IL-6R instrument with SBP and DBP in the present study are not in keeping with reported associations from previous studies which found lower SBP (Percentage mean difference per rs2228145-C [95% CI]: 0.1% [-0.3, -0.01]⁹⁶; mmHg per rs2228145-C: 0.21 mm Hg [-0.37, -0.05]⁹⁷) and lower DBP (mmHg per rs2228145-C: 0.11 mm Hg [-0.20, -0.02]⁹⁷) with the IL-6R instrument. The reason

for the discrepancy between the present study and previous studies concerning the association of the IL-6R instrument with blood pressure phenotypes is unclear. However, the present study had a combined (East Asian + European) sample size ($n = \sim 500,000$) five times the previous studies and therefore can provide more precise estimates of the association. For BMI, a previous MR study⁹⁶ observed null associations with BMI (Percentage mean difference per rs2228145-C [95% CI]: 0.01% [-0.1, 0.2]) consistent with the present study.

With the IL-6 instrument, like the IL-6R instrument, most of the associations were null on meta-analysis. However, the IL-6 instrument was associated with higher DBP on meta-analysis (Estimate in SD units [95% CI] per halving of genetically CRP concentration: 0.06 [0.02, 0.11]). This can potentially be the mechanism that might explain the higher risk of ICH observed with the IL-6 instrument in CKB since DBP is a risk factor for ICH in CKB⁴¹. As previously, the lack of studies using equivalent IL-6 instruments to assess associations with DBP limits comparison with external studies.

4.4.5 Interleukin-6 signalling and lipids

The present study reported largely consistent associations of the IL-6(R) instruments with most of the available clinical biochemistry measures of lipid traits. In CKB, the IL-6 instrument was associated with higher Apo-A1, HDL-C, Apo-B, and LDL-C. The estimates were consistent with effect estimates in Europeans using the equivalent IL-6 instrument in Europeans. Using MR methods, studies have shown that both LDL-C and Apo-B are causal risk factors for coronary heart disease^{51,389} but there was no evidence of causal association of HDL-C with CVD⁵¹. As above, the lack of studies using IL-6 instruments limits the comparison of these effect estimates with previous studies.

On meta-analysis of estimates from CKB with estimates of association in the European population, the associations of the IL-6 instrument with lipid traits in the present study were directionally consistent with the IL-6R instrument. For both

instruments, null associations were reported for triglycerides. The IL-6R instrument was associated with higher concentrations of Apo-A1 and HDL-C in CKB with consistent estimates in the Europeans. Apo-A1 was not assessed with an IL-6R instrument in previous studies, but null associations of the same IL-6R instrument with HDL-C were reported in two previous MR studies (mmol/L [95% CI] per rs2228145-C: 0.002 [-0.001, 0.006]⁹⁷, Percentage mean difference per rs2228145-C [95% CI]: 0.07% [-0.1, 0.4]⁹⁶). As with blood pressure previously, the discrepancy between the present study and the previous MR study may be due to the large sample size used in the present study. For example, a more recent MR study²⁸⁶ that used a larger sample size of the European population from UK Biobank (approximately twice the sample size of the previous studies, $n = 255,929$) reported associations of their IL-6R gene score (represented by sIL-6R concentrations) with HDL-C concentrations (Effect per SD unit increase of genetically predicted sIL-6R [95% CI]: 0.005 [0.001, 0.010]) consistent with meta-analysed (CKB + UKB, $n = 332,219$) estimates in the present study (Effect in SD units [95% CI] per halving of genetically predicted CRP: 0.07 [0.04, 0.11]). Although HDL-C is unlikely to be causal for CHD⁵¹ and does not mediate the protective CHD effects of IL-6R²⁸⁶ despite previous observational studies that showed inverse associations of HDL-C with CVD⁵⁵, plasma HDL-C (like plasma CRP) is likely a marker of CVD risk than being causal for CVD.

Separate from Apo-A1/HDL, there were largely null associations of the IL-6R instrument with Apo-B and triglycerides. Similar to Apo-A1, Apo-B association with the IL-6R instrument was not assessed in previous studies. But the null association of IL-6R with triglycerides in the present study is consistent with the previous studies that reported similar null associations^{96,97,286}. With LDL-C, there was weak evidence of association of IL-6R instrument with LDL-C in the present study (meta-analysed effect in SD units [95% CI] per halving of genetically predicted CRP: 0.04 [0.001, 0.07]) that was not observed in three previous studies which reported null associations with LDL-C (mmol/L [95% CI] per rs2228145-C: -0.003 [-0.012, 0.005]⁹⁷, percentage mean difference per rs2228145-C [95% CI]: -0.03%

$[-0.3, 0.3]$ ⁹⁶, effect per SD unit increase of genetically predicted sIL-6R [95% CI]: 0.002, 95% CI $[-0.001, 0.005]$ ²⁸⁶). As above, larger sample size in the present study for assessment of LDL-C ($n = 360,706$) compared to the previous studies ($n = \sim 100,000$ to $\sim 280,000$) enabled detection of this small effect that previous studies were underpowered to detect.

The association of higher HDL-C and LDL-C with IL-6R instrument is also consistent with RCTs of tocilizumab which observed similar associations with higher HDL-C/LDL-C due to tocilizumab therapy⁹⁷. The association of IL-6R instrument with HDL-C is supported by an association of a similar magnitude with its carrier protein Apo-A1 in the present study; this was not the case for LDL-C as the association of IL-6R instrument with its carrier protein Apo-B was null in the present study. The association patterns are similar to observations in a clinical case-control study of SAMURAI trial, where rheumatoid arthritis patients, after 3 months of tocilizumab therapy, had elevations (from baseline levels) of HDL-C, LDL-C, and Apo-A1, but not Apo-B³⁹⁰. The inconsistency of HDL-C/LDL-C associations between IL-6R instrument and what was observed in tocilizumab RCTs led some investigators in a previous study⁹⁷ to suggest it may be an off-target effect of tocilizumab or a general sign of inflammation suppression in individuals with excessive systemic inflammation. However, the present study, having assessed the association of IL-6R instrument with HDL-C/LDL-C in a larger number of samples, suggests that elevations of HDL-C/LDL-C observed with tocilizumab therapy may be an on-target effect of IL-6R blockade.

4.4.6 Strength and limitations

The present study has several strengths. To understand the role of interleukin-6 signalling on CVD, I used the MR approach to investigate causal associations of IL-6 receptor and IL-6 ligand on CVD risk for equivalent differences in CRP. Mendelian randomisation studies, in general, have several strengths over observational studies (these are discussed in Chapter 1.5.4). For example, provided instrumental variable assumptions are met³⁹¹, MR studies are generally robust to biases of reverse

causality and confounding - two major limitations of conventional observational epidemiology.

An extension of MR, the drug-target MR approach, uses genetic variants near or in genes encoding proteins that are of interest as pharmaceutical targets. Drug-target MR leverages the randomisation of genetic variants at conception at the population level in a manner analogous to the randomisation of study participants in an orthodox randomised controlled trial (RCT) (Figure 1.5). This random allocation of genetic variants allows genetic variants to be used to represent life-long effects of therapeutic target modulation, without necessarily incurring the high cost of an RCT^{392,393}. It needs to be emphasized, however, that drug-target MR studies are not a substitute for RCTs; rather, inferences generated using this approach can be used to prioritize RCTs of drugs that show promise in ‘human models’ of disease and are more likely to be approved for the treatment of human diseases²⁴².

To this end, a major strength of the MR approach used in the present study is the specificity of genetic instruments used to represent the different components of IL-6 signalling. Both the IL-6(R) instruments were constructed based on their shared associations with CRP - a general marker of systemic inflammation - that is produced in large amounts from the liver on stimulation by IL-6 protein. The present study further confirmed the legitimacy of rs7529229 as the IL-6R instrument by using IL-6 protein concentrations (available for a subset of the Chinese population in CKB) as indicators of pharmacological interference in the IL-6R mediated signalling. For the IL-6 instrument, I used genetic variants in the *IL6* locus that were associated with CRP in both Europeans (rs1880241) and East Asians (IL-6 gene score in CKB) to create a genetic instrument for IL-6 protein, validated the IL-6 instrument in CKB by confirming its association with IL-6 protein, and investigated whether effects of IL-6R instrument are representative of effects of IL-6 instrument, for equivalent differences in CRP as the IL-6R instrument. It is assumed that with the high specificity of the IL-6(R) instruments to represent modulation of their respective targets, the possibility of a horizontal pleiotropic association, i.e. associations of

the genetic variant are via pathways other than IL-6R or IL-6 pharmacological modulation, would be minimal (but horizontal pleiotropy is still possible, see next paragraph on limitations of the present study). The trade-off, however, with the use of locus-specific IL-6(R) instruments is power, i.e. a polygenic instrument (e.g. an allele score or genetic risk score of CRP) could potentially explain more of the variation of the protein abundance (e.g. CRP) than the locus-specific IL-6(R) instrument. But, as described before, if the aim of my thesis was to simply infer the causal relationship between CRP as a biomarker of systemic inflammation and CVD, the polygenic approach to genetic instrument selection would be more appropriate than use of locus-specific instruments²⁶⁹. However, when the objective is understanding whether pharmacological modulation of a target alters the course of a disease, locus-specific instruments work better than polygenic instruments (where associations may be driven via other unknown pathways). For example, using the polygenic instrument approach it was shown that CRP was associated with CHD²⁸⁷, although CRP locus-specific instruments showed that there was no evidence that direct pharmacological modulation of CRP is likely to have any cardioprotective benefits⁹⁴. Nevertheless, it remains possible (and plausible) that pathways beyond IL-6 signalling may regulate levels of CRP; however, a search for additional inflammation-associated genetic instruments (e.g. *IL1 β* , *JAK1*) did not yield any additional signals (Appendix Figure A.3) limiting my analyses to use only the IL-6(R) instruments. An additional strength of the present study is the large number of vascular outcomes and risk factors in CKB; reliability of the genetic associations of vascular outcomes in CKB is further strengthened by the inclusion of European studies. The public availability of summary statistics in European studies enabled me to determine whether findings in CKB concerning CVD associations of the IL-6(R) instruments were consistent between ancestries using equivalent European genetic instruments. Trans-ethnic meta-analysis of estimates provided higher power to confirm causal effects of IL-6 signalling on cardiovascular disease outcomes. The dense phenotyping in CKB study participants also enabled investigations of known CVD risk factors, lipids, lipidomics and proteomics with

the IL-6(R) instruments to understand if there are additional epidemiological and biological associations of IL-6 signalling.

The present MR study also has several limitations. First, the construction of the IL-6 instrument used different genetic variants from the *IL6* locus to represent IL-6 protein activity in East Asians and Europeans that may not be equivalent. This is because the IL-6 SNP used to represent IL-6 protein activity in Europeans (rs1880241) is rare in East Asians (allele frequency of effect allele G is 0.01) and 2 of the 3 IL-6 SNPs (rs13306436, rs80085481) used to create the IL-6 gene score in East Asians are at low frequency in the Europeans (minor allele frequencies < 0.01). The IL-6 instrument in Europeans (rs1880241), in particular, was selected based on its association with CRP in *IL6* locus in Europeans²⁸⁸ but not validated (as was done for the IL-6 gene score in CKB) by confirming its association to levels of IL-6 protein due to lack of access to publicly available data of IL-6 protein GWAS in Europeans at the time of writing this thesis; hence, this instrument, although linked to IL-6 gene expression (Table A.1) and CRP levels, may not reliably represent differences in IL-6 protein concentrations in Europeans. This was not the case for the IL-6 gene score used in CKB, as the instrument demonstrated clear associations in CKB with both IL-6 protein (in a small subset of study participants); further, one of the SNPs in the IL-6 gene score - rs13306436 - is located in the three prime untranslated region (3'-UTR) in IL-6 mRNA transcript³⁹⁴; 3'-UTR generally post-transcriptionally regulate gene expression. Second, instruments for East Asians were selected based on CRP measurements that were assayed in a subset of participants selected for a nested case-control study of cardiovascular disease in CKB (i.e. a non-random subset of participants). Hence, SNPs selected from a GWAS of CRP in CKB to create the IL-6 gene score may have been associated with CRP due to the selection being conditioned on cardiovascular disease (i.e. collider-stratification bias). However, I assessed the association of the selected SNPs, the IL-6(R) instruments in a random subset of study participants and observed consistent associations (as in the non-random subset) with CRP and interleukin-6 protein. Ideally, SNPs for IL-6 instrument should be selected from a GWAS of IL-6 protein, but given the

limited sample size that has IL-6 proteomic data in CKB, I chose to select SNPs from near the *IL6* locus based on their association with CRP (which were assayed in a larger number of participants with genetic data). Although I validated our instrument by ascertaining the association of IL-6 gene score with IL-6 protein, there remains a possibility that it may not reflect the effects of IL-6 protein. To further confirm these are effects of genetically predicted IL-6 protein concentrations, I assessed similar associations with one of the SNPs that is part of the IL-6 gene score (rs13306436) which, as mentioned previously, is likely a regulatory variant affecting IL-6 gene expression - and observed consistent associations (i.e. as seen with IL-6 gene score). Third, the weights for IL-6 gene score in CKB were generated internally, i.e. from the same data that is also used to assess outcomes (or ‘data under analysis’) and this can bias causal estimates³⁶⁸. To mitigate this bias, I generated weights using a 100-fold cross-validation approach, where weights for each 1% of the sample were derived iteratively from 99% of the data. Fourth, the lack of adequate phenotyping for ischaemic stroke subtypes in CKB at the time of writing this thesis limits the investigation of the association of IL-6(R) instruments with the subtypes. Finally, a further limitation concerns the use of the IL-6R instrument to represent IL-6 signalling. As stated, the IL-6R instrument represents the alteration of the ratio of sIL-6R to mIL-6R. The lack of sIL-6R data in CKB prevents assessment of the IL-6R instrument with sIL-6R; nonetheless, previous studies have consistently shown that the IL-6R instrument (represented by rs2228145-C non-synonymous variant) is associated with higher plasma sIL-6R concentrations^{96,97,286} similar to what was observed in patients on tocilizumab therapy⁹⁷. However, while elevations of circulating sIL-6R in tocilizumab therapy are due to the formation of higher sIL-6R:tocilizumab complexes²⁸⁴, elevations of sIL-6R observed using the IL-6R instrument may be due to higher sIL-6R:IL-6 complexes³⁷⁸. The accuracy of the SOMAScan platform to detect proteins complexed with other proteins (or discerning free proteins from complexed proteins) is yet to be determined³⁹⁵. The higher sIL-6R:IL-6 complexes associated with the IL-6R instrument can bind to the gp130 receptor which is present in most human tissues

(trans-signalling) and induce a pro-inflammatory state³⁹⁶. This raises the question, is higher trans-signalling induced inflammation associated with lower CVD risk observed with the IL-6R instrument in the present and previous studies? As stated at the beginning of this section, one hypothesis postulated by Garbers et al²⁸¹ to explain why the higher sIL-6R:IL-6 complexes do not result in higher trans-signalling is the disproportionately higher levels of plasma sgp130 concentrations which is typically the case in normal physiological states (that form sgp130:sIL-6R:IL-6 complex and prevent binding to gp130). The buffering activity of sgp130 is overwhelmed in pathological states when excessive production of IL-6 and the resulting sIL-6R:IL-6 complexes end up binding to membrane-bound gp130 receptors to induce the trans-signalling led pro-inflammatory state. However, this hypothesis requires experimental verification. Although the use of the IL-6 instrument enabled some differentiation, as the IL-6 instrument was used in a way that represented lower IL-6 protein activity using CRP as a marker of protein activity, to reliably distinguish between classic and trans-signalling, future research will require the use of additional genetic variants, e.g. genetic variants that encode sgp130 - that is known to inhibit trans-signalling specifically²⁸² (but also inhibits classic signalling to an extent depending on the ratio of free IL-6 and sIL-6R³⁹⁷) or genetic variants that represent the ratio of sIL-6R:IL-6 complex to the sIL-6R:IL-6:sgp130 complex. Finally, although I stated the specificity of genetic instruments as a strength in this thesis, it remains possible that exposure- and disease-associated variants do not share the same causal signal. The specificity of instruments simply based on how variants are selected (i.e. cis-acting rather than genome-wide) does not necessarily preclude horizontal pleiotropy via genetic confounding. Genetic confounding occurs when an MR association is driven by exposure-associated variants in LD with disease-associated variants. In these cases, additional analysis like genetic co-localisation can estimate the posterior probability that both exposure and disease share the same causal variant²⁴³. Furthermore, only about 3 SNPs were obtained after fine-mapping and conditional analysis of the signal in CRP GWAS near IL6 to construct the IL-6 instrument, precluding additional sensitivity analyses like use of MR-Egger,

which typically requires a large number of genetic instruments³⁹⁸. Additionally, for locus-specific genetic instruments – it is possible that the genetic associations of the exposure to be correlated to the genetic associations of the outcomes via genetic confounding due to linkage disequilibrium, violating one of the assumptions of MR-Egger (InSIDE assumption – Instrument Strength is independent of direct effects)³⁹⁹. As before, genetic colocalisation tests can be used to evaluate whether there is genetic confounding prior to using MR-Egger, provided there are sufficient number of strong genetic instruments. Future studies should report, alongside MR results, results of genetic co-localisation and, if the number of genetic variants permit, sensitivity analyses (MR-Egger and/or MR weighted median) to check for violations of instrumental variable assumptions and provide additional evidence for (or against) the specificity of a genetic instrument.

4.5 Conclusion

The work in this chapter provides evidence that reproducibly supports the association of the IL-6R instrument with lower CHD risk but, at present, does not support the association of the IL-6 instrument with CHD risk. The discrepant associations between the IL-6(R) instruments suggests that the cardioprotective association of the IL-6R instrument is likely not via IL-6 signalling inhibition, and reinforces the need for further pre-clinical investigations prior to large-scale investments in randomised controlled trials testing the efficacy of IL-6 blockade against CHD.

5

Phenome-wide associations of interleukin-6 signalling

5.1 Introduction

In the last chapter, I investigated associations of the IL-6R and IL-6 instruments (IL-6(R) instruments) with the risk of CVD outcomes. The present chapter reports associations of IL-6(R) instruments with prespecified and non-prespecified phenome-wide disease outcomes and quantitative traits in CKB. On the assessment of associations of the IL-6(R) instruments with blood biomarkers, I compare, where possible, the consistency of associations of the genetic instruments with effects observed in tocilizumab (an IL-6R antagonist) trials for the same biomarkers.

This approach to investigating endpoints beyond hypothesised outcomes has increasingly become standard practice in drug-target mendelian randomisation (MR) studies^{97,285,286}. This is due to two reasons. First, it is important to assess whether genetic instruments that mimic the effect of a potential therapy are associated with the risk of unintended disease outcomes. The risk of unintended disease outcomes could represent potential adverse events of the therapy (i.e. on-target adverse events) under investigation. The need to assess outcomes beyond efficacy endpoints is driven

by the concern that potential adverse drug events may compromise the purported therapeutic benefits of modulating a drug target. Human genetics can therefore serve as an important source of evidence to assess this balance between potential therapeutic benefits and safety endpoints for a drug, using genetic instruments that represent the effect of modulating the target of the drug. Furthermore, a number of drug-target MR studies have retrospectively recapitulated adverse drug events that were observed in clinical trials. For example, although several MR studies and clinical trials of statins demonstrated a reduced risk of atherosclerotic CVD associated with lower LDL-C, some MR studies that examined associations of genetically predicted lower LDL-C concentrations flagged a potentially higher risk with diabetes^{400,401}. The higher risk of diabetes with lower LDL-C also corresponded with findings in the statin trials⁴⁰², suggesting that the risk of diabetes associated with statins may be an on-target effect. The utility of drug-target MR for investigation of adverse drug events, therefore, has implications for the design of future RCTs that would use the appropriate efficacy and safety assessment parameters to investigate drugs for specific indications. Drug-target MR studies should thus form an integral part of the design of RCTs to inform what safety endpoints matter. Second, an additional output from MR-PheWAS is the potential opportunity for drug repositioning. For example, a phenome-wide association study of IL-6 signalling using the IL-6R functional variant rs2228145 suggested a reduced risk of aortic aneurysm²⁸⁵. Patients with giant cell arteritis (GCA) have approximately a two-fold higher risk of aortic aneurysm⁴⁰³ and GCA is the most recent FDA-approved indication for tocilizumab⁴⁰⁴.

Interleukin-6 is a highly pleiotropic cytokine⁴⁰⁵, i.e. it has a diverse repertoire of functions in multiple organ systems. It was shown that drug side effects were statistically more likely to occur in organ systems where there is genetic evidence of a link between the drug target and a phenotype involving that organ system⁴⁰⁶. In addition to CVD (Chapter 4), previous studies have reported that IL-6 signalling plays roles in cancers e.g. digestive cancers⁴⁰⁷ (especially gastric⁴⁰⁸, oesophageal⁴⁰⁹ and colorectal carcinoma⁴¹⁰), hepatocellular carcinoma⁴¹¹,

lung cancer⁴¹², breast cancer⁴¹³), respiratory disease⁴¹⁴ (e.g. chronic obstructive pulmonary disease⁴¹⁵, pneumonia⁴¹⁶, asthma⁴¹⁷, resistance to tuberculosis⁴¹⁸), and additionally in diabetes⁴¹⁹, and depression⁴²⁰. However, in most of these studies, the evidence for IL-6 signalling was generated from in vitro experiments or observational studies. The present study examines the causal role of IL-6 signalling for the aforementioned phenotypes using human genetics data from a large cohort of East Asians part of the China Kadoorie Biobank.

Unless otherwise stated, I used the same IL-6(R) instruments as in the previous chapter (Chapter 4). Throughout this chapter, to facilitate comparisons of effects between instruments, effects for the IL-6(R) instruments are expressed per halving of genetically predicted CRP.

5.2 Methods

The methods used for the present chapter are identical to what was used in Chapter 4. Additionally, I have also described these methods in brief in the relevant sections in the present chapter before discussion of the results.

5.2.1 Selection of cases and controls

Prespecified non-vascular disease outcomes

Pre-specified non-vascular outcomes [ICD-10 codes] were selected based on a non-systematic search for papers that have examined associations of IL-6 signalling and non-vascular disease phenotypes, and what was available in CKB (with a sample size of at least 500 cases in the genotyped set). This included individuals with cancers, respiratory disease, and other outcomes with prior evidence of association with interleukin-6 signalling.

Cancer outcomes included digestive tract cancers⁴⁰⁷ (oesophageal [C15]⁴⁰⁹, stomach [C16]⁴⁰⁸, colorectal [C18-C20]⁴¹⁰), lung cancer [C33-34]⁴¹² and breast cancer [C50]⁴¹³; respiratory disease outcomes included chronic obstructive pulmonary disease [J41-J44]⁴¹⁵, pneumonia [J12-J18]⁴¹⁶, asthma [J45-J46]⁴¹⁷, and tuberculosis [A15-A19,

J65, K23.0, K67.3, M01.1, M49.0, M90.0, N33.0, N74.0, N74.1, O98.0, P37.0]⁴¹⁸. Additionally, I had assessed association of the instruments with incident type 1 or type 2 diabetes [E10-E14] or diabetes diagnosed at baseline (using random and fasting blood glucose screens)⁴¹⁹, and individuals reporting symptoms of major depressive disorder at baseline or diagnosed during follow-up [F32]⁴²⁰. Controls for pre-specified non-vascular outcomes excluded individuals who self-reported with the respective disease outcomes at baseline or diagnosed during follow-up.

Power estimates for different sample sizes are provided in Appendix table A.4; in brief, for endpoints with >1000 cases, the IL-6 instrument and IL-6R instrument had more than 70% power to detect an OR of 0.85 at an adjusted alpha of 0.05/10 (adjusted for the number of pre-specified endpoints).

Non-prespecified vascular and non-vascular outcomes

For the phenome-wide analysis, I initially selected 35 outcomes with at least 500 *IL6R* genotyped cases for MR-PheWAS. This yielded 35 outcomes for PheWAS analysis. At an alpha of 0.05/35 and 500 cases, the MR-PheWAS analyses had more than 50% power to detect an OR of 0.85 and more than 80% power to detect an OR of 0.80. Although power was lower, outcomes selected for *IL6R* PheWAS analysis using the sample size parameter (i.e. $n = 500$ *IL6R* genotyped cases) were also selected for *IL6* PheWAS, irrespective of the number of *IL6* genotyped cases available, to allow comparisons of the results for *IL6R*.

5.3 Results

5.3.1 Associations of the IL-6(R) instruments with prespecified non-vascular diseases in CKB

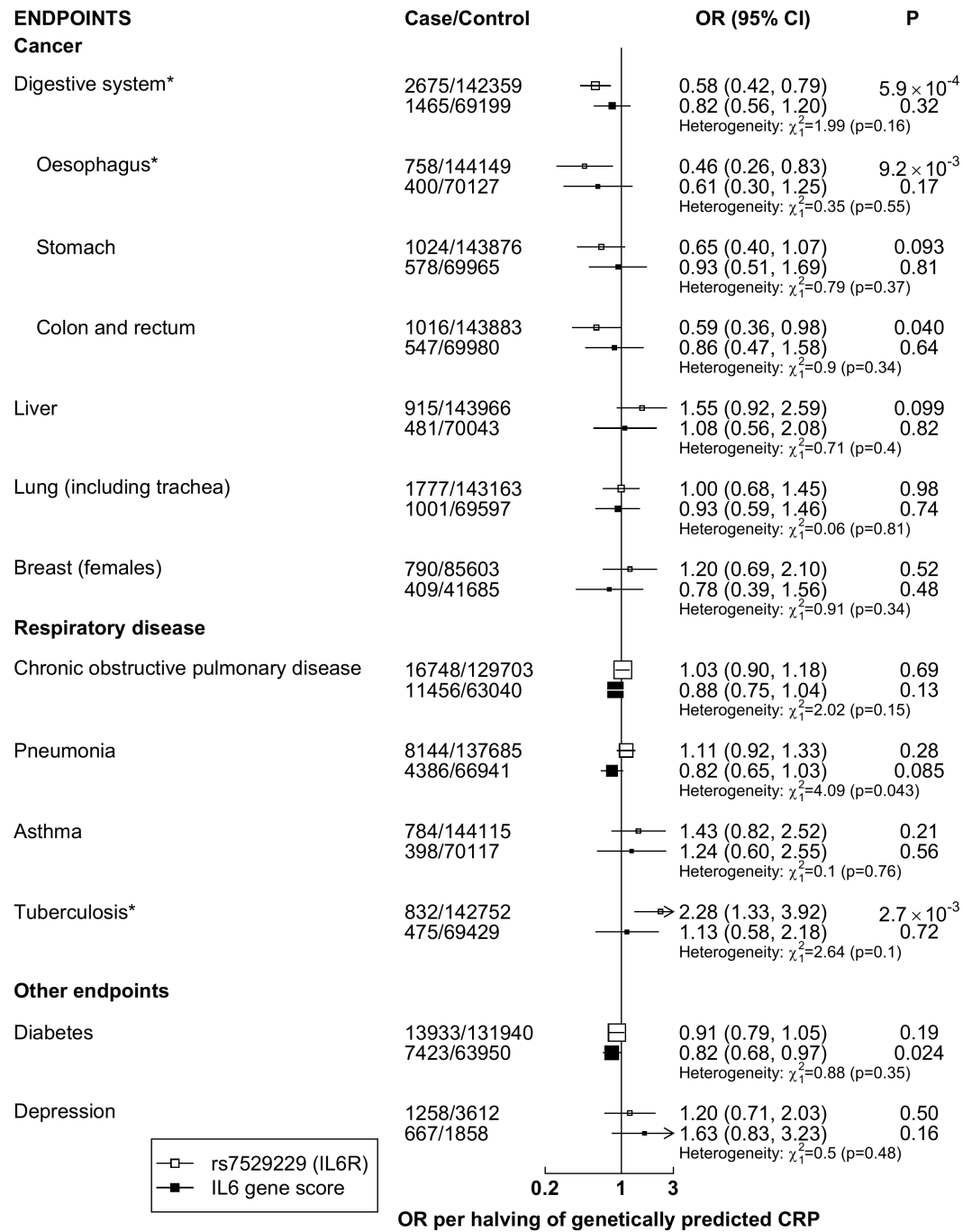
As for CVD (Chapter 4), I constructed logistic regression models to investigate associations of the IL-6(R) instruments with non-vascular disease outcomes. These models were adjusted for age, age², sex, and principal components of genetic ancestry; each model was stratified by 10 regions and the resulting estimates meta-analysed across using the fixed-effect method with inverse variance weighting.

The IL-6R instrument was associated with a lower risk of digestive cancers at 5% FDR (OR [95% CI] per halving of genetically predicted CRP: 0.58 [0.42, 0.79], $P = 5.9 \times 10^{-4}$), especially oesophageal cancer (OR [95% CI]: 0.46 [0.26, 0.83], $P = 9.2 \times 10^{-3}$) and a nominally significant reduced risk of colorectal cancer (OR [95% CI]: 0.59 [0.36, 0.98], $P = 0.04$). For equivalent differences in CRP, there was no evidence of associations of the IL-6 instrument with any digestive cancers but associations were consistent with those for IL-6R ($P_{het} = 0.16$). There was no evidence of associations observed with IL-6R or IL-6 instruments with non-digestive cancers (liver, lung and breast).

On assessment of associations with respiratory disease, there was a higher risk of tuberculosis associated with the IL-6R instrument (OR [95% CI] per halving of genetically predicted CRP: 2.28 [1.33, 3.92]), but no significant association was observed with the IL-6 instrument, although the two results were consistent ($P_{het} = 0.10$). Associations of the IL-6(R) instruments with asthma and COPD were largely null (i.e. no evidence of association).

Both IL-6R and IL-6 instruments had no evidence of association with depression. Although for the IL-6 instrument, there was a nominally significant lower risk of diabetes (OR [95% CI]: 0.82 [0.68, 0.97]), this did not survive multiple-testing correction at 5% FDR, and there was no association observed for the IL-6R instrument.

There is a possibility that the associations of the IL-6 instrument may not represent associations of altered IL-6 signalling. As a sensitivity analysis, I compared associations of the IL-6 instrument with rs13306436 (SNP with regulatory function on *IL6* gene expression³⁹⁴, which is part of the IL-6 instrument). When using rs13306436, I observed a similar pattern of associations with these prespecified non-vascular disease outcomes (Appendix Figure A.8).



*Indicates outcome significant at 5% FDR for IL-6R instrument

Figure 5.1: Association of the IL-6R and IL-6 instruments with prespecified non-vascular outcomes.

The figure illustrates associations of the IL-6R and IL-6 instruments with prespecified non-vascular outcomes in East Asians (CKB). Models for MR analysis in CKB adjusted for age, age², sex and principal components of genetic ancestry in each of the 10 regions, and meta-analysed across the regions using fixed-effect method with inverse variance weighting. Error bars represent 95% CI. Area of boxes are proportional to inverse of the variance of effect.

5.3.2 Associations of the IL-6(R) instruments with clinical chemistry measurements in CKB

To further clarify the phenotypic signatures of IL-6 signalling, I explored associations of the IL-6(R) instruments with additional clinical chemistry and amino acid measurements (Figure 5.2). Because this was a hypothesis-generating exercise, the selection of this panel of biological molecules were based on what was available in CKB. Covariates used in these models were identical to what was used for assessment of associations with non-vascular disease outcomes (Section 5.3.1), but with an additional covariate, ascertainment status, since clinical chemistry and amino acid profiles were obtained from a nested case-control cohort of CVD in CKB (Section 4.2.1).

As with the assessment of the IL-6(R) instruments with CVD risk factors and lipids in the previous chapter (Chapter 4), a useful output from this exercise would be to identify any potential associations of altered IL-6(R) signalling that might explain the association of the IL-6R instrument with lower risk of atherosclerotic CVD outcomes, lower risk of digestive cancers or a higher risk of tuberculosis; or alternatively, explain the null associations of the IL-6 instrument with similar outcomes. An additional aim was to draw comparisons between associations of the IL-6(R) instruments with these additional biomarkers, and associations of tocilizumab with the same biomarkers in RCTs; this would enable us to understand which instrument, IL-6R or IL-6, is better in reflecting associations of tocilizumab (Table 5.1).

Association of the IL-6(R) instruments with markers of inflammation

In Chapter 4.3.3, I reported associations of the IL-6(R) instruments with CRP. Here, I assessed associations of the IL-6(R) instruments with other markers of inflammation (Figure 5.2).

As expected, both the IL-6R and IL-6 instruments associated with lower levels of plasma fibrinogen (IL-6R, Effect [95% CI] in SD per halving of genetically predicted

CRP: -0.34 [-0.50, -0.19], $P = 1 \times 10^{-5}$; IL-6, -0.47 [-0.61, -0.34], $P = 1.7 \times 10^{-11}$). Contrary to expectations, a directionally discrepant association between the IL-6(R) instruments were observed with circulating albumin concentrations (P -het between instruments = 3.2×10^{-6}), i.e. the association of the IL-6R instrument with albumin was in the opposite direction (and included the null) to associations observed with the IL-6 instrument. The association of IL-6 with higher albumin was significant at 5% FDR (IL-6, Effect [95% CI] in SD per halving of genetically predicted CRP: 0.26 [0.16, 0.37], $P = 9.3 \times 10^{-7}$). Levels of albumin tend to reduce in systemic inflammation⁴²¹ and, therefore, the association of higher albumin with lower systemic inflammation (that the IL-6 instrument is orientated to) is biologically plausible. The associations of the IL-6R instrument with fibrinogen and IL-6 instrument with albumin were consistent with effects observed in RCTs of tocilizumab (Table 5.1).

There was no evidence of association of the IL-6 instrument with glycoprotein acetyl. The IL-6R instrument was associated with lower plasma concentrations of glycoprotein acetyl but the effect estimate of the IL-6R instrument included the null; hence any interpretation of the association of the IL-6R instrument with glycoprotein acetyl warrants additional studies with larger sample sizes. Glycoprotein acetyl, like CRP, is a recognised inflammation marker and observationally associated with incident CVD outcomes in CKB¹⁰¹.

Association of the IL-6(R) instruments with liver enzymes

Many inflammation biomarkers, including CRP and fibrinogen, are synthesized in the liver in response to IL-6 signalling^{205,302}. Whether IL-6R or IL-6 signalling is associated with the synthesis of other hepatic enzymes or proteins is unclear. All associations were directionally consistent between the IL-6(R) instruments (Figure 5.2). While associations of the IL-6R instrument with alanine transaminase (ALT) and aspartate transaminase (AST) were null, the IL-6 instrument was associated with higher levels of ALT (IL-6, Effect [95% CI] in SD per halving of genetically predicted CRP: 0.15 [0.04, 0.25], $P = 5.5 \times 10^{-3}$) that was significant at 5%

FDR. With the IL-6R instrument, altered IL-6R signalling was associated with lower gamma-glutamyl transferase (GGT) at nominal significance but largely null associations observed with IL-6. Clinical trials of tocilizumab reported higher levels of AST and ALT in tocilizumab-treated patients compared to placebo (Table 5.1), consistent with the association of the IL-6 instrument but not the IL-6R instrument (IL-6R instrument had null associations with both AST and ALT).

Association of the IL-6(R) instruments with markers of renal function

Observational studies have suggested higher levels of creatinine⁴²², cystatin-C⁴²³, and uric acid⁴²⁴ were associated with a higher risk of CVD. Mendelian randomisation studies were available only for cystatin-C⁴²⁵ and uric acid^{426,427}; evidence for a causal association of higher uric acid with risk of CVD was inconsistent, with a recent study demonstrating associations of genetically determined plasmic uric acid with CVD were likely acting via non-urate pleiotropic pathways. For cystatin-C, there was no evidence of the causal role of cystatin-C observed for CVD outcomes.

Three markers of renal function were assessed for association with the IL-6(R) instruments - creatinine, cystatin-C, and uric acid (Figure 5.2). Although the associations between IL-6R and IL-6 instruments were directionally consistent with respect to their associations with creatinine, cystatin-C, and uric acid, only the association of IL-6 instrument with lower cystatin-C was significant at 5% FDR (IL-6, Effect [95% CI] in SD per halving of genetically predicted CRP: -0.13 [-0.23, -0.03], $P = 8.3 \times 10^{-3}$). Furthermore, as with inflammation markers and liver enzymes above, null associations of the IL-6R instrument with markers of renal function suggest that it is unlikely that the cardio-protective benefits of altered IL-6R signalling can be explained by alteration of markers of renal function. The lack of studies examining the effect of tocilizumab on markers of renal function limits the comparison of the direction of associations of the IL-6(R) instruments with markers of renal function in the present study.

Association of the IL-6(R) instruments with amino acids

Given the role of some amino acids in inflammation⁴²⁸, I examined the associations of the IL-6(R) instruments with the available amino acid measurement data (as measured by NMR, Chapter 2.2.4) in CKB.

Only associations of IL-6 with histidine were significant at 5% FDR (IL-6, Effect [95% CI] in SD per halving of genetically predicted CRP: 0.28 [0.08, 0.48], $P = 6.1 \times 10^{-3}$) (Figure 5.2) which was directionally consistent between the IL-6R and IL-6 instruments. As with markers of renal function, the lack of studies examining the effect of tocilizumab on amino acids limit the comparison of direction of associations of the IL-6(R) instruments with amino acids in the present study

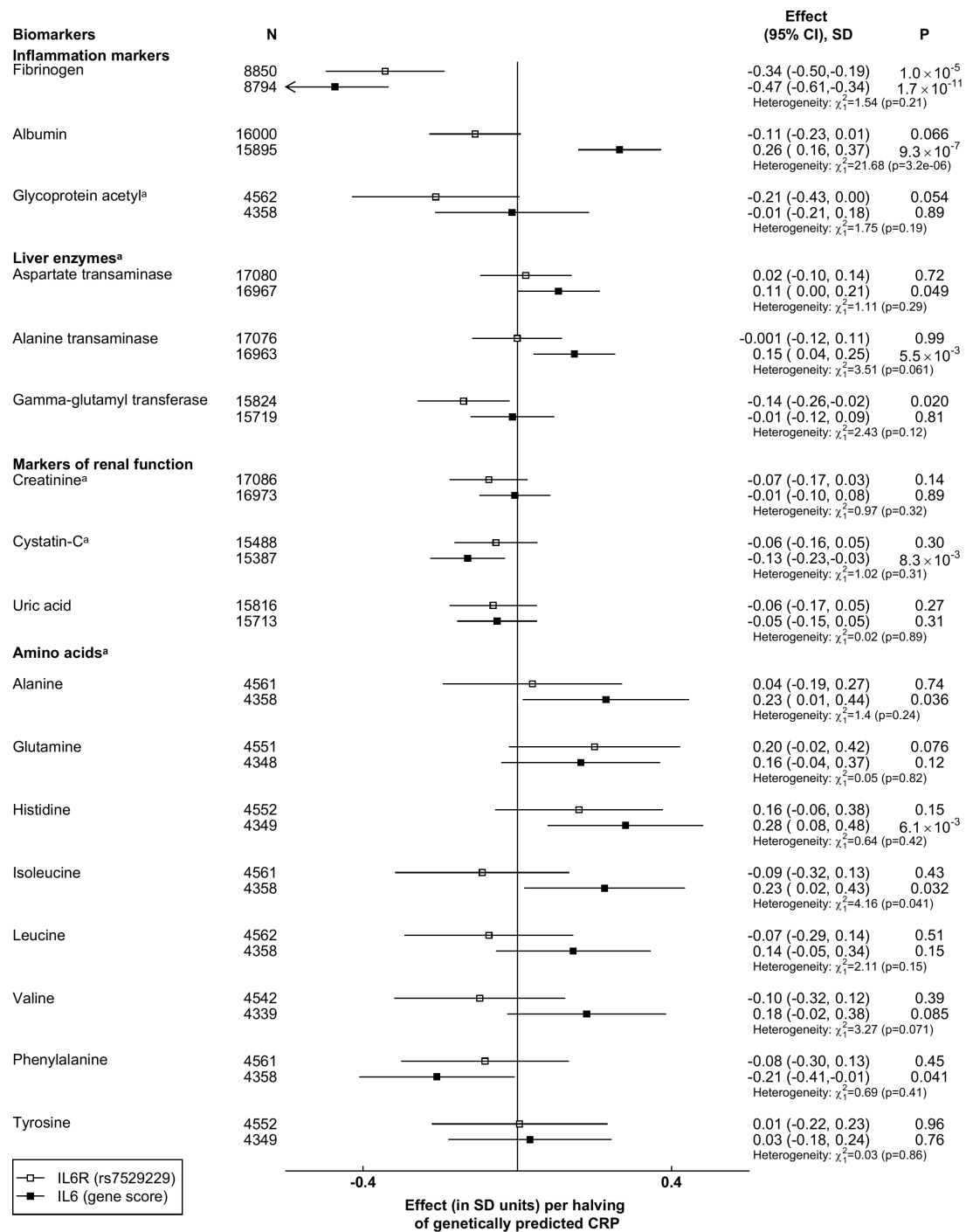
Table 5.1: Summary of comparisons of clinical chemistry measurements between effect of tocilizumab and IL-6R/IL-6 instruments

Trait	Effect of Tocilizumab ¹	Associations of	
		IL-6R instrument ²	IL-6 instrument ²
Inflammation markers			
CRP	lowers	lower*	lower*
Fibrinogen	lowers	lower*	lower*
Albumin	increases	lower	higher*
Liver enzymes			
ALT	increases	null	higher*
AST	increases	null	higher
Renal function markers			
Uric acid	-	null	null
Creatinine	-	null	null
Cystatin-C	-	null	lower*
Amino acids			
Alanine	-	null	higher
Glutamine	-	null	null
Histidine	-	null	higher*
Isoleucine	-	null	higher
Leucine	-	null	null
Valine	-	null	null
Phenylalanine	-	null	lower
Tyrosine	-	null	null

*Significant at 5% FDR

¹ Clinical trials tested tocilizumab at 8mg/kg and documented by Swerdlow et al [97]

² Orientated to represent lower systemic inflammation as indicated by lower plasma CRP concentrations



^aThese biomarkers were log-transformed prior to standardization

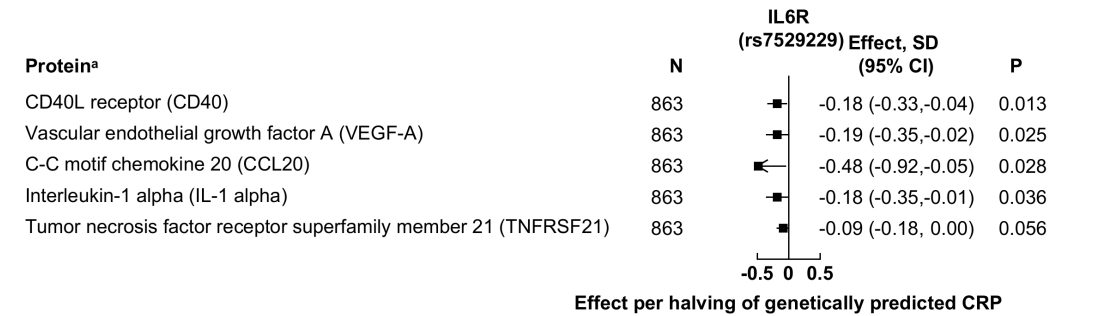
Figure 5.2: Associations of the IL-6(R) instruments with additional clinical chemistry and amino acid measurements.

Illustrates associations of the IL-6(R) instruments with other inflammation-related biomarkers, liver enzymes, markers of renal function and amino acids assayed in CKB. Models for MR analysis in CKB adjusted for age, age², sex, ascertainment status, and principal components of genetic ancestry in each of the 10 regions, and meta-analysed across the regions using fixed-effect method with inverse variance weighting. Error bars represent 95% CI. Area of boxes are not proportional to inverse of the variance of effect. Error bars represent 95% CI.

5.3.3 Proteomic associations of IL-6/IL-6R instruments

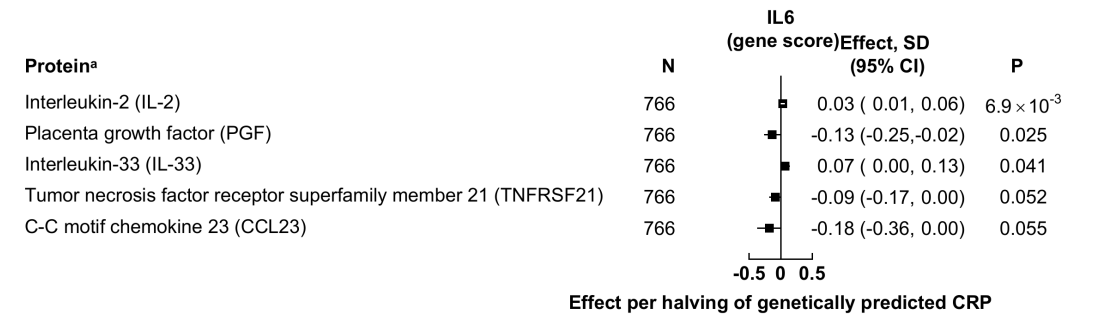
To characterize any systematic patterns of proteomic associations of the IL-6(R) instruments beyond association with IL-6 protein (Figure 4.4), I explored associations of these instruments with 91 proteins assayed by the immune-oncology panel of OLINK⁴²⁹ in a nested pancreatic cancer case-control subset in CKB (Details of proteomic measurement are provided in Chapter 2.2.4). Because the sample sizes were small compared to sample sizes for outcomes, models were not stratified by region. Instead, region was included as a covariate in the model.

There was no evidence of proteomic associations with the IL-6(R) instruments at 5% FDR. To illustrate the null associations for both instruments, associations with the IL-6(R) instruments were arranged by increasing P -values and the first 5 associations for each instrument are shown in Figure 5.3 and 5.4. Full associations with the 91 proteins are provided in the appendix Figure A.9.



^aProteins were log-transformed prior to standardization

Figure 5.3: Proteomics associations of the IL-6R instrument.
The figure illustrates the first 5 proteomic associations with the IL-6R instrument from among associations with 91 proteins (Appendix Figure A.9), when arranged in order of increasing *P*-values. Proteins were assayed by OLINK using a proximity extension assay technique (Chapter 2.2.4). For each protein, models for MR analysis in CKB adjusted for age, age², case status, sex and principal components of genetic ancestry and region. Error bars represent 95% CI. Area of boxes are not proportional to inverse of the variance of effect.



^aProteins were log-transformed prior to standardization

Figure 5.4: Proteomics associations of the IL-6 instrument.
The figure illustrates the first 5 proteomic associations with the IL-6 instrument from among associations with 91 proteins (Appendix Figure A.9), when arranged in order to increasing *P*-values. Proteins were assayed by OLINK using a proximity extension assay technique (Chapter 2.2.4). For each protein, models for MR analysis in CKB adjusted for age, age², case status, sex and principal components of genetic ancestry and region. Error bars represent 95% CI. Area of boxes are not proportional to inverse of the variance of effect.

5.3.4 Exploratory associations of IL-6R/IL-6 instrument with non-prespecified disease outcomes in CKB

As an exploratory analysis, I carried out a phenome-wide screen in CKB for associations of the IL-6(R) instruments, selecting 35 ICD-10 coded outcomes with at least 500 cases with genetic data (Figure 5.5). At an alpha of 0.05/35, the present MR-PheWAS analyses had more than 50% power to detect an OR of 0.85 and more than 80% power to detect an OR of 0.80. Covariates used in the MR-PheWAS models were identical to what was used for assessment of associations with prespecified non-vascular disease outcomes (Section 5.3.1), with models for each outcome stratified by 10 regions and estimates meta-analysed across using fixed-effect inverse variance weighting.

There was no evidence of associations for any of these endpoints with either the IL-6R or IL-6 instruments (Figure 5.5).

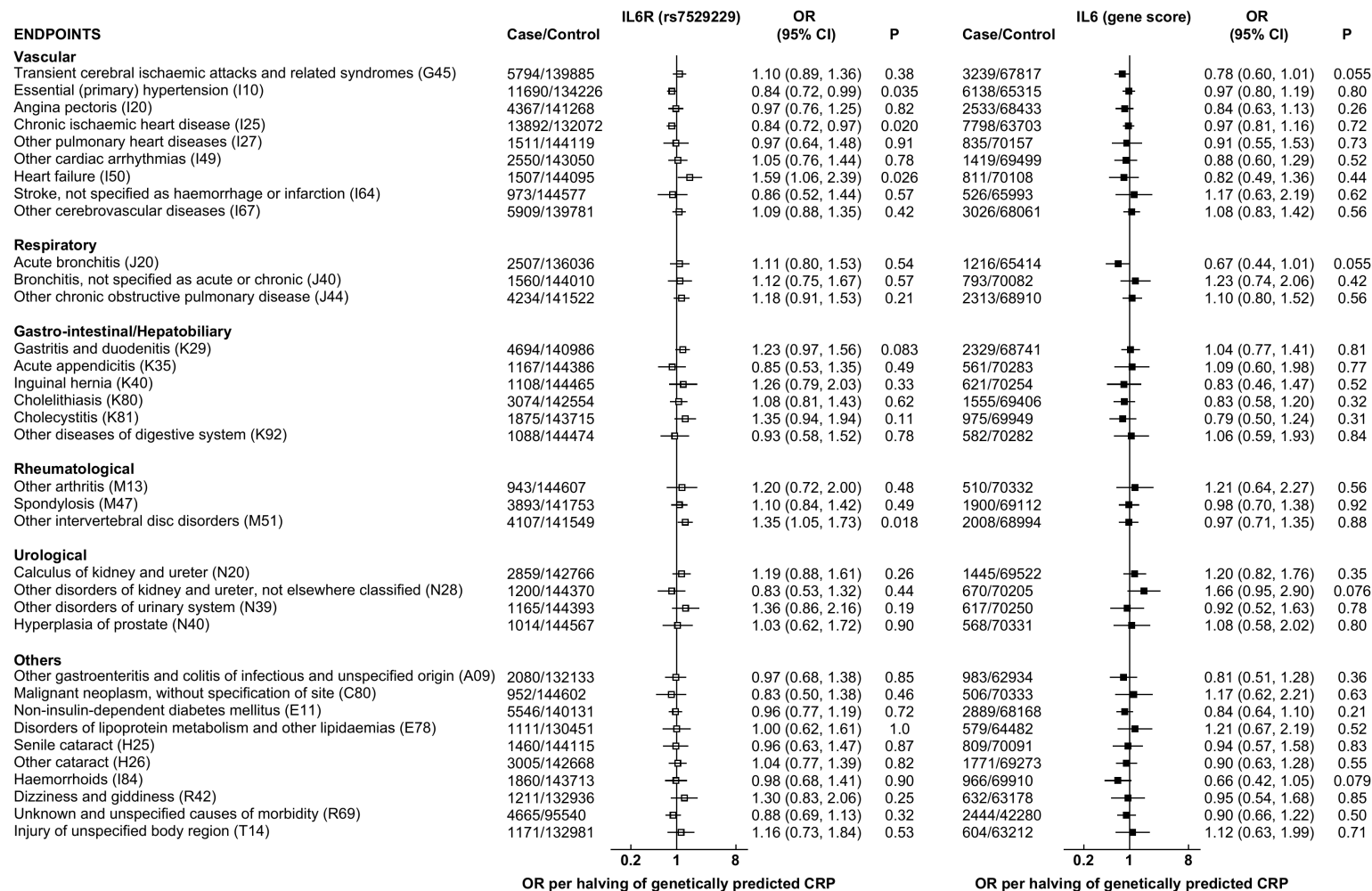


Figure 5.5: Phenome-wide associations of the IL-6R and the IL-6 instruments.

The figure illustrates phenome-wide associations of the IL-6R and the IL-6 instruments in CKB. Endpoints in the figure were annotated as outcomes (ICD-10 code) and those selected had at least 500 *IL6R* genotyped cases, offering more than 80% power to detect an OR as large as 0.80 (Chapter 5.2.1). For each outcomes, models for MR analysis in CKB adjusted for age, age², sex and principal components of genetic ancestry in each of the 10 regions, and meta-analysed across the regions using fixed-effect method with inverse variance weighting. Error bars represent 95% CI. Area of boxes are not proportional to inverse of the variance of effect.

5.4 Discussion

Interleukin-6 signalling is an important biological pathway and has been the subject of a number of experimental and epidemiological studies. Most epidemiological associations investigated associations of plasma IL-6R concentrations (or genetically predicted plasma IL-6R concentrations) with cardiovascular disease outcomes^{96,97,285,286,377}. Some of these studies have also performed PheWAS using the functional IL6R genetic variant (rs2228145) to assess associations beyond vascular outcomes^{97,285,286,377}. The work presented in this chapter builds on these previous works in three ways. First, the use of the IL-6 instrument alongside the IL-6R instrument is novel and has enabled comparison of estimates between the IL-6 receptor and ligand that is expected to provide additional confidence (or reduce confidence) on evidence linking the relevance of traits to the IL-6 signalling pathway. Second, the dual IL-6(R) phenome-wide assessment reported in this Chapter is the largest of its kind on East Asians, leveraging the dense phenotyping and the relatively large genotyped population in CKB. Third, the phenome-wide assessment revealed novel associations of the IL-6R instrument with digestive cancers (with consistent IL-6 instrument estimates but including the null), but confidence in this association is pending independent replication in other cohorts (currently limited by the lack of publicly available cancer GWAS). Furthermore, analysis of the IL-6(R) instruments with clinical chemistry measurements showed that both instruments associated similarly with fibrinogen (i.e. associated with lower fibrinogen levels) but directionally discrepant with respect to their associations with albumin. Both fibrinogen and albumin are produced by the liver and modulated in response to inflammation; in acute inflammation, levels of fibrinogen usually increase and albumin concentrations are usually reduced. The pattern of association of the IL-6 instrument (i.e. associated with lower circulating fibrinogen but higher albumin) is more consistent with the effect of tocilizumab on the same biomarkers than the pattern of association exhibited by the IL-6R instrument, suggesting the IL-6 instrument may be a better proxy for effects of tocilizumab than the IL-6R instrument. On exploratory analysis, there was no evidence of associations of

the IL6(R) instruments with any of the phenome-wide or proteomic outcomes in CKB.

5.4.1 Interleukin-6 signalling and cancers

For the first time to my knowledge, the present study showed that altered IL-6R signalling represented by the IL-6R instrument is associated with a reduced risk of digestive cancers in general, and oesophageal carcinoma in particular.

The association of altered IL-6R signalling with digestive cancers has implications for future studies that investigate drugs that inhibit components of the IL-6R signalling pathway for the prevention of digestive cancers. Furthermore, this observation is supported by a number of experimental studies that have documented the relevance of IL-6 signalling (especially the STAT3 signalling pathway) for tumour progression⁴³⁰. Specific to oral upper digestive cancers, *IL6* knockout suppressed tumorigenesis in physiologically relevant 3D organotypic and 3D tumoroid models and murine models of squamous cell variety of oesophageal cancer⁴³¹. The signalling pathways identified in the study for the development of oesophageal cancer were the STAT3 and ERK1/2 signalling pathways, and administration of tocilizumab (that blocks both mIL-6R and sIL-6R) suppressed tumour growth in 100% of the *in vitro* 3D tumoroid models of oesophageal cancer⁴³¹. The association of the IL-6R instrument with digestive cancer traits is also supported by the generally higher IL6R gene expression, compared to IL6 gene expression, in digestive tract tissues (Figure A.10) and the association of the IL-6R instrument with the relatively higher IL6R gene expression in digestive tract tissues in a cross-tissue comparison (Figure A.11).

Additionally, I attempted to examine the association of the IL-6R instrument with digestive, oesophageal, and colorectal cancer in other studies. To assess well-powered GWAS of equivalent phenotypes in UKBB (as processed by Zhou et al⁴³²), I was only able to find the colorectal cancer phenotype that had a sample size of cases over 1000 (oesophageal cancer had only 162 cases and therefore was not assessed

Table 5.2: Look-up and comparison of IL-6R instrument (rs7529229-C) signal in other colorectal cancer case-control cohorts

Study	Ancestry	Case/Control	OR ¹ (95% CI)	P-value ²	P for heterogeneity ³
CKB	Chinese	1016/143883	0.91 (0.83, 1.00)	0.04	0.14
Tanikawa et al [434]	Japanese	10405/34511	0.99 (0.95, 1.03)	0.72	
Zhou et al [432]	Europeans (UKB)	4562/382756	1.01 (0.96, 1.05)	0.76	

¹ Odds ratio per 7529229-C allele² P-value testing the null hypothesis of no association between rs7529229-C and colorectal cancer risk.³ P-value testing the null hypothesis of no heterogeneity across the three effect estimates tabulated.

due to the potential unreliability of associations as a result of low statistical power). I also identified several GWAS of digestive cancer traits in the GWAS catalog (EFO codes: EFO_1000218 for digestive cancer, EFO_0000478 for oesophageal adenocarcinoma, and EFO_0005842 for colorectal cancer)⁴³³; however, summary statistics were publicly accessible only for two colorectal cancer GWAS (one of them from the UKBB study identified above in Zhou et al⁴³²), that showed no evidence of replication of association of rs7529229-C with colorectal cancer (Table 5.2). The limited accessibility of summary statistics for digestive cancer and oesophageal cancer GWAS prevents any assessment of replication of observed signals for digestive cancer and oesophageal cancer in CKB with other studies. The lack of replication of signals of the IL-6R instrument with digestive cancer traits in other studies is a reason to exercise caution on the reliability of the signal observed in the present study.

With other cancer types, a previous study that assessed associations of the IL-6R instrument with cancer outcomes reported null findings with all cancer, breast cancer and colorectal cancer⁹⁷. The present investigation also reported no evidence of association of the IL-6R instrument with breast or colorectal cancer, consistent with the previous report⁹⁷. However, in contrast to a previous Mendelian randomisation study that observed a lower risk of liver cancer associated with genetically predicted lower plasma IL-6 concentrations⁴³⁵, I did not observe any association of IL-6R or IL-6 instrument with liver cancer in CKB. This may be due to the use of an IL-6 instrument by Tian et al⁴³⁵ constructed using a single SNP rs1800795, that is not in linkage disequilibrium (LD) with any of the 3 *IL6* SNPs used in CKB; the effect

of rs1800795 on plasma IL-6 concentrations as reported by the different studies collated by Tian et al was, as acknowledged in the study, highly heterogeneous (I^2 more than 99%) limiting the reliability of its use as an IL-6 instrument to assess the association of IL-6 signalling on liver cancer. Furthermore, most of the cohorts that contributed cases of liver cancer for meta-analysis by Tian et al were predominantly Europeans, and the observed null association in East Asians in the present study raises the possibility of a gene by environment interaction of IL-6 signalling influencing the development of liver cancer.

5.4.2 Interleukin-6 signalling and prespecified non-cancer outcomes

For the non-cancer non-vascular endpoints, the present study does not support a causal role of IL-6R or IL-6 in COPD, asthma, pneumonia or depression, conditions which were previously linked to dysregulated IL-6 signalling^{416,417,420,436}.

Null associations with COPD and overall asthma in the present study are consistent with previous MR studies^{280,437}; using data from the UK Biobank and ECLIPSE COPD cohort, Farahi et al used the same IL-6R instrument as the present study and showed a higher risk of atopic asthma (OR [95% CI] per minor allele of rs2228145: 1.07 [1.01, 1.13]) but null associations of overall asthma (OR [95% CI]: 1.02 [0.99, 1.05]) and COPD (OR [95% CI]: 1.02 [0.96, 1.07])²⁸⁰ consistent with null associations of COPD (OR [95% CI] per halving of genetically predicted CRP: 1.03 [0.90, 1.18]) and asthma (OR [95% CI]: 1.43 [0.82, 2.52]) observed in the present study.

With the IL-6 instrument, there was a nominally significant reduced risk of diabetes (OR [95% CI] per halving of genetically predicted CRP: 0.82 [0.68, 0.97]), supporting previous studies linking IL-6 with type 1 and type 2 diabetes^{419,438}. Recent studies also suggested a role of interleukin-6 in the pathogenesis of depression especially in rheumatoid arthritis (RA) patients⁴³⁹. Using an MR approach, a study showed a protective effect of the IL-6R instrument for severe depression and/or psychosis (OR [95% CI]: 0.65 [0.44, 0.95]) but associations were null for specific depression

or psychosis outcomes⁴⁴⁰. The null findings with specific depression outcome are consistent with the present study which also showed null associations of IL-6R with overall depression (OR [95% CI] per halving of genetically predicted CRP: 1.20 [0.71, 2.03]), suggesting either there is no evidence of a causal association of IL-6 signalling in depression or the effect may be limited to patients with high background levels of inflammation (e.g. RA patients). An RCT of tocilizumab (IL-6 receptor blocker) is currently ongoing that will clarify the MR findings and evaluate the potential of tocilizumab to treat depression in patients with evidence of low-grade systemic inflammation⁴⁴¹.

5.4.3 Interleukin-6 signalling and blood biomarkers

To further clarify phenotypic signatures of IL-6 signalling, I explored associations of the IL-6(R) instruments with additional clinical chemistry and amino acid measurements.

I observed both instruments to be associated with lower fibrinogen concentrations (associations of IL-6R consistent with previous MR studies^{96,97}) but directionally discordant associations with albumin - albumin is considered a ‘negative’ acute phase protein, which means that levels of albumin are inversely correlated with inflammation⁴⁴². Association of the IL-6R instrument with albumin in the present study, although null, is directionally opposite to associations of IL-6R observed for albumin in Europeans⁹⁷. Whether this directional inconsistency between Europeans and East Asians with respect to the association of the IL-6R instrument with albumin is due to a gene by environment interaction is presently unclear. The IL-6 instrument, however, was positively associated with albumin and, therefore, consistent with the negative correlation of albumin with CRP⁴⁴². Furthermore, the association of the IL-6 instrument with albumin was directionally consistent with the effect on albumin reported in tocilizumab clinical trials, suggesting the IL-6 instrument may be a better genetic instrument representing the effects of tocilizumab than the IL-6R instrument.

In addition, I observed suggestive (nominal) associations of the IL-6R instrument with lower levels of gamma-glutamyl transferase (GGT), consistent with the observational linear trends of GGT with CRP⁴⁴³. The IL-6 instrument, on the other hand, exhibited nominally significant associations with plasma histidine. Dietary histidine is hypothesized to play roles in the prevention of obesity and diabetes⁴⁴⁴, and associations of IL-6 with higher levels of histidine observed in the present study may underlie the association of IL-6 with a lower risk of diabetes as observed in Figure 5.1. The lack of studies exploring associations of the IL-6(R) instruments with GGT and amino acids limited comparisons of associations observed in the present study with external studies.

5.4.4 Phenome-wide and proteomic associations of interleukin-6 signalling

To explore any additional associations of the IL-6(R) instruments that can be used to generate new hypotheses, I carried out a phenome-wide screen in CKB selecting ICD-10 coded outcomes that have at least 500 cases with genetic data and also assessed associations of the instruments with a panel of 91 proteins. There was no evidence of additional phenome-wide or proteomic associations with either the IL-6R and IL-6 instruments observed in the present study.

5.4.5 Strength and limitations

Because the present MR study used a similar approach to Chapter 4, many of the strengths and limitations of the MR approach were already described in Chapter 4.4.6.

A major strength of the present MR study was the unprecedented depth of investigations, made possible by the dense phenotyping in CKB, carried out using a dual receptor-ligand MR approach. From our understanding of the IL-6 signalling pathway, we would expect to observe consistent estimates between IL-6R and IL-6 instruments, given that both these instruments represent lower systemic inflammation via lower classic signalling (that results in lower CRP concentrations).

The IL-6 instrument, additionally, represents lower trans-signalling, but the direction of trans-signalling with the IL-6R instrument is unclear (Chapter 4.4.1). Hence, while inter-instrument consistency of associations are helpful to understand what associations are likely operating via lower classic signalling pathway, interpretation of associations that are heterogeneous between the IL6R and IL6 instruments remains unresolved. Research on instruments specifically representing trans-signalling (e.g. genetic instruments from a GWAS of the ratio of IL-6:sIL-6R complex to IL-6:sIL-6R:sgp130 complex) are warranted in future studies. From the point of view of safety of putative drugs that would mimic the IL-6R or IL-6 instrument, except for the higher risk of tuberculosis observed with the IL-6R instrument, there was no evidence of additional associations of IL-6R or IL-6 instrument with any phenotypes or proteomic measurements in CKB at 5% FDR level of significance.

An additional limitation (that is typically common to all PheWAS and the present study is no exception) is the high number of non-vascular outcomes assessed that can increase the number of false-positive associations; although I have applied the Benjamini-Hochberg method⁴⁴⁵ for multiple hypotheses testing corrections to control the false discovery rate (FDR), false-positive associations may remain. Hence independent replication of non-vascular associations of IL-6 and IL-6R are warranted in future studies. Finally, MR associations of the instruments with traits (or associations with gene expression) reported in the present study may be confounded by linkage disequilibrium (LD) of the SNPs used in the instrument with SNPs associated with the trait. Future investigations using tests such as genetic co-localisation (a fairly recent development that occurred at the final stages of drafting my thesis²⁴³) are warranted to verify whether associations were confounded by LD. Nevertheless, given how the genetic instruments in the present study were constructed, i.e based on association with CRP and whether there was any prior biological evidence for their use as genetic instruments, it appears unlikely that reported associations would be confounded by LD.

5.5 Conclusion

The use of the IL-6(R) instruments enabled a dual receptor-ligand based phenome-wide assessment of IL-6 signalling in a large East Asian cohort. Genetically predicted lower plasma CRP concentrations, represented by the IL-6(R) instruments, was associated with a lower risk of digestive cancers and a higher risk of tuberculosis using the IL-6R instrument (with consistent estimates using the IL-6 instrument but estimates included the null). Assessment of associations of the IL-6(R) instruments with blood biomarkers revealed directionally consistent associations with effects of tocilizumab on the same biomarkers (where comparisons were possible), except in the case of circulating albumin: only association of the IL-6 instrument with circulating albumin was consistent with effects of tocilizumab on plasma albumin concentration (association of the IL-6R instrument with albumin included the null), suggesting that IL-6 instrument may be a better proxy for tocilizumab than the IL-6R instrument. The present work demonstrates the utility of dual receptor-ligand based MR studies to clarify epidemiological associations of genetic instruments representing the receptor and ligand, and critically appraise potential therapeutic and adverse effects of modulating targets part of the same signalling pathways.

6

Discussion

My DPhil research aimed to understand the epidemiological relevance of components and markers of systemic inflammation to CVD outcomes. In particular, I wanted to address several knowledge gaps that remained with regards to observational associations of CRP and fibrinogen, which are fairly well-established markers of systemic inflammation, and genetic associations of components of interleukin-6 signalling, which influence the hepatic synthesis of CRP and fibrinogen, with CVD, CVD-related traits, and any relevant non-CVD outcomes. In this section, I discuss the knowledge and evidence gaps that motivated the research presented in my thesis and summarise evidence in the context of the wider literature.

6.1 CRP, Fibrinogen, and CVD outcomes

Observational epidemiological studies over the years have consistently shown that higher concentrations of plasma CRP and fibrinogen were associated with higher risks of CHD, IS, and ICH^{92,93,304–310}. However, only a few observational studies have evaluated these CRP-/fibrinogen - CVD associations in non-Europeans^{230,231,311}.

Although the underlying biology of diseases or traits may be shared across different ethnic groups, epidemiological associations informing biology of diseases in one

ethnic group may not necessarily be portable to a different ethnic group^{233,234}; one reason this can occur is due to differences in the distribution of potential effect measure modifiers²³². Moreover, while many studies have investigated associations of CRP and fibrinogen with CHD and IS, far fewer studies (that were largely likely underpowered^{230,231,311}) have assessed associations of the inflammatory markers with ICH - a subtype of haemorrhagic stroke, that is becoming increasingly prevalent²³⁷ in non-European countries, like China.

The present study showed, similar to studies in Europeans⁹², higher usual levels of plasma CRP were associated with higher risks of CHD, IS, and ICH. The associations of CRP with the CVD outcomes were independent of several well-established CVD risk factors, including BMI, SBP, and plasma LDL-C concentrations, among others. There was also no evidence of effect modification of the CRP-CVD associations by pre-specified subgroups.

However, unlike reports in European studies⁹³, higher usual levels of plasma fibrinogen were not associated with higher risks of CHD or the stroke types in my analysis. A major limitation of my analysis of plasma fibrinogen with CVD outcomes was that only half the study participants who had measurements of CRP, had measurements of plasma fibrinogen ($n = \sim 9,000$), suggesting one possible reason for the largely null fibrinogen-CVD association may be due to low statistical power. The alternative explanation of the null fibrinogen-CVD association is that it is a real observational association specific to East Asians and was not observed in Europeans due to the aforementioned trans-ethnic differences in the distribution of potential effect measure modifiers. Larger East Asian cohorts are required to resolve whether the present scenario of neutral fibrinogen-CVD association reported in my thesis represents a case of an underpowered study or illustrates a true trans-ethnic difference in the epidemiological association due to differences in the distribution of anthropometric traits (e.g. height, BMI) that are associated with CVD risk or known CVD risk factors (e.g. LDL-C, SBP)^{353,356}.

Nevertheless, observational associations of CRP and fibrinogen with CVD outcomes in the present and the previous studies are discordant with their respective genetic epidemiological associations. Mendelian randomisation studies, using genetic variants in or near the relevant gene encoding CRP or fibrinogen, have reliably shown that observational associations of CRP and fibrinogen with CVD outcomes were unlikely to be causal^{94,289,360,446}. This means that whilst CRP and fibrinogen remain useful as markers of general inflammation and as predictors of CVD outcomes, they are unlikely to be useful as therapeutic targets. This led me to search for inflammation-associated therapeutic targets that influence the synthesis of circulating CRP and fibrinogen that are likely to have translational value.

The studies by Ridker et al^{205,309} and the CANTOS³⁰³ and CIRT²⁵⁹ trials (that were published during my DPhil) demonstrated the potential therapeutic value of targeting the IL-1 β -IL-6 axis for CHD. While the CIRT study using low-dose methotrexate showed no effect on circulating levels of IL-6 or CRP and no evidence of a therapeutic effect in recurrent CVD, the CANTOS trial using canakinumab (IL-1 β antagonist) showed reductions in plasma IL-6 and CRP concentrations and a commensurate lowering of risk of recurrent CVD events compared to placebo. This motivated my research on IL-1 β and IL-6 signalling components and resulted in the subsequent chapters where I report genetic epidemiological investigations of components of the IL-6 signalling pathway. It should be noted that the subsequent chapters focus on IL-6 and not on IL-1 β because I was not able to find genetic variants in the *IL1B* locus (and other inflammation-associated loci) (Figure A.3) associated with plasma CRP concentrations in CKB whereas there were several genetic variants in *IL6R* and *IL6* loci associated with circulating CRP in CKB and also reported in various European studies^{288,371}.

6.2 Interleukin-6 signalling and CVD outcomes

Evidence from MR studies has documented associations of the IL-6R protein with CVD in Europeans^{96,97,286,385}. The basis of these MR studies was the use of a genetic variant (rs2228145-C, *Asp358Ala*) that was reported to alter IL-6 signalling by enhancing proteolytic cleavage of mIL-6R to sIL-6R²⁸¹. The *Asp358Ala* variant (will be referred to as the IL-6R instrument henceforth), in addition to being associated with higher plasma sIL-6R concentrations, was also associated with lower plasma concentrations of CRP and fibrinogen, and higher plasma concentration of IL-6 protein. The direction of association of the IL-6R instrument with sIL-6R, IL-6 protein, CRP, and fibrinogen was consistent with the direction of effects of tocilizumab (an IL-6R antagonist) on the same circulating proteins reported in tocilizumab clinical trials⁹⁷, raising the possibility that the IL-6R instrument can be used to investigate therapeutic and adverse effects of tocilizumab (and similar IL-6R antagonists).

However, given the two modes of IL-6 signalling (Figure 1.9), the higher concentrations of sIL-6R and IL-6 protein observed with tocilizumab and the IL-6R instrument would tend to promote IL-6 trans-signalling, the mode of signalling that was shown in animal models to underlie the progression of several inflammatory diseases³⁶⁷. One explanation of higher sIL-6R and IL-6 protein observed with tocilizumab was that tocilizumab competes with free IL-6 protein to bind to sIL-6R to form tocilizumab:sIL-6R complexes and prolongs the half-life of sIL-6R²⁸⁴. Similar to tocilizumab, the higher plasma sIL-6R concentrations in individuals harbouring the IL-6R instrument is likely to result in higher sIL-6R:IL-6 complexes that prolong the half-life of both sIL-6R and IL-6 protein³⁷⁸ and this can lead protein assays to detect higher concentrations of sIL-6R and IL-6 protein. It remains unclear to what extent protein assays, e.g. SomaLogic, can reliably distinguish between bound and free forms of a protein³⁷¹.

Nevertheless, while tocilizumab blocks both mIL-6R and sIL-6R (preventing IL-6 classic and trans-signalling, respectively), the higher circulating sIL-6R:IL-6 complex

associated with the IL-6R instrument, unless bound to soluble gp130 (sgp130), can still (plausibly) trigger trans-signalling. To better profile the effects of lower IL-6 classic and trans-signalling, I used genetic variants in the *IL6* locus associated with lower plasma CRP in CKB as the IL-6 instrument. An alternative design would be to select genetic variants associated with CRP from across the genome and this polygenic approach would explain more of the plasma CRP variation than the IL-6(R) instruments. This polygenic approach, however, would be susceptible to horizontal pleiotropy as observed associations may be mediated by unknown pathways. Furthermore, the polygenic approach is not appropriate when the research question concerns associations related to pharmacological perturbation of a biological target as it may yield associations that are not target-mediated (e.g. differences in polygenic vs. locus-specific associations of CRP with CHD²⁸⁷). In these cases, where investigations aim to infer the efficacy or safety of target modulation, it is important to understand whether the observed associations are target-mediated and use of cis-acting and biologically valid trans-pQTLs tend to be more informative approaches than the polygenic approach to instrument selection²⁶⁹.

Additionally, to assess whether associations are generalisable, I used the same IL-6R instrument and an equivalent IL-6 instrument from European studies to investigate associations in Europeans. An important implication of this work, beyond characterising the consistency of associations between the IL-6(R) instruments and between the ethnic groups, is whether the potential effects of drugs targeting the IL-6R receptor is equivalent to drugs targeting the IL-6 ligand, informing the design and relevance of clinical trials targeting components of the IL-6 signalling pathway.

In this chapter of my thesis, I reported associations of the IL-6R and the IL-6 instruments with CVD outcomes and CVD risk factors. In particular, I identified some associations with the IL-6R instrument in East Asians that were not consistent with the IL-6 instrument. A key finding was the association of the IL-6R instrument with lower risks of CHD and MI in East Asians and Europeans (consistent with

previous studies in largely European populations) was not consistent with the null association of the IL-6 instrument with CHD/MI that was observed in both East Asians and Europeans and persisted even when IL-6 instrument:CHD/MI estimates were meta-analysed across the two ethnic groups. Construction and analysis using the IL-6 instrument was motivated not just by the need to understand associations due to lower levels of the IL-6 ligand informative for IL-6 blockers, but also to resolve whether the associations observed with the IL-6R instrument (given its paradoxical associations with higher sIL-6R and IL-6) are representative of IL-6 signalling inhibition (i.e. relevant for IL-6R blockers). The null associations of the IL-6 instrument with CHD may, therefore, indicate that pharmacological agents that inhibit the IL-6 signalling pathway may not reduce the risk of CHD. The discrepant IL-6 receptor-ligand association for CHD/MI has implications for the development and trials of drugs targeting IL-6 receptor and IL-6 protein for these CVD outcomes. For example, although preliminary trials of ziltivekimab (IL-6 blocker)³⁸¹ increased myocardial salvage in patients with acute MI and reduced biomarkers of inflammation respectively, the present work questions their efficacy in preventing incident CVD events.

By contrast, the IL-6R instrument was also associated with a lower risk of IS and a higher risk of ICH, and in this case, these associations were consistent with the IL-6 instrument and with meta-analysed estimates consistent across both ethnic groups. The lower risk of IS associated with the IL-6R instrument was also demonstrated in a more recent study in Europeans, along with the observation that the lower risk extends to different subtypes of IS (specifically, large-artery, small-vessel, and cardio-embolic subtypes)²⁸⁶. The corresponding higher risk of ICH with both the IL-6R and IL-6 instrument, however, has not been reported by previous studies to the best of my knowledge.

The lack of evidence of an association of the IL-6 instrument with CHD and the corresponding higher risk of ICH observed with both the IL-6(R) instruments should serve as an impetus for deeper pre-clinical investigations thoroughly examining the

genetic associations in different ethnic groups in well-powered phenotype-linked bioresources. Provided there is clear evidence of replication of associations observed in my thesis, we should reconsider the value of large investments in clinical trials testing IL-6 inhibition for primary prevention of CHD.

Furthermore, although both the IL-6(R) instruments are associated with lower CRP and fibrinogen, there may be key differences in their signalling mechanisms (with one [IL-6R] altering the ratio of trans- to classical signalling and the other [IL-6] lowering both) that underlies the differences in their distinct associations (or lack of an association) with CHD. Experimental work, which is beyond the scope of my thesis, is required to disentangle the precise molecular signatures that result from the different signalling pathways. Additionally, it remains possible, despite use of locus-specific instruments, and especially for the IL-6 instrument (since the IL-6R instrument was based on a biological evidence of a mutation directly affecting the levels of the sIL-6R protein), that the associations (or the lack of associations) of the IL-6 instrument may be mediated via unknown pathways proximal to the IL-6 ligand (i.e. horizontal pleiotropy). However, the association of the CKB IL-6 instrument with the IL-6 protein in CKB provides additional confidence that the associations are likely mediated via the IL-6 protein; future studies, however, need to use larger GWAS of the IL-6 protein to develop better IL-6 instruments in Europeans to ascertain the observed null associations with CVD.

In addition to CHD^{97,384}, the IL-6R instrument was noted to also reduce risk of aortic aneurysm in a separate study³⁸⁵; due to the lack of phenotypes related to aortic aneurysm in CKB at the time of analyses, I was not able to verify the association of the IL-6(R) instruments with incidence of aortic aneurysm in CKB.

To investigate potential mechanisms underlying these associations, I examined associations of both instruments with carotid atherosclerosis, well-established CVD risk factors, lipids, and lipidomic traits (Table 6.1). The IL-6R instrument was not associated with any of the selected CVD risk factors. However, both the IL-6R

and IL-6 instruments showed directionally consistent associations with carotid atherosclerosis, consistent with a previous report in Europeans³⁸⁵. The directionally consistent associations of both instruments with carotid atherosclerosis potentially explains the consistent association between IL-6R and IL-6 instruments with lower risk of IS, but does not explain the discrepant associations between the instruments with CHD/MI.

To identify any potential biology that could explain the inconsistent associations of the IL-6(R) instruments with CHD/MI, I examined associations of both instruments with lipid traits available in CKB and compared, where possible, equivalent associations in Europeans (Table 6.1). Both the IL-6(R) instruments were associated with higher plasma concentrations of HDL-C in both East Asians and Europeans. The association of HDL-C with the IL-6R instrument, however, does not explain the lower risk of CHD/MI with the IL-6R instrument, as MR studies using instruments specific to HDL-C have shown that genetically predicted plasma HDL-C concentrations were not associated with CHD⁵¹. Additionally, there was evidence that both instruments were associated with higher plasma LDL-C concentrations. A previous IL-6R MR study reported largely null associations of the IL-6R instrument with plasma concentrations HDL-C and LDL-C, a finding that was inconsistent with clinical trials of tocilizumab⁹⁷. This led investigators to speculate that higher HDL-C/LDL-C may be an off-target effect of tocilizumab or a general sign of inflammation suppression in individuals with excessive systemic inflammation⁹⁷. However, the present study (and a more recent MR study for HDL-C²⁸⁶), using larger sample sizes, demonstrate associations of the IL-6R instrument with both HDL-C and LDL-C, that are directionally consistent with tocilizumab trials and suggest the HDL-C/LDL-C elevations may be an on-target effect of IL-6R blockade. To understand if the associations of the instruments with HDL-C/LDL-C were due to associations of the instruments with any specific lipidomic trait (e.g. HDL/LDL particles), I investigated associations of the instruments with lipidomic data available in CKB. However, there was no evidence of associations of either instrument with any lipidomic traits.

These dual receptor-ligand analyses provide three important insights. First, blockade of IL-6 signalling, although associated with lower markers of systemic inflammation, is unlikely to prevent incident CHD. There is some evidence that IL-6 signalling antagonism may reduce the risk of IS, but at the cost of higher risk of ICH (similar to the effect of LDL-lowering therapies⁵⁴) and future research should aim to perform a net-benefit analysis..

Table 6.1: Summary of trans-ethnic (East Asians and Europeans) and inter-instrument (IL6(R)) consistency of associations

Traits	Figure	Evidence of heterogeneity*	
		East Asians and Europeans	IL-6R and IL-6 instruments
CVD outcomes	Figure 4.6	Null	Null, except for CHD and MI
CVD risk factors	Figure 4.7	Null, except SBP and BMI	Null
Lipids	Figure 4.8	Null	Null

*5% FDR of meta-analysed estimates

Second, the association of both the IL-6(R) instruments with HDL-C and LDL-C are directionally consistent with reports from tocilizumab trials and suggest the HDL-C/LDL-C elevations may be an on-target effect of IL-6R blockade. Third, the thesis makes a general case for dual receptor-ligand analyses as standard practice, where appropriate, for therapeutic appraisal of drug targets. If associations of the IL-6 instrument with CHD/MI are real, it suggests that the effects of drugs that target IL-6R may not be readily portable to drugs that target the IL-6 ligand or, more generally, block the IL-6 signalling pathway. This means that investigators should not assume that the effects of drugs targeting the receptor as equivalent to targeting the ligand (or vice versa) and should perform in-depth receptor-ligand assessments, for example using human genetics. Additionally, many soluble receptors are known to be decoy receptors⁴⁴⁷, i.e. higher levels of these soluble receptors sequester the circulating ligand and, rather than activating, antagonizes the signalling pathway that typically operates when the ligand binds to the membrane-bound form of the receptor. This may not always be the case, and like soluble IL-6R, many soluble receptors may operate via pathways distinct from the ligand. The present work exemplifies one approach to appraise the associations of the soluble

receptors. Using genetic instruments that represent the circulating receptor and ligand enables a more comprehensive assessment of the likely efficacy and safety of pharmacologically modulating the biological pathways that operate via the receptor and ligand molecules.

6.3 Phenome-wide associations of interleukin-6 signalling

In addition to investigating the therapeutic relevance of IL-6 signalling for CVD, I leveraged the extensive phenotyping in CKB to investigate pre-specified non-vascular (i.e. for diseases, where the role of inflammation is known or suspected), proteomic, and phenome-wide screens to elucidate potential adverse effects or discover additional potential therapeutic signals with each instrument.

Among the pre-specified non-vascular outcomes, the IL-6R instrument was associated with a reduced risk of digestive cancer outcomes (driven largely by the association of the IL-6R instrument with oesophageal cancer) and a higher risk of tuberculosis. Like CHD/MI analysis in the previous chapter, these associations were not consistent with the IL-6 instrument which showed largely neutral associations with digestive cancer and tuberculosis. The restricted access to full summary statistics of digestive cancer traits in the European population and the lack of studies examining the association of IL-6R/IL-6 instruments with tuberculosis, limits the comparison of effects reported in my thesis. The association of the IL-6R instrument with a higher risk of tuberculosis is consistent with the immunological role of IL-6 signalling⁴¹⁸ and was also identified as a possible adverse effect of tocilizumab which was reported to re-activate latent tuberculosis⁴⁴⁸.

To investigate potential mechanisms, I assessed associations of the IL-6(R) instruments with clinical chemistry measurements (in addition to the clinical chemistry examined in the previous chapter) and, for a smaller subset of the study cohort, with 92 proteins part of OLINK's immuno-oncology panel. The IL-6R instrument was associated with lower plasma fibrinogen and lower plasma albumin concentrations,

whereas the IL-6 instrument was also associated with lower levels of fibrinogen but, unlike the IL-6R instrument, associated with higher plasma albumin concentrations. Albumin is a ‘negative’ acute phase protein and is inversely correlated with CRP concentrations³⁴⁶. The directionally discordant associations of both instruments with albumin suggest that the IL-6(R) instruments may represent different IL-6 signalling pathways and speaks to the relevance of dual receptor-ligand MR investigations as part of comprehensive pharmacological assessments of drugs targeting components of a signalling pathway. For example, in this case, the association of the IL-6 instrument (and not the IL-6R instrument) with albumin was consistent with observations in clinical trials of tocilizumab^{449,450}.

Additional proteomic and phenome-wide assessments did not yield any additional biological insights, indicating the need for larger studies for more reliable assessments in the future.

6.4 Research implications

Emerging evidence suggests that targeting inflammation pathways related to the IL-1 β signalling pathway is of therapeutic relevance for reducing the residual risk of CVD (residual risk refers to the CVD risk that remains despite treatment with lipid-lowering, antihypertensive, and anti-platelet agents)³⁰⁹. The IL-1 β signalling pathway regulates IL-6 signalling which contributes to the hepatic synthesis of CRP and fibrinogen. The CANTOS trial of ~10,000 patients with prior MI demonstrated a beneficial effect of therapeutic inhibition of IL-1 β on the primary endpoint of major vascular events (nonfatal myocardial infarction, nonfatal stroke, or cardiovascular death)³⁰³. Compared to placebo, canakinumab (the IL-1 β blocker tested in CANTOS) reduced levels of plasma IL-6 and CRP, with the magnitude of CVD risk reduction proportional to the reduction of CRP plasma concentrations^{258,451}. However, a separate assessment of CVD risk in study participants on canakinumab showed a residual CVD risk still remains and this risk was highest in those with the highest levels of IL-6⁴⁵¹. This suggested the

possibility that CVD risk can be lowered further by drugs that inhibit IL-6 signalling pathways.

Human genetic evidence that used genetic instruments which partially mimicked the effects of tocilizumab showed CVD risk reductions prompting the possibility of therapeutic repurposing tocilizumab or tocilizumab-like drugs for CVD^{96,97,286,385}. The work in my thesis builds on the previous human genetic evidence, by extending the IL-6 signalling assessment to a large cohort of East Asians, and, using separate genetic instruments to represent altered IL-6R and IL-6 signalling, showed that CHD risk reductions associated with the IL-6R instrument were not observed with the IL-6 instrument. While both instruments were constructed using genetic data from a GWAS of CRP in CKB, the IL-6 instrument was designed to mimic the effect of a drug that would specifically lower CRP concentrations via inhibition of the IL-6 protein (e.g. siltuximab), as opposed to reductions of CRP observed with alteration of the ratio of sIL-6R to mIL-6R (the effect of the IL-6R genetic instrument).

Aside from CHD, there were largely consistent associations between both instruments. Indeed, some IL-6 instrument associations with clinical chemistry measurements (e.g. albumin, Table 5.1) suggested that the IL-6 instrument may represent biological effects of tocilizumab better than the IL-6R instrument that was assumed in previous studies to represent effects of tocilizumab. Given the partial consistency of effects between IL-6R instrument and tocilizumab observed in the present and previous studies, it would be appropriate to state that the IL-6R instrument only partially mimics the effects of tocilizumab. A potential drug that would sufficiently represent the effect of the IL-6R instrument would be therapeutically beneficial not just for CVD, but potentially also for digestive cancer traits. However, given the null findings of the IL-6 instrument with CHD, it is unlikely that IL-6 signalling antagonists would be therapeutically useful to prevent incident CHD events. The results should guide the need for deeper pre-clinical research on the cardiovascular implications

of IL-6 signalling blockade before major investments are made to conduct large clinical trials.

The work in my thesis, therefore, provides a proof-of-concept that dual receptor-ligand assessments using human genetic evidence can provide important pre-trial data to guide investments for therapies that have the highest chance of success in well-powered clinical trials.

Appendices

A

APPENDIX

A.1 Supplemental contents for Chapter 3

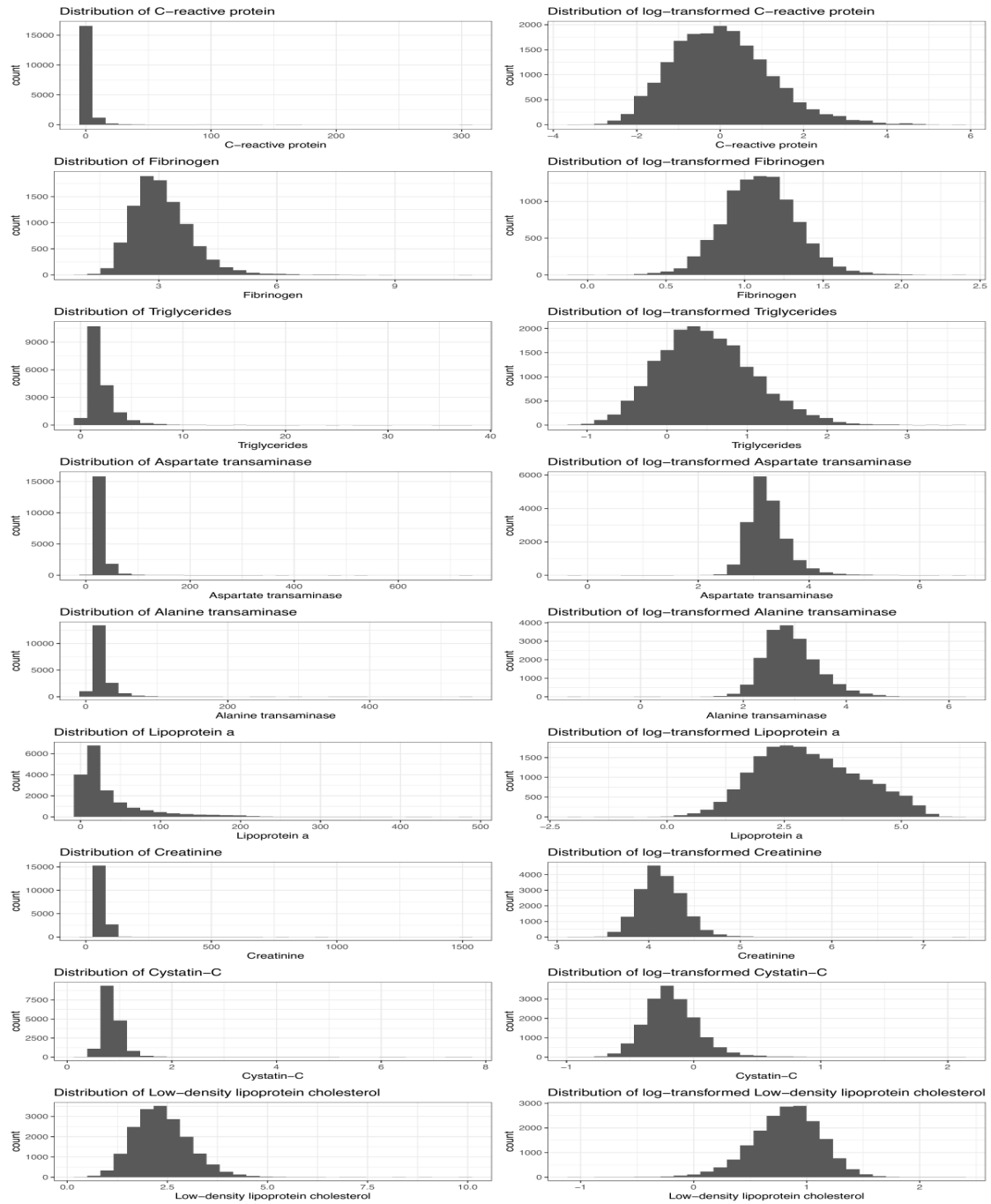


Figure A.1: Distributions of biomarkers studied in this thesis before and after log-transformation

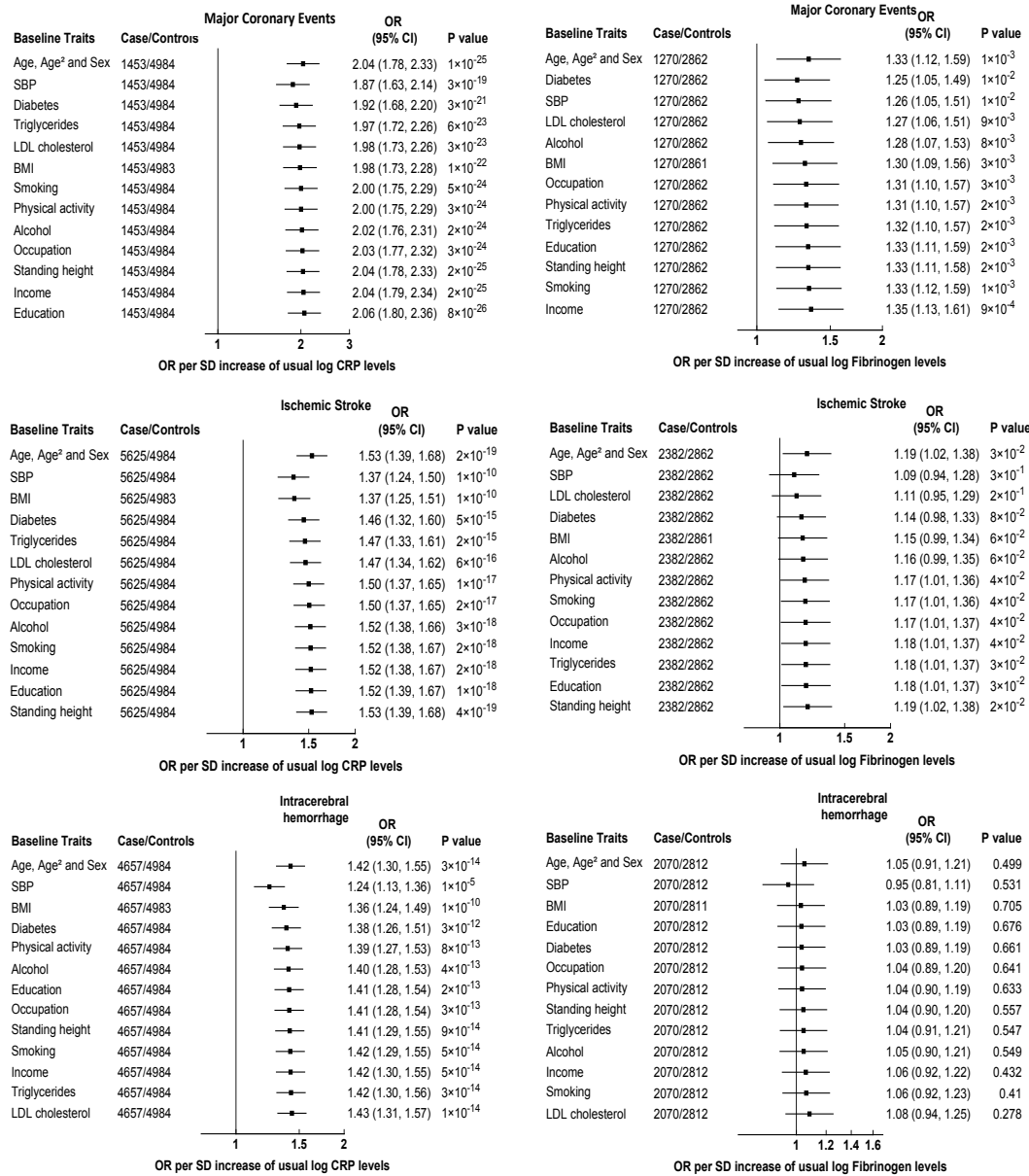


Figure A.2: Associations of CRP and fibrinogen with MCE, IS and ICH on adjustments for individual risk factors

Table A.1: Association of rs1880241-G on *IL6* gene expression across several tissues or cell lines in different eQTL studies curated from the Open Targets Genetics portal filtered at $p < 0.05$

Study	Tissue or cell line ¹	beta ²	se	pval
FAIRFAX_2014 [380]	Monocyte (LPS2)	-0.22	0.04	3.18E-07
QUACH_2016 [452]	Monocyte (R848)	-0.07	0.02	1.36E-03
GTEX_v7 [453]	Cells transformed fibroblasts	-0.12	0.04	7.32E-03
GTEX_v7 [453]	Spleen	-0.22	0.08	8.57E-03
NEDELEC_2016 [454]	Macrophage (Listeria)	-0.34	0.13	9.81E-03
GTEX_v7 [453]	Brain caudate basal ganglia	0.17	0.07	2.41E-02
NEDELEC_2016 [454]	Macrophage (Salmonella)	-0.20	0.09	2.82E-02
ALASOO_2018 [455]	Macrophage (Salmonella)	-0.11	0.05	3.05E-02
TWINSUK [456]	Lymphoblastoid cell lines	-0.17	0.08	3.44E-02
CEDAR [457]	Platelet	0.03	0.02	4.10E-02
TWINSUK [456]	Skin	-0.09	0.04	4.16E-02
GTEX_v7 [453]	Skeletal muscle	-0.09	0.04	4.20E-02
GTEX_v7 [453]	Tibial nerve	-0.09	0.04	4.46E-02

LPS2: Lipopolysaccharide 2; R848: Toll-like receptor 7/8 ligand

¹ beta represents effect per rs1880241-G allele² Included in parenthesis are agents or microbes used to stimulate gene expression in cells

A.2 Supplemental contents for Chapter 4

Table A.2: Power estimates for primary vascular endpoints for IL6R analyses

CVD Outcome	Case N	Controls N per Case	Power			
			OR = 0.95	OR = 0.90	OR = 0.85	OR = 0.80
MCE	6987	20	0.82	1.00	1.00	1.00
MI	3554	20	0.54	0.99	1.00	1.00
IS	16124	10	0.99	1.00	1.00	1.00
ICH	6757	20	0.81	1.00	1.00	1.00
Plaque	8427	2	0.75	1.00	1.00	1.00

MCE: Major Coronary Events; MI: Myocardial Infarction; IS: Ischaemic stroke; ICH: Intracerebral haemorrhage;

Table A.3: Power estimates for primary vascular endpoints for IL6 analyses

CVD Outcome	Case N	Controls N per Case	Power			
			OR = 0.95	OR = 0.90	OR = 0.85	OR = 0.80
MCE	5113	20	0.69	1.00	1.00	1.00
MI	2844	20	0.45	0.96	1.00	1.00
IS	11618	10	0.95	1.00	1.00	1.00
ICH	5998	20	0.76	1.00	1.00	1.00
Plaque	8383	2	0.75	1.00	1.00	1.00

MCE: Major Coronary Events; MI: Myocardial Infarction; IS: Ischaemic stroke; ICH: Intracerebral haemorrhage;

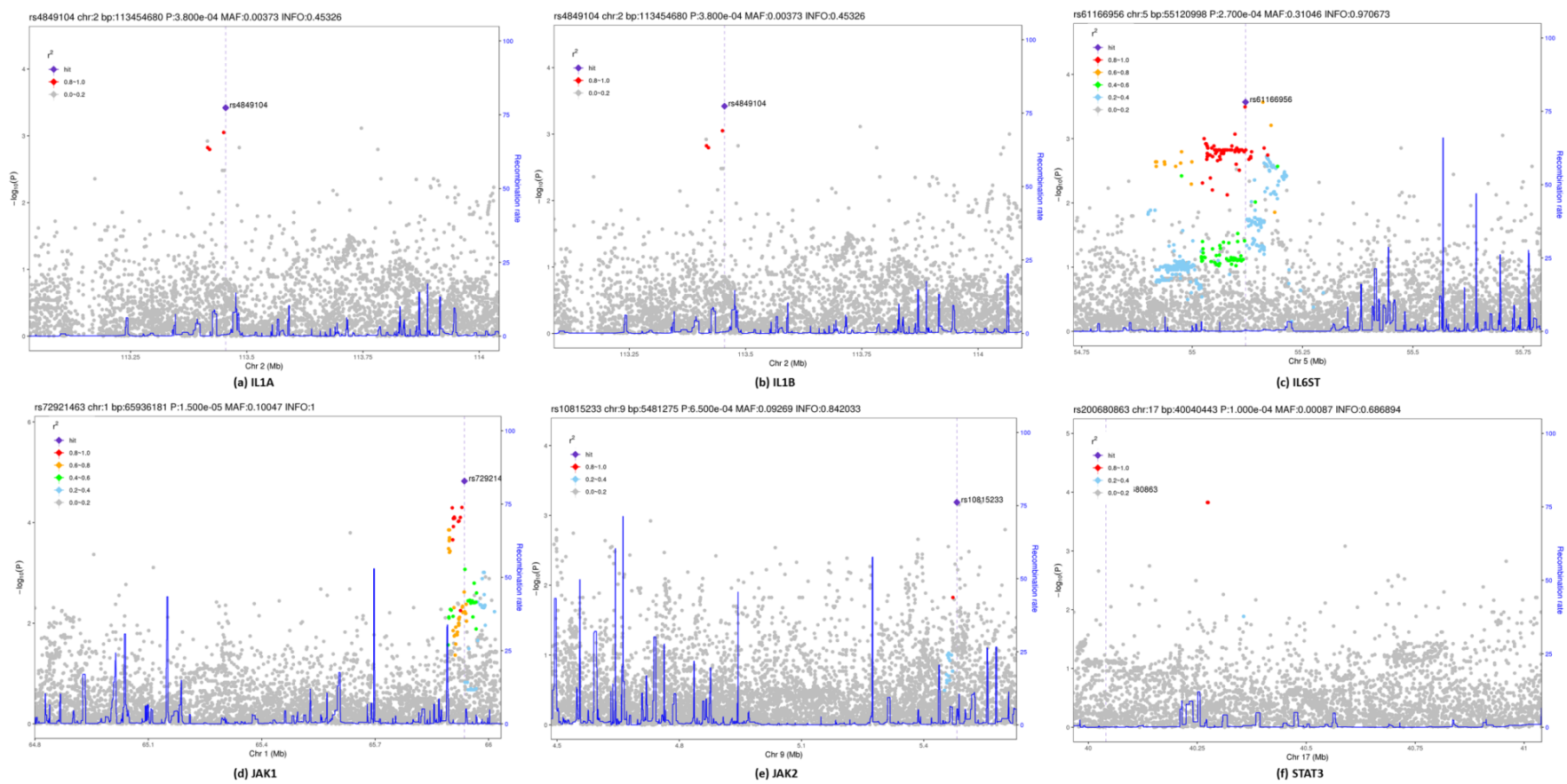


Figure A.3: LosusZoom plots of inflammation loci from CRP GWAS

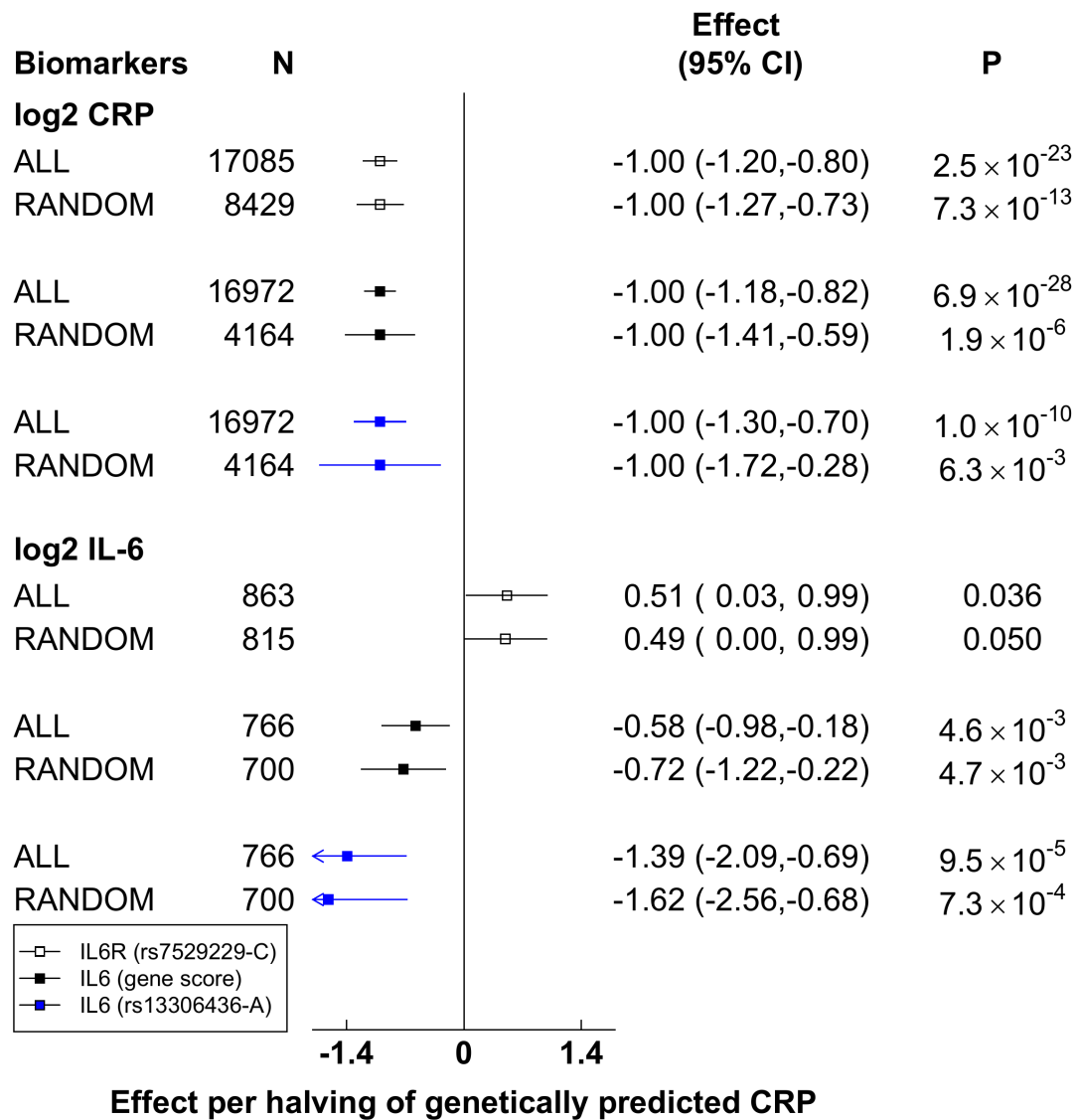
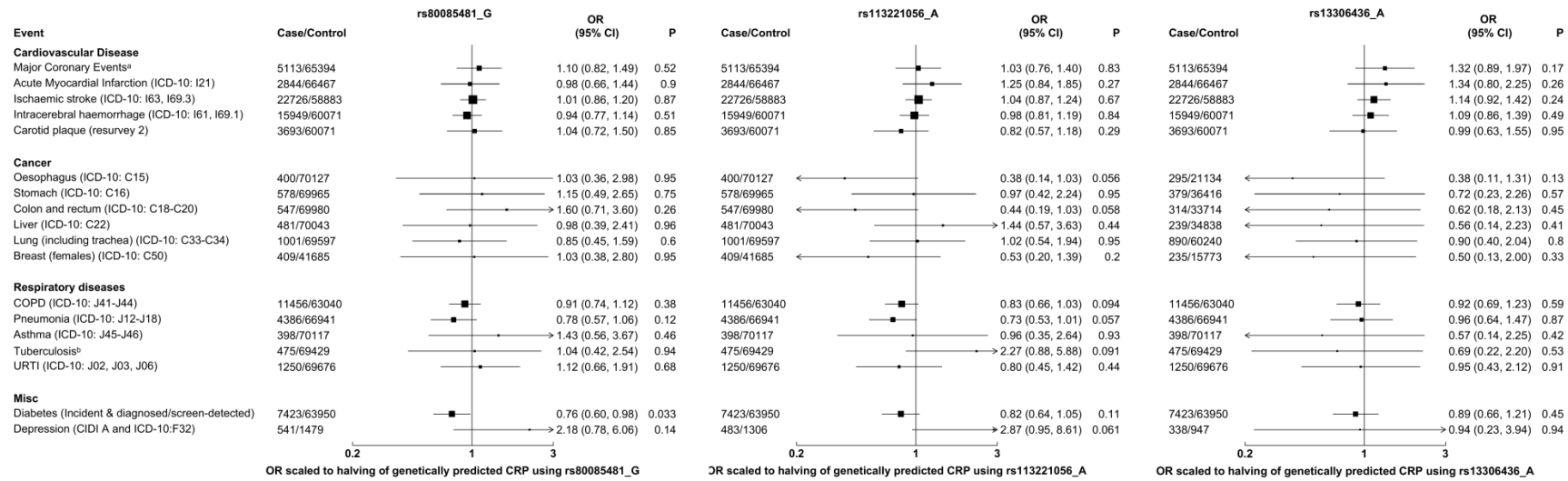


Figure A.4: Association of the IL-6(R) instruments and *IL6* regulatory SNP (rs13306436-A) with CRP and IL-6 in all and random subsets.

The figure compares associations of *IL6* and *IL6R* instruments between all individuals with genetic, biochemistry and proteomics data (model adjusted for age, age², sex and case ascertainment in each region and estimates meta-analysed across region using fixed effect method with inverse weighting) and individuals part of a random study population representative subset (model adjusted for age, age², and sex in each region and estimates meta-analysed across region using fixed effect method with inverse weighting). Error bars represent 95% CI. Area of boxes are not proportional to inverse of the variance of effect.

Figure A.5: Association of *IL6* SNPs with vascular and non-vascular outcomes

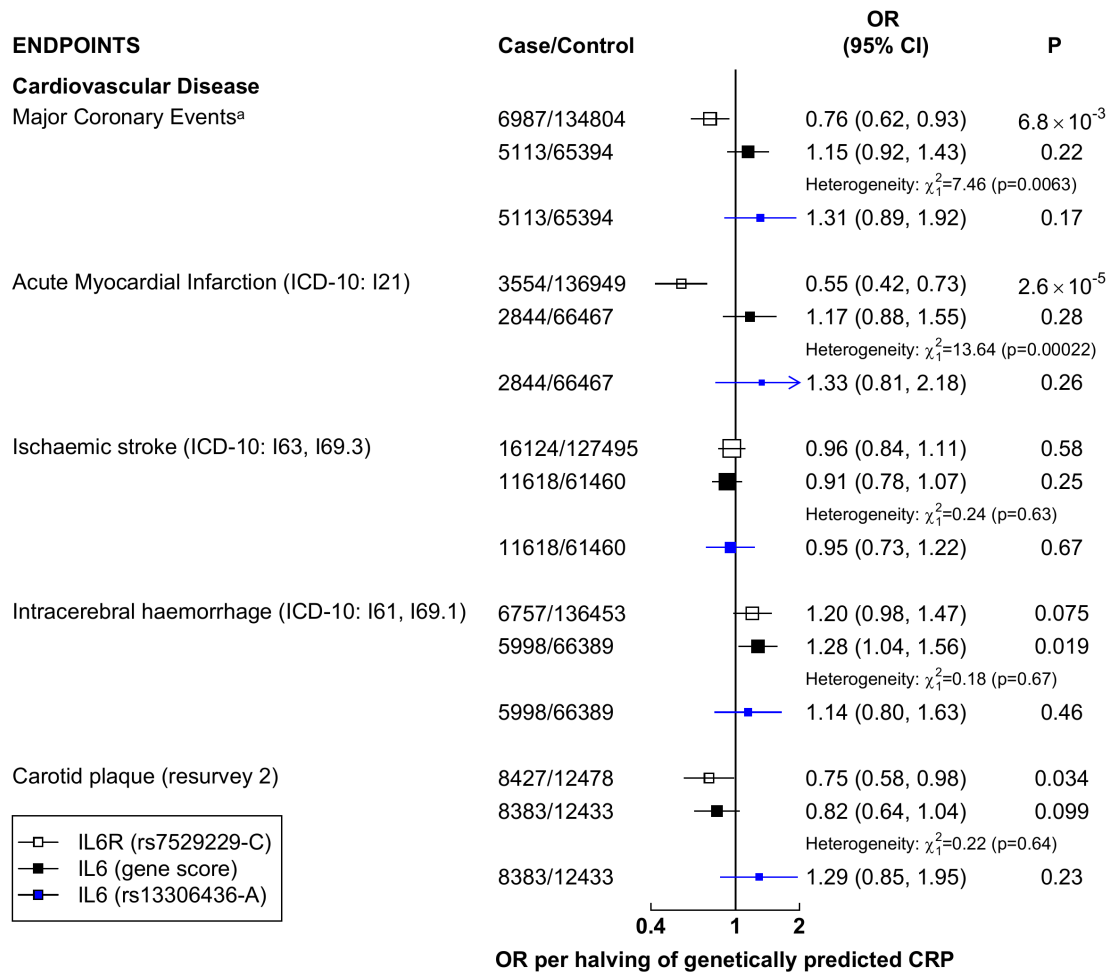
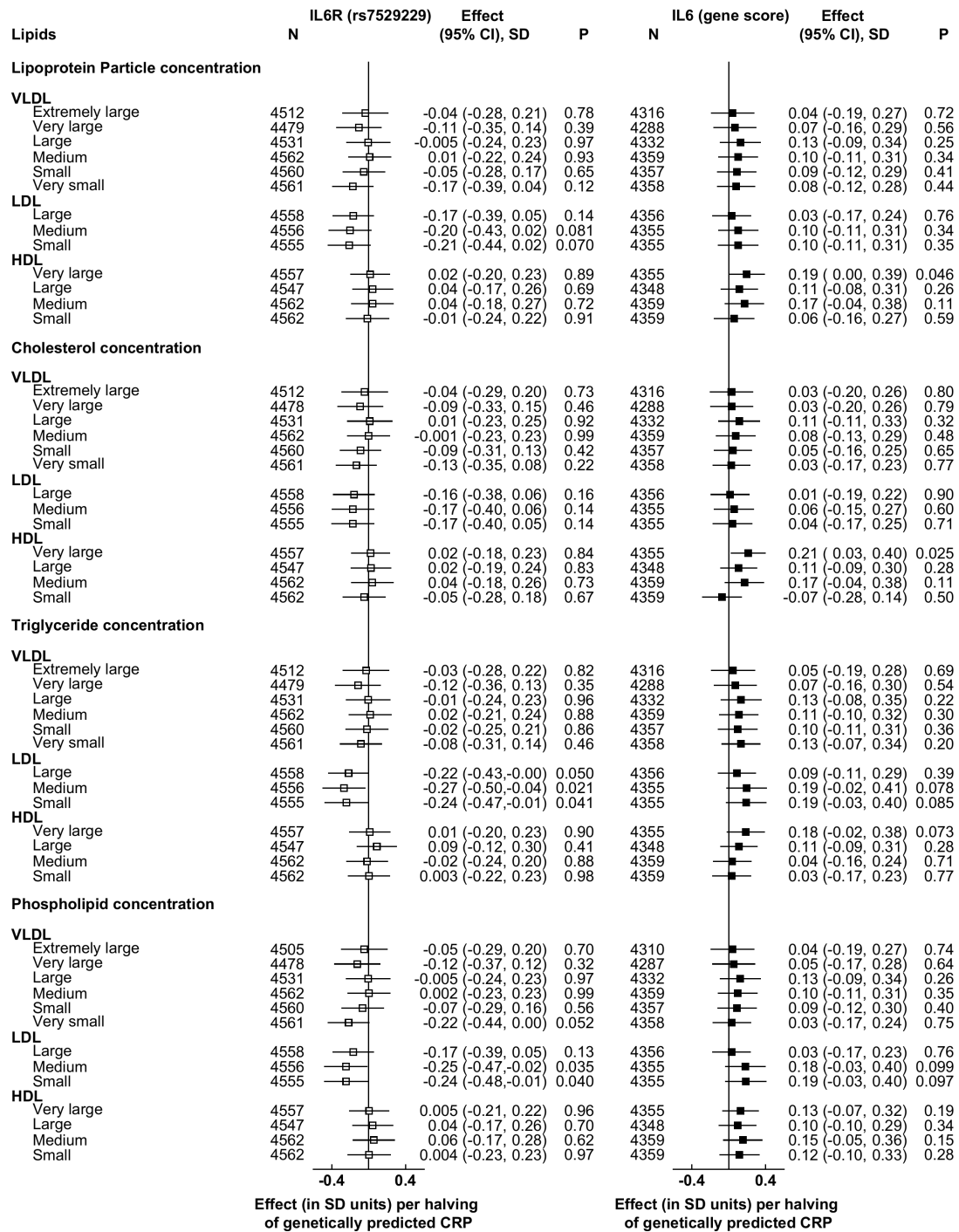


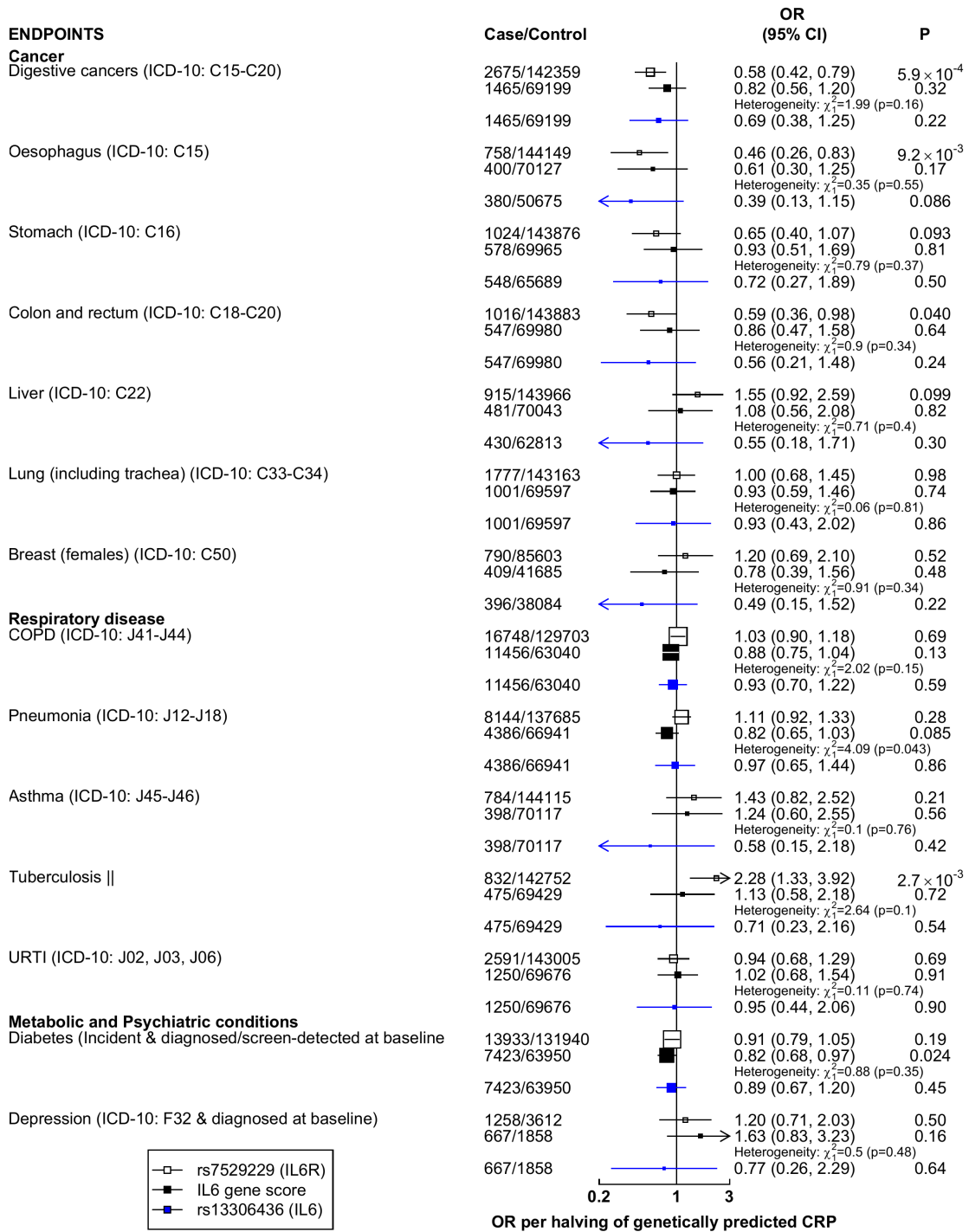
Figure A.6: Association of *IL6* SNP rs13306436 with CVD traits and outcomes

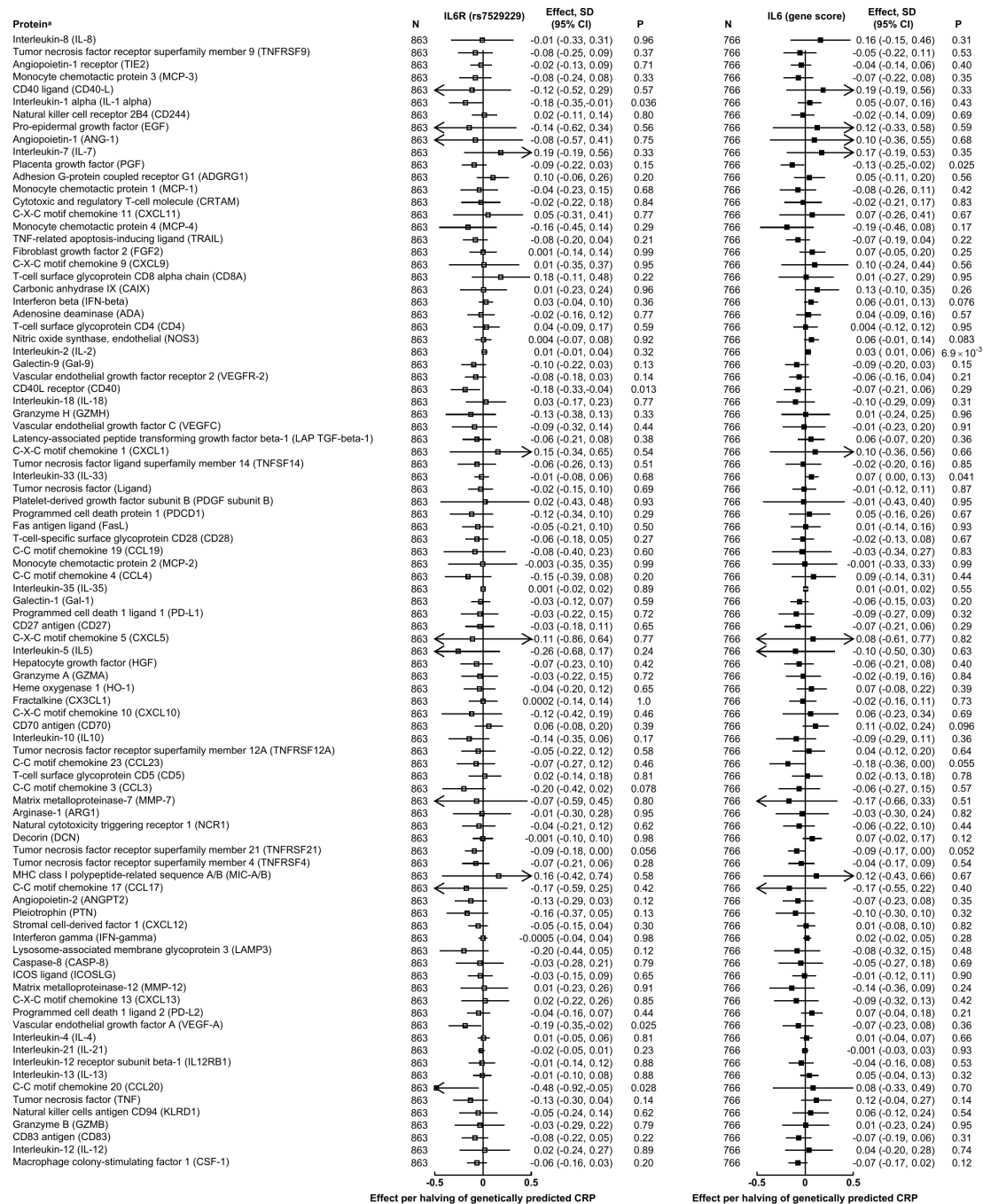
Figure A.7: Associations of *IL6R*, *IL6* gene score with lipidomic traits

A.3 Supplemental contents for Chapter 5

Table A.4: Power estimates for pre-specified non-vascular endpoints with an $\alpha = 0.05/10$ (10 is the number of secondary non-vascular endpoints)

Case N	Power			
	OR = 0.95	OR = 0.90	OR = 0.85	OR = 0.80
500	0.02	0.11	0.35	0.69
1000	0.04	0.28	0.73	0.97
5000	0.35	0.98	1.00	1.00
15000	0.92	1.00	1.00	1.00

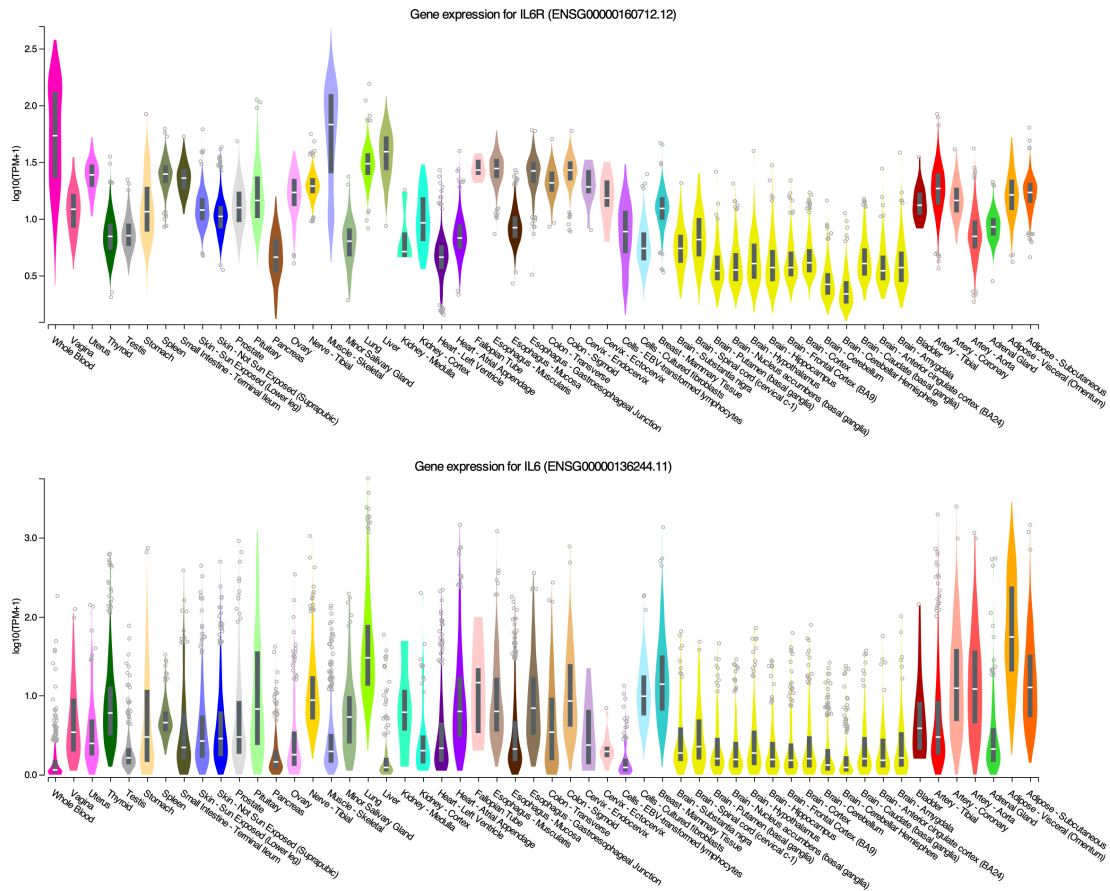
Figure A.8: Association of *IL6* SNP rs13306436 with non-vascular outcomes



Notes: Error bars represent 95% CI. Area of boxes are not proportional to inverse of the variance of effect.

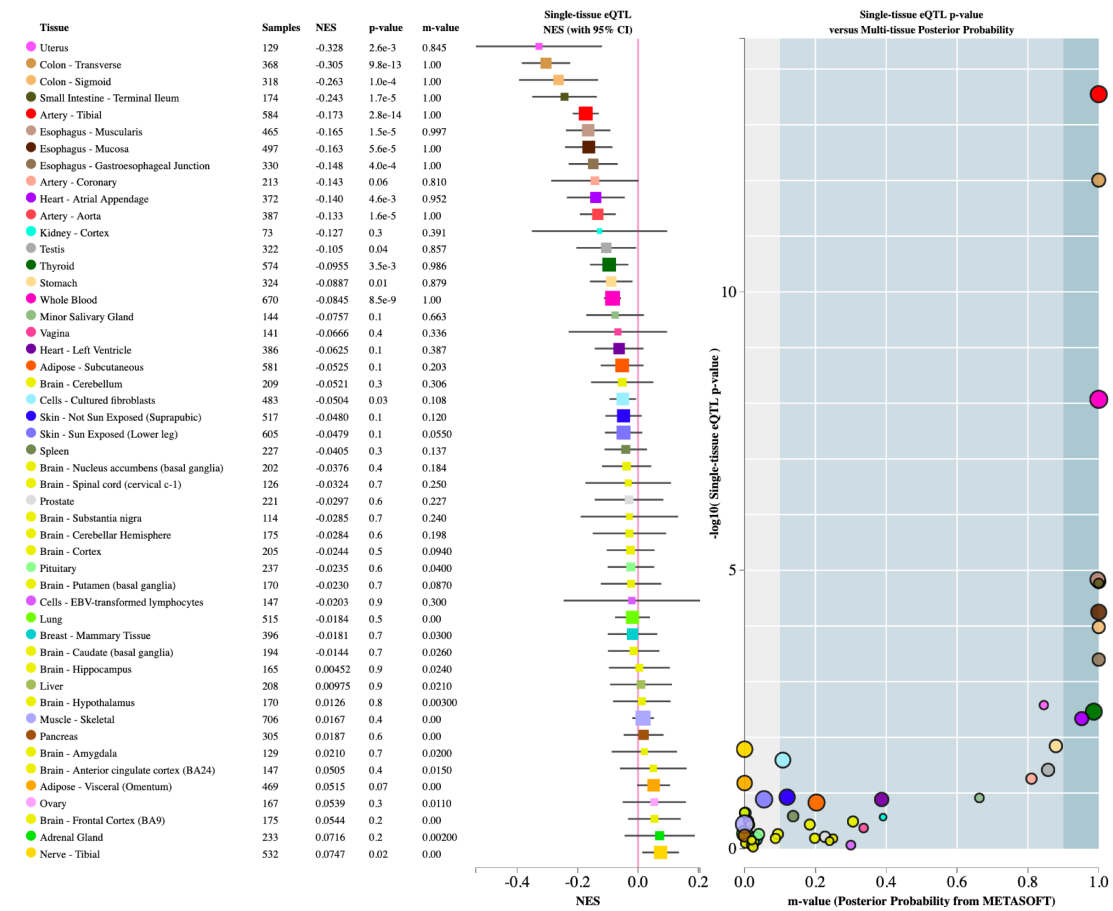
*Proteins were log-transformed prior to standardization

Figure A.9: Proteome-wide associations of the IL-6(R) instruments



TPM: Transcripts per million

Figure A.10: *IL6R* and *IL6* gene expression across post-mortem tissues (listed alphabetically in reverse). Please note, because these were extracted directly from the GTEx v8 web interface, the y-axes are on different intervals of the same scale, $\log(\text{TPM}+1)$. The figure provides a qualitative measure of relative expression of IL-6R and IL-6 transcripts across different tissues.



NES (Normalized effect size) is defined as the slope of the linear regression, and is computed as the effect of the alternative allele (ALT) on gene expression.

m-value is the posterior probability that an eQTL effect exists in each tissue tested in the cross-tissue meta-analysis, and ranges from 0 (tissue is predicted to not have an eQTL effect) to 1 (tissue is predicted to have an eQTL effect)

p-value is derived from a t-test that tests the null hypothesis of no effect of rs7529229-C on *IL6R* gene expression in a given tissue.

Figure A.11: Effect per rs7529229-C on *IL6R* gene expression across tissues in GTEx v8 arranged by effect size. The figure shows that some of the strongest evidence of association of rs7529229-C was on *IL6R* gene expression in arterial and digestive tract tissues.

REFERENCES

1. Abbafati, C., Abbas, K. M. & Abbasi-Kangevari, M. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet* **396**, 1204–1222 (Oct. 2020).
2. WHO. Cardiovascular diseases (CVDs) fact sheets. *Who*, 1 (2016).
3. Holligan, S. D., Berryman, C. E., Wang, L., Flock, M. R., Harris, K. A., *et al.* Atherosclerotic Cardiovascular Disease. *Present Knowledge in Nutrition: Tenth Edition*, 745–805 (2012).
4. Sanchis-Gomar, F., Perez-Quilis, C., Leischik, R. & Lucia, A. Epidemiology of coronary heart disease and acute coronary syndrome. *Annals of Translational Medicine* **4**, 13 (2016).
5. Overbaugh, K. J. Acute coronary syndrome. *American Journal of Nursing* **109**, 42–52 (2009).
6. Ford, T. J., Corcoran, D. & Berry, C. Stable coronary syndromes: Pathophysiology, diagnostic advances and therapeutic need. *Heart* **104**, 284–292 (2018).
7. Noel Bairey Merz, C., Pepine, C. J., Walsh, M. N. & Fleg, J. L. Ischemia and No Obstructive Coronary Artery Disease (INOCA): Developing Evidence-Based Therapies and Research Agenda for the Next Decade. *Circulation* **135**, 1075–1092 (2017).
8. Abu Daya, H. & Hage, F. G. Guidelines in review: ACC/AATS/AHA/ASE/ASNC/SCAI/SCCT/STS 2017 appropriate use criteria for coronary revascularization in patients with stable ischemic heart disease. *Journal of Nuclear Cardiology* **24**, 1793–1799 (2017).
9. Frequency of negative coronary arteriographic findings in patients with chest pain is related to community practice patterns. *American Journal of Managed Care* **8**, 643–648 (2002).
10. Jespersen, L., Hvelplund, A., Abildstrøm, S. Z., Pedersen, F., Galatius, S., *et al.* Stable angina pectoris with no obstructive coronary artery disease is associated with increased risks of major adverse cardiovascular events. *European Heart Journal* **33**, 734–744 (2012).
11. Hasdai, D., Holmes, D. R., Higano, S. T., Burnett, J. C. & Lerman, A. Prevalence of coronary blood flow reserve abnormalities among patients with nonobstructive coronary artery disease and chest pain. *Mayo Clinic Proceedings* **73**, 1133–1140 (1998).
12. Aho, K. *et al.* Cerebrovascular disease in the community: results of a WHO Collaborative Study. *Bulletin of the World Health Organization* **58**, 113–130 (1980).
13. Bhalla, A., Wang, Y., Rudd, A. & Wolfe, C. D. Differences in outcome and predictors between ischemic and intracerebral hemorrhage: The South London Stroke Register. *Stroke* **44**, 2174–2181 (2013).

14. Mittal, S. H. & Goel, D. Mortality in ischemic stroke score: A predictive score of mortality for acute ischemic stroke. *Brain Circulation* **3**, 29–34 (2017).
15. An, S. J., Kim, T. J. & Yoon, B. W. Epidemiology, risk factors, and clinical features of intracerebral hemorrhage: An update. *Journal of Stroke* **19**, 3–10 (2017).
16. Hop, J. W., Rinkel, G. J., Algra, A. & Van Gijn, J. Case-fatality rates and functional outcome after subarachnoid hemorrhage: A systematic review. *Stroke* **28**, 660–664 (1997).
17. Love, B. B. & Bendixen, B. H. Classification of subtype of acute ischemic stroke definitions for use in a multicenter clinical trial. *Stroke* **24**, 35–41 (1993).
18. Kamel, H. & Healey, J. S. Cardioembolic Stroke. *Circulation Research* **120**, 514–526 (2017).
19. Chaturvedi, S. & Bhattacharya, P. Large artery atherosclerosis: Carotid stenosis, vertebral artery disease, and intracranial atherosclerosis. *CONTINUUM Lifelong Learning in Neurology* **20**, 323–334 (2014).
20. Kannel, W. B. & Vasan, R. S. Is Age Really a Non-Modifiable Cardiovascular Risk Factor? *American Journal of Cardiology* **104**, 1307–1310 (2009).
21. Lloyd-Jones, D. M., Leip, E. P., Larson, M. G., D’Agostino, R. B., Beiser, A., *et al.* Prediction of lifetime risk for cardiovascular disease by risk factor burden at 50 years of age. *Circulation* **113**, 791–798 (2006).
22. Jousilahti, P., Vartiainen, E., Tuomilehto, J. & Puska, P. Sex, age, cardiovascular risk factors, and coronary heart disease: A prospective follow-up study of 14 786 middle-aged men and women in Finland. *Circulation* **99**, 1165–1172 (1999).
23. Bots, S. H., Peters, S. A. & Woodward, M. Sex differences in coronary heart disease and stroke mortality: A global assessment of the effect of ageing between 1980 and 2010. *BMJ Global Health* **2**, 2 (2017).
24. Leening, M. J., Ferket, B. S., Steyerberg, E. W., Kavousi, M., Deckers, J. W., *et al.* Sex differences in lifetime risk and first manifestation of cardiovascular disease: Prospective population based cohort study. *BMJ (Online)* **349** (2014).
25. Lloyd-Jones, D. M., Larson, M. G., Beiser, A. & Levy, D. Lifetime risk of developing coronary heart disease. *Lancet* **353**, 89–92 (1999).
26. Barrett-Connor, E. Sex differences in coronary heart disease: Why are women so superior? The 1995 Ancel Keys Lecture. *Circulation* **95**, 252–264 (1997).
27. Vaidya, D., Becker, D. M., Bittner, V., Mathias, R. A. & Ouyang, P. Ageing, menopause, and ischaemic heart disease mortality in England, Wales, and the United States: Modelling study of national mortality data. *BMJ (Online)* **343** (2011).
28. Lang, T., Lepage, B., Schieber, A.-C., Lamy, S. & Kelly-Irving, M. Social Determinants of Cardiovascular Diseases. *Public Health Reviews* **33**, 601–622 (2011).
29. Wild, S. & McKeigue, P. Cross sectional analysis of mortality by country of birth in England and Wales, 1970–92. *Bmj* **314**, 705 (1997).
30. Tillin, T., Hughes, A. D., Mayet, J., Whincup, P., Sattar, N., *et al.* The relationship between metabolic risk factors and incident cardiovascular disease in Europeans, South Asians, and African Caribbeans: SABRE (Southall and Brent Revisited) - A prospective population-based study. *Journal of the American College of Cardiology* **61**, 1777–1786 (2013).

31. Feder, G., Crook, A. M., Magee, P., Banerjee, S., Timmis, A. D., *et al.* Ethnic differences in invasive management of coronary disease: Prospective cohort study of patients undergoing angiography. *British Medical Journal* **324**, 511–516 (2002).
32. McKeigue, P. M., Shah, B. & Marmot, M. G. Relation of central obesity and insulin resistance with high diabetes prevalence and cardiovascular risk in South Asians. *The Lancet* **337**, 382–386 (1991).
33. Patel, D. J., Winterbotham, M., Britt, R. P., Sutton, G. C., Bhatnagar, D., *et al.* Coronary risk factors in people from the Indian subcontinent living in West London and their siblings in India. *The Lancet* **345**, 405–409 (1995).
34. Chambers, J. C., Eda, S., Bassett, P., Karim, Y., Thompson, S. G., *et al.* C-reactive protein, insulin resistance, central obesity, and coronary heart disease risk in Indian Asians from the United Kingdom compared with European whites. *Circulation* **104**, 145–150 (2001).
35. Kain, K., Catto, A. J. & Grant, P. J. Impaired fibrinolysis and increased fibrinogen levels in South Asian subjects. *Atherosclerosis* **156**, 457–461 (2001).
36. Sirugo, G., Williams, S. M. & Tishkoff, S. A. The Missing Diversity in Human Genetic Studies. *Cell* **177**, 26–31 (2019).
37. Manolio, T. A. Using the Data We Have: Improving Diversity in Genomic Research. *American Journal of Human Genetics* **105**, 233–236 (2019).
38. Deaton, C., Froelicher, E. S., Wu, L. H., Ho, C., Shishani, K., *et al.* The global burden of cardiovascular disease. *European Journal of Cardiovascular Nursing* **10** (2011).
39. Lewington, S., Clarke, R., Qizilbash, N., Peto, R. & Collins, R. Age-specific relevance of usual blood pressure to vascular mortality: A meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet* **360**, 1903–1913 (2002).
40. Rapsomaniki, E., Timmis, A., George, J., Pujades-Rodriguez, M., Shah, A. D., *et al.* Blood pressure and incidence of twelve cardiovascular diseases: Lifetime risks, healthy life-years lost, and age-specific associations in 1 · 25 million people. *The Lancet* **383**, 1899–1911 (2014).
41. Lacey, B., Lewington, S., Clarke, R., Kong, X. L., Chen, Y., *et al.* Age-specific association between blood pressure and vascular and non-vascular chronic diseases in 0 · 5 million adults in China: a prospective cohort study. *The Lancet Global Health* **6**, e641–e649 (2018).
42. Ettehad, D., Emdin, C. A., Kiran, A., Anderson, S. G., Callender, T., *et al.* Blood pressure lowering for prevention of cardiovascular disease and death: A systematic review and meta-analysis. *The Lancet* **387**, 957–967 (2016).
43. World Health Organization. Fact Sheet Obesity and overweight. *World Health Organization*, 1–5 (2011).
44. Cornier, M. A., Després, J. P., Davis, N., Grossniklaus, D. A., Klein, S., *et al.* Assessing adiposity: A scientific statement from the american heart association. *Circulation* **124**, 1996–2019 (2011).
45. MacMahon, S., Baigent, C., Duffy, S., Rodgers, A., Tominaga, S., *et al.* Body-mass index and cause-specific mortality in 900 000 adults: Collaborative analyses of 57 prospective studies. *The Lancet* **373**, 1083–1096 (2009).
46. Chen, Z., Yang, G., Zhou, M., Smith, M., Offer, A., *et al.* Body mass index and mortality from ischaemic heart disease in a lean population: 10 year prospective

- study of 220 000 adult men. *International Journal of Epidemiology* **35**, 141–150 (2006).
47. Chen, Z., Iona, A., Parish, S., Chen, Y., Guo, Y., *et al.* Adiposity and risk of ischaemic and haemorrhagic stroke in 0·5 million Chinese men and women: a prospective cohort study. *The Lancet Global Health* **6**, e630–e640 (2018).
 48. KANNEL, W. B., DAWBER, T. R., FRIEDMAN, G. D., GLENNON, W. E. & MCNAMARA, P. M. Risk Factors in Coronary Heart Disease. an Evaluation of Several Serum Lipids As Predictors of Coronary Heart Disease; the Framingham Study. *Annals of internal medicine* **61**, 888–899 (1964).
 49. MacMahon, S., Duffy, S., Rodgers, A., Tominaga, S., Chambless, L., *et al.* Blood cholesterol and vascular mortality by age, sex, and blood pressure: A meta-analysis of individual data from 61 prospective studies with 55 000 vascular deaths. *The Lancet* **370**, 1829–1839 (2007).
 50. Ference, B. A., Ginsberg, H. N., Graham, I., Ray, K. K., Packard, C. J., *et al.* Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. Evidence from genetic, epidemiologic, and clinical studies. A consensus statement from the European Atherosclerosis Society Consensus Panel. *European Heart Journal* **38**, 2459–2472 (2017).
 51. Holmes, M. V., Asselbergs, F. W., Palmer, T. M., Drenos, F., Lanktree, M. B., *et al.* Mendelian randomization of blood lipids for coronary heart disease. *European Heart Journal* **36**, 539–550 (2015).
 52. Nissen, S. E., Tuzcu, E. M., Schoenhagen, P., Crowe, T., Sasiela, W. J., *et al.* Statin Therapy, LDL Cholesterol, C-Reactive Protein, and Coronary Artery Disease. *New England Journal of Medicine* **352**, 29–38 (2005).
 53. Collins, R., Reith, C., Emberson, J., Armitage, J., Baigent, C., *et al.* Interpretation of the evidence for the efficacy and safety of statin therapy. *The Lancet* **388**, 2532–2561 (2016).
 54. Sun, L., Clarke, R., Bennett, D., Guo, Y., Walters, R. G., *et al.* Causal associations of blood lipids with risk of ischemic stroke and intracerebral hemorrhage in Chinese adults. *Nature Medicine* **25**, 569–574 (2019).
 55. Di Angelantonio, E., Sarwar, N., Perry, P., Kaptoge, S., Ray, K. K., *et al.* Major lipids, apolipoproteins, and risk of vascular disease. *JAMA - Journal of the American Medical Association* **302**, 1993–2000 (2009).
 56. Hindy, G., Engström, G., Larsson, S. C., Traylor, M., Markus, H. S., *et al.* Role of blood lipids in the development of ischemic stroke and its subtypes: A mendelian randomization study. *Stroke* **49**, 820–827 (2018).
 57. Park, C., Guallar, E., Linton, J. A., Lee, D. C., Jang, Y., *et al.* Fasting glucose level and the risk of incident atherosclerotic cardiovascular diseases. *Diabetes Care* **36**, 1988–1993 (2013).
 58. Rodgers, A. Blood glucose and risk of cardiovascular disease in the Asia Pacific region. *Diabetes Care* **27**, 2836–2842 (2004).
 59. Sarwar, N., Gao, P., Kondapally Seshasai, S. R., Gobin, R., Kaptoge, S., *et al.* Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: A collaborative meta-analysis of 102 prospective studies. *The Lancet* **375**, 2215–2222 (2010).
 60. Einarson, T. R., MacHado, M. & Henk Hemels, M. E. Blood glucose and subsequent cardiovascular disease: Update of a meta-analysis. *Current Medical Research and Opinion* **27**, 2155–2163 (2011).

61. Bragg, F., Li, L., Smith, M., Guo, Y., Chen, Y., *et al.* Associations of blood glucose and prevalent diabetes with risk of cardiovascular disease in 500 000 adult Chinese: The China Kadoorie Biobank. *Diabetic Medicine* **31**, 540–551 (2014).
62. Bragg, F., Li, L., Bennett, D., Guo, Y., Lewington, S., *et al.* Association of random plasma glucose levels with the risk for cardiovascular disease among Chinese adults without known diabetes. *JAMA Cardiology* **1**, 813–823 (2016).
63. Jung, H., Joo, J., Jeon, Y., Lee, J., In, J., *et al.* Advanced glycation end products downregulate glucokinase in mice. *Diabetes/Metabolism Research and Reviews* **27**, 557–563 (2011).
64. Zhao, Z., Zhao, C., Xu, H. Z., Zheng, F., Cai, W., *et al.* Advanced glycation end products inhibit glucose-stimulated insulin secretion through nitric oxide-dependent inhibition of cytochrome c oxidase and adenosine triphosphate synthesis. *Endocrinology* **150**, 2569–2576 (2009).
65. Wright, E., Scism-Bacon, J. L. & Glass, L. C. Oxidative stress in type 2 diabetes: The role of fasting and postprandial glycaemia. *International Journal of Clinical Practice* **60**, 308–314 (2006).
66. Hegab, Z. Role of advanced glycation end products in cardiovascular disease. *World Journal of Cardiology* **4**, 90 (2012).
67. Vlassara, H. & Uribarri, J. Advanced glycation end products (AGE) and diabetes: Cause, effect, or both? *Current Diabetes Reports* **14** (2014).
68. Burns, D. M. Epidemiology of smoking-induced cardiovascular disease. *Progress in Cardiovascular Diseases* **46**, 11–29 (2003).
69. Doll, R., Peto, R., Boreham, J. & Sutherland, I. Mortality in relation to smoking: 50 Years' observations on male British doctors. *British Medical Journal* **328**, 1519–1528 (2004).
70. Huxley, R. R. & Woodward, M. Cigarette smoking as a risk factor for coronary heart disease in women compared with men: A systematic review and meta-analysis of prospective cohort studies. *The Lancet* **378**, 1297–1305 (2011).
71. Jha, P. & Peto, R. Global Effects of Smoking, of Quitting, and of Taxing Tobacco. *New England Journal of Medicine* **370**, 60–68 (2014).
72. Messner, B. & Bernhard, D. Smoking and cardiovascular disease: Mechanisms of endothelial dysfunction and early atherogenesis. *Arteriosclerosis, Thrombosis, and Vascular Biology* **34**, 509–515 (2014).
73. Barnoya, J. & Glantz, S. A. Cardiovascular effects of secondhand smoke: Nearly as large as smoking. *Circulation* **111**, 2684–2698 (2005).
74. U.S.DHSS. *How Tobacco Smoke Causes Disease The Biology and Behavioral Basis for Smoking-Attributable Disease A Report of the Surgeon General* (2010).
75. Holmes, M. V., Dale, C. E., Zuccolo, L., Silverwood, R. J., Guo, Y., *et al.* Association between alcohol and cardiovascular disease: Mendelian randomisation analysis based on individual participant data. *BMJ (Online)* **349** (2014).
76. Millwood, I. Y., Walters, R. G., Mei, X. W., Guo, Y., Yang, L., *et al.* Conventional and genetic evidence on alcohol and vascular disease aetiology: a prospective study of 500 000 men and women in China. *The Lancet* **393**, 1831–1842 (2019).
77. Wood, A. M., Kaptoge, S., Butterworth, A., Nietert, P. J., Warnakula, S., *et al.* Risk thresholds for alcohol consumption: combined analysis of individual-participant data for 599 912 current drinkers in 83 prospective studies. *The Lancet* **391**, 1513–1523 (2018).

78. Abbafati, C., Abbas, K. M., Abbasi-Kangevari, M., Abd-Allah, F., Abdelalim, A., *et al.* Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet* **396**, 1223–1249 (2020).
79. Afshin, A., Sur, P. J., Fay, K. A., Cornaby, L., Ferrara, G., *et al.* Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet* **393**, 1958–1972 (2019).
80. Christ, A., Günther, P., Lauterbach, M. A., Duewell, P., Biswas, D., *et al.* Western Diet Triggers NLRP3-Dependent Innate Immune Reprogramming. *Cell* **172**, 162–175.e14 (2018).
81. Du, H., Li, L., Bennett, D., Guo, Y., Key, T. J., *et al.* Fresh Fruit Consumption and Major Cardiovascular Disease in China. *New England Journal of Medicine* **374**, 1332–1343 (2016).
82. Estruch, R., Ros, E., Salas-Salvadó, J., Covas, M.-I., Corella, D., *et al.* Primary Prevention of Cardiovascular Disease with a Mediterranean Diet Supplemented with Extra-Virgin Olive Oil or Nuts. *New England Journal of Medicine* **378**, e34 (2018).
83. Key, T. J., Appleby, P. N., Bradbury, K. E., Sweeting, M., Wood, A., *et al.* Consumption of Meat, Fish, Dairy Products, and Eggs and Risk of Ischemic Heart Disease: A Prospective Study of 7198 Incident Cases among 409 885 Participants in the Pan-European EPIC Cohort. *Circulation* **139**, 2835–2845 (2019).
84. Landrigan, P. J., Fuller, R., Acosta, N. J., Adeyi, O., Arnold, R., *et al.* The Lancet Commission on pollution and health. *The Lancet* **391**, 462–512 (2018).
85. Mitter, S. S., Vedanthan, R., Islami, F., Pourshams, A., Khademi, H., *et al.* Household fuel use and cardiovascular disease mortality. *Circulation* **133**, 2360–2369 (2016).
86. Yu, K., Qiu, G., Chan, K. H., Lam, K. B. H., Kurmi, O. P., *et al.* Association of solid fuel use with risk of cardiovascular and all-cause mortality in rural China. *JAMA - Journal of the American Medical Association* **319**, 1351–1361 (2018).
87. Bartelt, L. A., Bolick, D. T. & Guerrant, R. L. Disentangling Microbial Mediators of Malnutrition: Modeling Environmental Enteric Dysfunction. *Cmgh* **7**, 692–707 (2019).
88. Guerrant, R. L., Deboer, M. D., Moore, S. R., Scharf, R. J. & Lima, A. A. The impoverished gut - A triple burden of diarrhoea, stunting and chronic disease. *Nature Reviews Gastroenterology and Hepatology* **10**, 220–229 (2013).
89. Deboer, M. D., Lima, A. A., Oría, R. B., Scharf, R. J., Moore, S. R., *et al.* Early childhood growth failure and the developmental origins of adult disease: Do enteric infections and malnutrition increase risk for the metabolic syndrome? *Nutrition Reviews* **70**, 642–653 (2012).
90. Lee, G. O., Olortegui, M. P., Salas, M. S., Yori, P. P., Trigoso, D. R., *et al.* Environmental enteropathy is associated with cardiometabolic risk factors in Peruvian children. *Journal of Developmental Origins of Health and Disease* **8**, 337–348 (2017).
91. Ryan, G. B. & Majno, G. *Acute inflammation* **1**, 183–276 (Churchill Livingstone, 1977).
92. Kaptoge, S., Di Angelantonio, E., Lowe, G., Pepys, M. B., Thompson, S. G., *et al.* C-reactive protein concentration and risk of coronary heart disease, stroke, and

- mortality: An individual participant meta-analysis. *The Lancet* **375**, 132–140 (2010).
93. Danesh, J., Lewington, S., Thompson, S. G., Lowe, G. D., Collins, R., *et al.* Plasma fibrinogen level and the risk of major cardiovascular diseases and nonvascular mortality: An individual participant meta-analysis. *Journal of the American Medical Association* **294**, 1799–1809 (2005).
94. Eiriksdottir, G., Harris, T. B., Launer, L. J., Gudnason, V., Folsom, A. R., *et al.* Association between C reactive protein and coronary heart disease: Mendelian randomisation analysis based on individual participant data. *Bmj* **342**, 425 (2011).
95. Keavney, B., Danesh, J., Parish, S., Palmer, A., Clark, S., *et al.* Fibrinogen and coronary heart disease: Test of causality by 'Mendelian randomization'. *International Journal of Epidemiology* **35**, 935–943 (2006).
96. Sarwar, N., Butterworth, A. S., Freitag, D. F., Gregson, J., Willeit, P., *et al.* Interleukin-6 receptor pathways in coronary heart disease: A collaborative meta-analysis of 82 studies. *The Lancet* **379**, 1205–1213 (2012).
97. Swerdlow, D. I., Holmes, M. V., Kuchenbaecker, K. B., Engmann, J. E., Shah, T., *et al.* The interleukin-6 receptor as a target for prevention of coronary heart disease: A mendelian randomisation analysis. *The Lancet* **379**, 1214–1224 (2012).
98. Chen, Z., Peto, R., Zhou, M., Iona, A., Smith, M., *et al.* Contrasting male and female trends in tobacco-attributed mortality in China: evidence from successive nationwide prospective cohort studies. *The Lancet* **386**, 1447–1456 (2015).
99. Millwood, I. Y., Li, L., Smith, M., Guo, Y., Yang, L., *et al.* Alcohol consumption in 0.5 million people from 10 diverse regions of china: Prevalence, patterns and socio-demographic and health-related correlates. *International Journal of Epidemiology* **42**, 816–827 (2013).
100. Du, H., Li, L., Whitlock, G., Bennett, D., Guo, Y., *et al.* Patterns and socio-demographic correlates of domain-specific physical activities and their associations with adiposity in the China Kadoorie Biobank study. *BMC Public Health* **14** (2014).
101. Holmes, M. V., Millwood, I. Y., Kartsonaki, C., Hill, M. R., Bennett, D. A., *et al.* Lipids, Lipoproteins, and Metabolites and Risk of Myocardial Infarction and Stroke. *Journal of the American College of Cardiology* **71**, 620–632 (2018).
102. Lewington, S., Lacey, B., Clarke, R., Guo, Y., Kong, X. L., *et al.* The burden of hypertension and associated risk for cardiovascular mortality in China. *JAMA Internal Medicine* **176**, 524–532 (2016).
103. Biasucci, L. M. C-reactive protein and other markers of inflammation in the prediction of cardiovascular disease in women. *Italian heart journal. Supplement : official journal of the Italian Federation of Cardiology* **1**, 1066–1067 (2000).
104. Libby, P. & Ridker, P. M. Inflammation and Atherothrombosis. From Population Biology and Bench Research to Clinical Practice. *Journal of the American College of Cardiology* **48**, A33–A46 (2006).
105. Du, Y., Xu, X., Chu, M., Guo, Y. & Wang, J. Air particulate matter and cardiovascular disease: The epidemiological, biomedical and clinical evidence. *Journal of Thoracic Disease* **8**, E8–E19 (2016).
106. Keusch, G. T., Rosenberg, I. H., Denno, D. M., Duggan, C., Guerrant, R. L., *et al.* Implications of acquired environmental enteric dysfunction for growth and stunting in infants and children living in low-and middle-income countries. *Food and Nutrition Bulletin* **34**, 357–364 (2013).

107. Franceschi, C., Garagnani, P., Parini, P., Giuliani, C. & Santoro, A. Inflammaging: a new immune–metabolic viewpoint for age-related diseases. *Nature Reviews Endocrinology* **14**, 576–590 (2018).
108. Liberale, L., Montecucco, F., Tardif, J. C., Libby, P. & Camici, G. G. Inflamm-aging: The role of inflammation in age-dependent cardiovascular disease. *European Heart Journal* **41**, 2974–2982 (2020).
109. Pilling, L. C., Kuo, C. L., Sicinski, K., Tamosauskaite, J., Kuchel, G. A., *et al.* Human longevity: 25 genetic loci associated in 389,166 UK biobank participants. *Aging* **9**, 2504–2520 (2017).
110. Camell, C. D., Günther, P., Lee, A., Goldberg, E. L., Spadaro, O., *et al.* Aging Induces an Nlrp3 Inflammasome-Dependent Expansion of Adipose B Cells That Impairs Metabolic Homeostasis. *Cell Metabolism* **30**, 1024–1039.e6 (2019).
111. Sebastian-Valverde, M. & Pasinetti, G. M. The NLRP3 Inflammasome as a Critical Actor in the Inflammaging Process. *Cells* **9**, 6 (2020).
112. Youm, Y.-H., Grant, R. W., McCabe, L. R., Albarado, D. C., Nguyen, K. Y., *et al.* Canonical Nlrp3 Inflammasome Links Systemic Low-Grade Inflammation to Functional Decline in Aging. *Cell Metabolism* **18**, 519–532 (Oct. 2013).
113. Li, H., Mitchell, L., Zhang, X., Heiselman, D. & Motosko, S. Testosterone Therapy and Risk of Acute Myocardial Infarction in Hypogonadal Men: An Administrative Health Care Claims Study. *Journal of Sexual Medicine* **14**, 1307–1317 (2017).
114. Martinez, C., Suissa, S., Rietbrock, S., Katholing, A., Freedman, B., *et al.* Testosterone treatment and risk of venous thromboembolism: Population based case-control study. *BMJ (Online)* **355** (2016).
115. Vigen, R., O'Donnell, C. I., Barón, A. E., Grunwald, G. K., Maddox, T. M., *et al.* Association of testosterone therapy with mortality, myocardial infarction, and stroke in men with low testosterone levels. *JAMA - Journal of the American Medical Association* **310**, 1829–1836 (2013).
116. Luo, S., Au Yeung, S. L., Zhao, J. V., Burgess, S. & Schooling, C. M. Association of genetically predicted testosterone with thromboembolism, heart failure, and myocardial infarction: Mendelian randomisation study in UK Biobank. *BMJ (Online)* **364** (2019).
117. Alves, J. V., da Costa, R. M., Pereira, C. A., Fedoce, A. G., Silva, C. A. A., *et al.* Supraphysiological Levels of Testosterone Induce Vascular Dysfunction via Activation of the NLRP3 Inflammasome. *Frontiers in Immunology* **11** (2020).
118. Ellulu, M. S., Patimah, I., Khaza'ai, H., Rahmat, A. & Abed, Y. Obesity & inflammation: The linking mechanism & the complications. *Archives of Medical Science* **13**, 851–863 (2017).
119. Nakayama, T. & Wang, Z. Inflammation, a link between obesity and cardiovascular disease. *Mediators of Inflammation* **2010** (2010).
120. Visser, M., Bouter, L. M., McQuillan, G. M., Wener, M. H. & Harris, T. B. Elevated C-reactive protein levels in overweight and obese adults. *Journal of the American Medical Association* **282**, 2131–2135 (1999).
121. Welsh, P., Polisecki, E., Robertson, M., Jahn, S., Buckley, B. M., *et al.* Unraveling the directional link between adiposity and inflammation: A bidirectional mendelian randomization approach. *Journal of Clinical Endocrinology and Metabolism* **95**, 93–99 (2010).
122. Nicklas, B. J., Ambrosius, W., Messier, S. P., Miller, G. D., Penninx, B. W., *et al.* Diet-induced weight loss, exercise, and chronic inflammation in older, obese adults:

- A randomized controlled clinical trial. *American Journal of Clinical Nutrition* **79**, 544–551 (2004).
123. Weisberg, S. P., McCann, D., Desai, M., Rosenbaum, M., Leibel, R. L., *et al.* Obesity is associated with macrophage accumulation in adipose tissue. *Journal of Clinical Investigation* **112**, 1796–1808 (2003).
124. Rajala, M. W. & Scherer, P. E. Minireview: The adipocyte - At the crossroads of energy homeostasis, inflammation, and atherosclerosis. *Endocrinology* **144**, 3765–3773 (2003).
125. Öhman, M. K., Shen, Y., Obimba, C. I., Wright, A. P., Warnock, M., *et al.* Visceral adipose tissue inflammation accelerates atherosclerosis in apolipoprotein E-deficient mice. *Circulation* **117**, 798–805 (2008).
126. Bautista, L. E., López-Jaramillo, P., Vera, L. M., Casas, J. P., Otero, A. P., *et al.* Is C-reactive protein an independent risk factor for essential hypertension? *Journal of Hypertension* **19**, 857–861 (2001).
127. Sesso, H. D., Buring, J. E., Rifai, N., Blake, G. J., Gaziano, J. M., *et al.* C-Reactive Protein and the Risk of Developing Hypertension. *Journal of the American Medical Association* **290**, 2945–2951 (2003).
128. Smith, G. D., Lawlor, D. A., Harbord, R., Timpson, N., Rumley, A., *et al.* Association of C-reactive protein with blood pressure and hypertension: Life course confounding and Mendelian randomization tests of causality. *Arteriosclerosis, Thrombosis, and Vascular Biology* **25**, 1051–1056 (2005).
129. Chae, C. U., Lee, R. T., Rifai, N. & Ridker, P. M. Blood pressure and inflammation in apparently healthy men. *Hypertension* **38**, 399–403 (2001).
130. Hackshaw, A., Morris, J. K., Boniface, S., Tang, J. L. & Milenkovi, D. Low cigarette consumption and risk of coronary heart disease and stroke: Meta-analysis of 141 cohort studies in 55 study reports. *BMJ (Online)* **360** (2018).
131. Bergmann, S., Siekmeier, R., Mix, C. & Jaross, W. Even moderate cigarette smoking influences the pattern of circulating monocytes and the concentration of sICAM-1. *Respiration Physiology* **114**, 269–275 (1998).
132. Kalra, V. K., Ying, Y., Deemer, K., Coates, T. D., Natarajan, R., *et al.* Mechanism of cigarette smoke condensate induced adhesion of human monocytes to cultured endothelial cells. *Journal of Cellular Physiology* **160**, 154–162 (1994).
133. Shen, Y., Rattan, V., Sultana, C. & Kalra, V. K. Cigarette smoke condensate-induced adhesion molecule expression and transendothelial migration of monocytes. *American Journal of Physiology - Heart and Circulatory Physiology* **270** (1996).
134. Tracy, R. P., Psaty, B. M., Macy, E., Bovill, E. G., Cushman, M., *et al.* Lifetime smoking exposure affects the association of C-reactive protein with cardiovascular disease risk factors and subclinical disease in healthy elderly subjects. *Arteriosclerosis, Thrombosis, and Vascular Biology* **17**, 2167–2176 (1997).
135. Tappia, P. S., Troughton, K. L., Langley-Evans, S. C. & Grimble, R. F. Cigarette smoking influences cytokine production and antioxidant defences. *Clinical Science* **88**, 485–489 (1995).
136. Bell, S., Mehta, G., Moore, K. & Britton, A. Ten-year alcohol consumption typologies and trajectories of C-reactive protein, interleukin-6 and interleukin-1 receptor antagonist over the following 12 years: a prospective cohort study. *Journal of Internal Medicine* **281**, 75–85 (2017).

137. Volpato, S., Pahor, M., Ferrucci, L., Simonsick, E. M., Guralnik, J. M., *et al.* Relationship of Alcohol Intake with Inflammatory Markers and Plasminogen Activator Inhibitor-1 in Well-Functioning Older Adults: The Health, Aging, and Body Composition Study. *Circulation* **109**, 607–612 (2004).
138. King, D. E., Mainous, A. G., Buchanan, T. A. & Pearson, W. S. C-reactive protein and glycemic control in adults with diabetes. *Diabetes Care* **26**, 1535–1539 (2003).
139. Estruch, R. Anti-inflammatory effects of the Mediterranean diet: The experience of the PREDIMED study. *Proceedings of the Nutrition Society* **69**, 333–340 (2010).
140. Schwingshackl, L. & Hoffmann, G. Mediterranean dietary pattern, inflammation and endothelial function: A systematic review and meta-analysis of intervention trials. *Nutrition, Metabolism and Cardiovascular Diseases* **24**, 929–939 (2014).
141. Estruch, R., Ros, E., Salas-Salvadó, J., Covas, M. I., Corella, D., *et al.* Primary prevention of cardiovascular disease with a mediterranean diet. *Zeitschrift für Gefassmedizin* **10**, 28 (2013).
142. Shivappa, N., Steck, S. E., Hurley, T. G., Hussey, J. R. & Hébert, J. R. Designing and developing a literature-derived, population-based dietary inflammatory index. *Public Health Nutrition* **17**, 1689–1696 (2014).
143. Garcia-Arellano, A., Ramallal, R., Ruiz-Canela, M., Salas-Salvadó, J., Corella, D., *et al.* Dietary inflammatory index and incidence of cardiovascular disease in the PREDIMED study. *Nutrients* **7**, 4124–4138 (2015).
144. Bennett, D. A., Du, H., Clarke, R., Guo, Y., Yang, L., *et al.* Association of physical activity with risk of major cardiovascular diseases in Chinese men and women. *JAMA Cardiology* **2**, 1349–1358 (2017).
145. Lee, C. D., Folsom, A. R. & Blair, S. N. Physical activity and stroke risk: A meta-analysis. *Stroke* **34**, 2475–2481 (2003).
146. Li, J. & Siegrist, J. Physical activity and risk of cardiovascular disease-a meta-analysis of prospective cohort studies. *International Journal of Environmental Research and Public Health* **9**, 391–407 (2012).
147. Mora, S., Cook, N., Buring, J. E., Ridker, P. M. & Lee, I. M. Physical activity and reduced risk of cardiovascular events: Potential mediating mechanisms. *Circulation* **116**, 2110–2118 (2007).
148. Rothenbacher, D., Hoffmeister, A., Brenner, H. & Koenig, W. Physical activity, coronary heart disease, and inflammatory response. *Archives of Internal Medicine* **163**, 1200–1205 (2003).
149. Hamer, M., Sabia, S., Batty, G. D., Shipley, M. J., Tabák, A. G., *et al.* Physical activity and inflammatory markers over 10 years: Follow-up in men and women from the whitehall II cohort study. *Circulation* **126**, 928–933 (2012).
150. Liu, Y., Wheaton, A. G., Chapman, D. P. & Croft, J. B. Sleep duration and chronic diseases among us adults age 45 years and older: Evidence from the 2010 Behavioral Risk Factor Surveillance System. *Sleep* **36**, 1421–1427 (2013).
151. Partinen, M., Putkonen, P. T., Kaprio, J., Koskenvuo, M. & Hilakivi, I. Sleep Disorders in Relation to Coronary Heart Disease. *Acta Medica Scandinavica* **211**, 69–83 (1982).
152. Sabanayagam, C. & Shankar, A. Sleep duration and cardiovascular disease: Results from the National Health Interview Survey. *Sleep* **33**, 1037–1042 (2010).
153. Ge, B. & Guo, X. Short and long sleep durations are both associated with increased risk of stroke: A meta-analysis of observational studies. *International Journal of Stroke* **10**, 177–184 (2015).

154. Patel, S. R., Zhu, X., Storfer-Isser, A., Mehra, R., Jenny, N. S., *et al.* Sleep duration and biomarkers of inflammation. *Sleep* **32**, 200–204 (2009).
155. Banks, S. & Dinges, D. F. Behavioral and physiological consequences of sleep restriction. *Journal of Clinical Sleep Medicine* **3**, 519–528 (2007).
156. Grandner, M. A., Hale, L., Moore, M. & Patel, N. P. Mortality associated with short sleep duration: The evidence, the possible mechanisms, and the future. *Sleep Medicine Reviews* **14**, 191–203 (2010).
157. Biasucci, L. M. C-reactive protein and other markers of inflammation in the prediction of cardiovascular disease in women. *Italian heart journal. Supplement : official journal of the Italian Federation of Cardiology* **1**, 1066–1067 (Mar. 2000).
158. Guerrant, R. L., Leite, A. M., Pinkerton, R., Medeiros, P. H., Cavalcante, P. A., *et al.* Biomarkers of environmental enteropathy, inflammation, stunting, and impaired growth in children in northeast Brazil. *PLoS ONE* **11** (2016).
159. Van Eeden, S. F., Tan, W. C., Suwa, T., Mukae, H., Terashima, T., *et al.* Cytokines involved in the systemic inflammatory response induced by exposure to particulate matter air pollutants (PM10). *American Journal of Respiratory and Critical Care Medicine* **164**, 826–830 (2001).
160. Bourey, R. E., Kaw, M. K., Lester, S. G., Ghanem, S. S. & Najjar, S. M. Diabetes. *Diet, Exercise, and Chronic Disease: The Biological Basis of Prevention*, 101–132 (2014).
161. Bentzon, J. F., Otsuka, F., Virmani, R. & Falk, E. Mechanisms of plaque formation and rupture. *Circulation Research* **114**, 1852–1866 (2014).
162. Zhang, J., Yang, Y., Sun, H. & Xing, Y. Hemorrhagic transformation after cerebral infarction: Current concepts and challenges. *Annals of Translational Medicine* **2** (2014).
163. Poon, M. T., Bell, S. M. & Al-Shahi Salman, R. Epidemiology of Intracerebral Haemorrhage. *Frontiers of Neurology and Neuroscience* **37**, 1–12 (2015).
164. Mayer, S. & Hillis, L. D. Prinzmetal’s variant angina. *Clinical Cardiology* **21**, 243–246 (1998).
165. Hansson, G. K. Inflammation and atherosclerosis the end of a controversy. *Circulation* **136**, 1875–1877 (Nov. 2017).
166. Libby, P. Inflammation in atherosclerosis. *Nature* **420**, 868–874 (2002).
167. Libby, P. Inflammation in atherosclerosis. *Arteriosclerosis, Thrombosis, and Vascular Biology* **32**, 2045–2051 (Sept. 2012).
168. Denburg, J. A. The inflammatory response. *American Journal of Respiratory and Critical Care Medicine* **153** (1996).
169. Tabas, I. & Lichtman, A. H. Monocyte-Macrophages and T Cells in Atherosclerosis. *Immunity* **47**, 621–634 (2017).
170. Virchow, R. *Cellular pathology. As based upon physiological and pathological histology. Lecture XVI—Atheromatous affection of arteries.* 23–25 (1989).
171. Ruparel, N., Chai, J. T., Fisher, E. A. & Choudhury, R. P. Inflammatory processes in cardiovascular disease: a route to targeted therapies. *Nat Rev Cardiol* **14**, 133–144 (2017).
172. Cunningham, K. S. & Gotlieb, A. I. The role of shear stress in the pathogenesis of atherosclerosis. *Laboratory Investigation* **85**, 9–23 (2005).
173. Chiu, J. J. & Chien, S. Effects of disturbed flow on vascular endothelium: Pathophysiological basis and clinical perspectives. *Physiological Reviews* **91**, 327–387 (2011).

174. Borish, L. C. & Steinke, J. W. 2. Cytokines and chemokines. *Journal of Allergy and Clinical Immunology* **111**, S460–S475 (2003).
175. Han, K. H., Hong, K. H., Park, J. H., Ko, J., Kang, D. H., *et al.* C-reactive protein promotes monocyte chemoattractant protein-1-mediated chemotaxis through upregulating CC chemokine receptor 2 expression in human monocytes. *Circulation* **109**, 2566–2571 (2004).
176. Nakajima, K., Nakano, T. & Tanaka, A. The oxidative modification hypothesis of atherosclerosis: The comparison of atherogenic effects on oxidized LDL and remnant lipoproteins in plasma. *Clinica Chimica Acta* **367**, 36–47 (2006).
177. Witztum, J. L. & Steinberg, D. The oxidative modification hypothesis of atherosclerosis: Does it hold for humans? *Trends in Cardiovascular Medicine* **11**, 93–102 (2001).
178. Mestas, J. & Ley, K. Monocyte-Endothelial Cell Interactions in the Development of Atherosclerosis. *Trends in Cardiovascular Medicine* **18**, 228–232 (2008).
179. Galis, Z. S., Muszynski, M., Sukhova, G. K., Simon-Morrissey, E., Unemori, E. N., *et al.* Cytokine-stimulated human vascular smooth muscle cells synthesize a complement of enzymes required for extracellular matrix digestion. *Circulation Research* **75**, 181–189 (1994).
180. Duewell, P., Kono, H., Rayner, K. J., Sirois, C. M., Vladimer, G., *et al.* NLRP3 inflammasomes are required for atherogenesis and activated by cholesterol crystals. *Nature* **464**, 1357–1361 (Apr. 2010).
181. Keller, M., Rüegg, A., Werner, S. & Beer, H. D. Active Caspase-1 Is a Regulator of Unconventional Protein Secretion. *Cell* **132**, 818–831 (2008).
182. Bevilacqua, M. P., Pober, J. S. & Wheeler, M. E. Interleukin-1 activation of vascular endothelium. Effects on procoagulant activity and leukocyte adhesion. *American Journal of Pathology* **121**, 393–403 (1985).
183. Dinarello, C. A., Ikejima, T., Warner, S. J., Orencole, S. F., Lonnemann, G., *et al.* Interleukin 1 induces interleukin 1. I. Induction of circulating interleukin 1 in rabbits in vivo and in human mononuclear cells in vitro. *Journal of immunology (Baltimore, Md. : 1950)* **139**, 1902–10 (1987).
184. Libby, P., Warner, S. J. & Friedman, G. B. Interleukin 1: A mitogen for human vascular smooth muscle cells that induces the release of growth-inhibitory prostanoids. *Journal of Clinical Investigation* **81**, 487–498 (1988).
185. Loppnow, H. & Libby, P. Proliferating or interleukin 1-activated human vascular smooth muscle cells secrete copious interleukin 6. *Journal of Clinical Investigation* **85**, 731–738 (1990).
186. Klouche, M., Bhakdi, S., Hemmes, M. & Rose-John, S. Novel path to activation of vascular smooth muscle cells: up-regulation of gp130 creates an autocrine activation loop by IL-6 and its soluble receptor. *Journal of immunology (Baltimore, Md. : 1950)* **163**, 4583–9 (1999).
187. Kraakman, M. J., Kammoun, H. L., Allen, T. L., Deswaerte, V., Henstridge, D. C., *et al.* Blocking IL-6 trans-signaling prevents high-fat diet-induced adipose tissue macrophage recruitment but does not improve insulin resistance. *Cell Metabolism* **21**, 403–416 (2015).
188. Ganapathi, M. K., May, L. T., Schultz, D., Brabenec, A., Weinstein, J., *et al.* Role of interleukin-6 in regulating synthesis of C-reactive protein and serum amyloid A in human hepatoma cell lines. *Biochemical and Biophysical Research Communications* **157**, 271–277 (1988).

189. Heinrich, P. C., Castell, J. V. & Andus, T. Interleukin-6 and the acute phase response. *Biochemical Journal* **265**, 621–636 (1990).
190. Reiss, A. B., Siegert, N. M. & De Leon, J. Interleukin-6 in atherosclerosis: atherogenic or atheroprotective? *Clinical Lipidology* **12**, 14–23 (2017).
191. Scheller, J., Chalaris, A., Schmidt-Arras, D. & Rose-John, S. The pro- and anti-inflammatory properties of the cytokine interleukin-6. *Biochimica et Biophysica Acta - Molecular Cell Research* **1813**, 878–888 (2011).
192. Aguilar, M. I. & Brott, T. G. Update in Intracerebral Hemorrhage. *The Neurohospitalist* **1**, 148–159 (2011).
193. Chen, S., Yang, Q., Chen, G. & Zhang, J. H. An Update on Inflammation in the Acute Phase of Intracerebral Hemorrhage. *Translational Stroke Research* **6**, 4–8 (2015).
194. Sheth, K. N. & Rosand, J. Targeting the immune system in intracerebral hemorrhage. *JAMA Neurology* **71**, 1083–1084 (2014).
195. Zhou, Y., Wang, Y., Wang, J., Anne Stetler, R. & Yang, Q. W. Inflammation in intracerebral hemorrhage: From mechanisms to clinical translation. *Progress in Neurobiology* **115**, 25–44 (2014).
196. Xiong, X. Y. & Yang, Q. W. Rethinking the Roles of Inflammation in the Intracerebral Hemorrhage. *Translational Stroke Research* **6**, 339–341 (2015).
197. Wang, J. Preclinical and clinical research on inflammation after intracerebral hemorrhage. *Progress in Neurobiology* **92**, 463–477 (2010).
198. Mendelow, A. D., Gregson, B. A., Rowan, E. N., Murray, G. D., Gholkar, A., *et al.* Early surgery versus initial conservative treatment in patients with spontaneous supratentorial lobar intracerebral haematomas (STICH II): A randomised trial. *The Lancet* **382**, 397–408 (2013).
199. Shen, J., Xie, L., Mao, X. O., Zhou, Y., Zhan, R., *et al.* Neurogenesis after primary intracerebral hemorrhage in adult human brain. *Journal of Cerebral Blood Flow and Metabolism* **28**, 1460–1468 (2008).
200. Yang, P., Wen, H., Ou, S., Cui, J. & Fan, D. IL-6 promotes regeneration and functional recovery after cortical spinal tract injury by reactivating intrinsic growth program of neurons and enhancing synapse formation. *Experimental Neurology* **236**, 19–27 (2012).
201. Kerr, R., Stirling, D. & Ludlam, C. A. Interleukin 6 and haemostasis. *British Journal of Haematology* **115**, 3–12 (2001).
202. Casolla, B., Moulin, S., Kyheng, M., Hénon, H., Labreuche, J., *et al.* Five-Year Risk of Major Ischemic and Hemorrhagic Events After Intracerebral Hemorrhage. *Stroke* **50**, 1100–1107 (2019).
203. Gabay, C. & Kushner, I. Acute-Phase Proteins and Other Systemic Responses to Inflammation. *New England Journal of Medicine* **340**, 448–454 (1999).
204. Masek, Z., Maruna, P. & Schreiber, V. Acute phase proteins. *Sborník lékařský* **96**, 451–454 (Apr. 1995).
205. Ridker, P. M. From C-Reactive Protein to Interleukin-6 to Interleukin-1: Moving Upstream to Identify Novel Targets for Atheroprotection. *Circulation Research* **118**, 145–156 (2016).
206. Thompson, D., Pepys, M. B. & Wood, S. P. The physiological structure of human C-reactive protein and its complex with phosphocholine. *Structure* **7**, 169–177 (1999).

207. Tillett, W. S. & Francis, T. Serological reactions in pneumonia with a nonprotein somatic fraction of pneumococcus. *Journal of Experimental Medicine* **52**, 561–571 (1930).
208. Landry, A., Docherty, P., Ouellette, S. & Cartier, L. J. Causes and outcomes of markedly elevated C-reactive protein levels. *Canadian Family Physician* **63**, e316–e323 (2017).
209. Costa-Filho, R., Hochleitner, G., Wendt, M., Teruya, A. & Spahn, D. R. Over 50 years of fibrinogen concentrate. *Clinical and Applied Thrombosis/Hemostasis* **22**, 109–114 (2016).
210. Auron, P. E., Webb, A. C., Rosenwasser, L. J., Mucci, S. F., Rich, A., *et al.* Nucleotide sequence of human monocyte interleukin 1 precursor cDNA. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 7907–7911 (1984).
211. Aggarwal, B. B., Kohr, W. J., Hass, P. E., Moffat, B., Spencer, S. A., *et al.* Human tumor necrosis factor: Production, purification, and characterization. *Journal of Biological Chemistry* **260**, 2345–2354 (1985).
212. HAEGEMAN, G., CONTENT, J., VOLCKAERT, G., DERYNCK, R., TAVERNIER, J., *et al.* Structural analysis of the sequence coding for an inducible 26-kDa protein in human fibroblasts. *European Journal of Biochemistry* **159**, 625–632 (1986).
213. Hirano, T., Yasukawa, K., Harada, H., Taga, T., Watanabe, Y., *et al.* Complementary DNA for a novel human interleukin (BSF-2) that induces B lymphocytes to produce immunoglobulin. *Nature* **324**, 73–76 (1986).
214. Zilberstein, A., Ruggieri, R., Korn, J. H. & Revel, M. Structure and expression of cDNA and genes for human interferon-beta-2, a distinct species inducible by growth-stimulatory cytokines. *The EMBO journal* **5**, 2529–2537 (Oct. 1986).
215. Gauldie, J., Richards, C., Harnish, D., Lansdorp, P. & Baumann, H. Interferon beta 2/B-cell stimulatory factor type 2 shares identity with monocyte-derived hepatocyte-stimulating factor and regulates the major acute phase protein response in liver cells. *Proceedings of the National Academy of Sciences of the United States of America* **84**, 7251–7255 (1987).
216. Castell, J. V., Gómez-lechón, M. J., David, M., Fabra, R., Trullenque, R., *et al.* Acute-phase response of human hepatocytes: Regulation of acute-phase protein synthesis by interleukin-6. *Hepatology* **12**, 1179–1186 (1990).
217. Libby, P., Ordovas, J. M., Birinyi, L. K., Auger, K. R. & Dinarello, C. A. Inducible interleukin-1 gene expression in human vascular smooth muscle cells. *Journal of Clinical Investigation* **78**, 1432–1438 (1986).
218. VAN DAMME, J., CAYPHAS, S., VAN SNICK, J., CONINGS, R., PUT, W., *et al.* Purification and characterization of human fibroblast-derived hybridoma growth factor identical to T-cell-derived B-cell stimulatory factor-2 (interleukin-6). *European Journal of Biochemistry* **168**, 543–550 (1987).
219. Sironi, M., Breviario, F., Proserpio, P., Biondi, A., Vecchi, A., *et al.* IL-1 stimulates IL-6 production in endothelial cells. *Journal of immunology (Baltimore, Md. : 1950)* **142**, 549–53 (1989).
220. LUGER, T. A., SCHWARZ, T., KRUTMANN, J., KIRNBAUER, R., NEUNER, P., *et al.* Interleukin-6 Is Produced by Epidermal Cells and Plays an Important Role in the Activation of Human T-Lymphocytes and Natural Killer Cells. *Annals of the New York Academy of Sciences* **557**, 405–414 (1989).

221. Tosato, G. & Jones, K. D. Interleukin-1 induces interleukin-6 production in peripheral blood monocytes. *Blood* **75**, 1305–1310 (1990).
222. Swerdlow, D. I., Holmes, M. V., Kuchenbaecker, K. B., Engmann, J. E., Shah, T., *et al.* The interleukin-6 receptor as a target for prevention of coronary heart disease: A mendelian randomisation analysis. *The Lancet* **379**, 1214–1224 (2012).
223. Ridker, P. M., Thuren, T., Zalewski, A. & Libby, P. Interleukin-1 β inhibition and the prevention of recurrent cardiovascular events: Rationale and Design of the Canakinumab Anti-inflammatory Thrombosis Outcomes Study (CANTOS). *American Heart Journal* **162**, 597–605 (2011).
224. Ridker, P. M. From C-Reactive Protein to Interleukin-6 to Interleukin-1: Moving Upstream to Identify Novel Targets for Atheroprotection. *Circulation Research* **118**, 145–156 (Jan. 2016).
225. Kaptoge, S., Seshasai, S. R. K., Gao, P., Freitag, D. F., Butterworth, A. S., *et al.* Inflammatory cytokines and risk of coronary heart disease: New prospective study and updated meta-analysis. *European Heart Journal* **35**, 578–589 (Mar. 2014).
226. Ridker, P. M., Rifai, N., Stampfer, M. J. & Hennekens, C. H. Plasma concentration of interleukin-6 and the risk of future myocardial infarction among apparently healthy men. *Circulation* **101**, 1767–1772 (2000).
227. Stec, J. J., Silbershatz, H., Tofler, G. H., Matheney, T. H., Sutherland, P., *et al.* Association of fibrinogen with cardiovascular risk factors and cardiovascular disease the Framingham Offspring Population. *Circulation* **102**, 1634–1638 (2000).
228. Zakai, N. A., Katz, R., Jenny, N. S., Psaty, B. M., Reiner, A. P., *et al.* Inflammation and hemostasis biomarkers and cardiovascular risk in the elderly: The Cardiovascular Health Study. *Journal of Thrombosis and Haemostasis* **5**, 1128–1135 (2007).
229. Danesh, J., Kaptoge, S., Mann, A. G., Sarwar, N., Wood, A., *et al.* Long-term interleukin-6 levels and subsequent risk of coronary heart disease: Two new prospective studies and a systematic review. *PLoS Medicine* **5**, 0600–0610 (2008).
230. Saito, I., Maruyama, K. & Eguchi, E. C-reactive protein and cardiovascular disease in east asians: A systematic review. *Clinical Medicine Insights: Cardiology* **8**, 35–42 (2014).
231. Liu, Y., Wang, J., Zhang, L., Wang, C., Wu, J., *et al.* Relationship between C - Reactive protein and stroke: A large prospective community based study. *PLoS ONE* **9**, e107017 (2014).
232. Keyes, K. M. & Westreich, D. UK Biobank, big data, and the consequences of non-representativeness. *The Lancet* **393**, 1297 (2019).
233. Lesko, C. R., Buchanan, A. L., Westreich, D., Edwards, J. K., Hudgens, M. G., *et al.* Generalizing Study Results: A Potential Outcomes Perspective. *Epidemiology* **28**, 553–561 (2017).
234. Westreich, D., Edwards, J. K., Lesko, C. R., Cole, S. R. & Stuart, E. A. Target Validity and the Hierarchy of Study Designs. *American Journal of Epidemiology* **188**, 438–443 (2019).
235. Jiménez, M. C., Rexrode, K. M., Kotler, G., Everett, B. M., Glynn, R. J., *et al.* Association Between Markers of Inflammation and Total Stroke by Hypertensive Status Among Women. *American Journal of Hypertension* **29**, 1117–1124 (2016).
236. Clarke, R., Valdes-Marquez, E., Hill, M., Gordon, J., Farrall, M., *et al.* Plasma cytokines and risk of coronary heart disease in the PROCARDIS study. *Open Heart* **5**, e000807 (2018).

237. Wang, W., Jiang, B., Sun, H., Ru, X., Sun, D., *et al.* Prevalence, Incidence, and Mortality of Stroke in China: Results from a Nationwide Population-Based Survey of 480 687 Adults. *Circulation* **135**, 759–771 (2017).
238. Kola, I. & Landis, J. Can the pharmaceutical industry reduce attrition rates? *Nature Reviews Drug Discovery* **3**, 711–715 (2004).
239. Hwang, T. J., Lauffenburger, J. C., Franklin, J. M. & Kesselheim, A. S. Temporal Trends and Factors Associated With Cardiovascular Drug Development, 1990 to 2012. *JACC: Basic to Translational Science* **1**, 301–308 (2016).
240. DiMasi, J. A., Grabowski, H. G. & Hansen, R. W. The Cost of Drug Development. *New England Journal of Medicine* **372**, 1972–1972 (2015).
241. Hingorani, A. D., Kuan, V., Finan, C., Kruger, F. A., Gaulton, A., *et al.* Improving the odds of drug development success through human genomics: modelling study. *Scientific Reports* **9**, 2045–2322 (2019).
242. Nelson, M. R., Tipney, H., Painter, J. L., Shen, J., Nicoletti, P., *et al.* The support of human genetic evidence for approved drug indications. *Nature Genetics* **47**, 856–860 (2015).
243. Zheng, J., Haberland, V., Baird, D., Walker, V. M., Haycock, P. M., *et al.* Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat Genet* **52**, 1122–1131 (2020).
244. Yokoyama, A., Tsutsumi, E., Imazeki, H., Suwa, Y., Nakamura, C., *et al.* Polymorphisms of alcohol dehydrogenase-1b and aldehyde dehydrogenase-2 and the blood and salivary ethanol and acetaldehyde concentrations of Japanese alcoholic men. *Alcoholism: Clinical and Experimental Research* **34**, 1246–1256 (2010).
245. Ridker, P. M., Hennekens, C. H., Roitman-Johnson, B., Stampfer, M. J. & Allen, J. Plasma concentration of soluble intercellular adhesion molecule 1 and risks of future myocardial infarction in apparently healthy men. *Lancet* **351**, 88–92 (Jan. 1998).
246. Ridker, P. M., Rifai, N., Pfeifer, M. A., Sacks, F. & Braunwald, E. Long-term effects of pravastatin on plasma concentration of C-reactive protein. *Circulation* **100**, 230–235 (1999).
247. Nissen, S. E., Tuzcu, E. M., Schoenhagen, P., Crowe, T., Sasiela, W. J., *et al.* Statin Therapy, LDL Cholesterol, C-Reactive Protein, and Coronary Artery Disease. *New England Journal of Medicine* **352**, 29–38 (Jan. 2005).
248. Ridker, P. M., Cannon, C. P., Morrow, D., Rifai, N., Rose, L. M., *et al.* C-Reactive Protein Levels and Outcomes after Statin Therapy. *New England Journal of Medicine* **352**, 20–28 (Jan. 2005).
249. Bohula, E. A., Giugliano, R. P., Cannon, C. P., Zhou, J., Murphy, S. A., *et al.* Achievement of dual low-density lipoprotein cholesterol and high-sensitivity c-reactive protein targets more frequent with the addition of ezetimibe to simvastatin and associated with better outcomes in IMPROVE-IT. *Circulation* **132**, 1224–1233 (2015).
250. Ridker, P. M., Danielson, E., Fonseca, F. A., Genest, J., Gotto, A. M., *et al.* Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. *Obstetrical and Gynecological Survey* **64**, 168–170 (Nov. 2009).
251. Arnaud, C., Burger, F., Steffens, S., Veillard, N. R., Nguyen, T. H., *et al.* Statins reduce interleukin-6-induced C-reactive protein in human hepatocytes: New evidence for direct antiinflammatory effects of statins. *Arteriosclerosis, Thrombosis, and Vascular Biology* **25**, 1231–1236 (2005).

252. Mayer, C., Gruber, H. J., Landl, E. M., Pailer, S., Scharnagl, H., *et al.* Rosuvastatin reduces interleukin-6-induced expression of C-reactive protein in human hepatocytes in a STAT3- and C/EBP-dependent fashion. *International Journal of Clinical Pharmacology and Therapeutics* **45**, 319–327 (2007).
253. Everett, B. M., Pradhan, A. D., Solomon, D. H., Paynter, N., MacFadyen, J., *et al.* Rationale and design of the Cardiovascular Inflammation Reduction Trial: A test of the inflammatory hypothesis of atherothrombosis. *American Heart Journal* **166**, 199–207.e15 (Aug. 2013).
254. Rondeau, J. M., Ramage, P., Zurini, M. & Gram, H. The molecular mode of action and species specificity of canakinumab, a human monoclonal antibody neutralizing IL-1 β . *mAbs* **7**, 1151–1160 (2015).
255. Reiss, A. B., Coomes, E. & Chan, E. S. Methotrexate in atherogenesis and cholesterol metabolism. *Cholesterol* **2011**, 1–8 (Feb. 2011).
256. Reiss, A. B., Carsons, S. E., Anwar, K., Rao, S., Edelman, S. D., *et al.* Atheroprotective effects of methotrexate on reverse cholesterol transport proteins and foam cell transformation in human THP-1 monocyte/macrophages. *Arthritis & Rheumatism* **58**, 3675–3683 (Dec. 2008).
257. Ridker, P. M., Everett, B. M., Thuren, T., MacFadyen, J. G., Chang, W. H., *et al.* Antiinflammatory Therapy with Canakinumab for Atherosclerotic Disease. *New England Journal of Medicine* **377**, 1119–1131 (2017).
258. Ridker, P. M., MacFadyen, J. G., Everett, B. M., Libby, P., Thuren, T., *et al.* Relationship of C-reactive protein reduction to cardiovascular event reduction following treatment with canakinumab: a secondary analysis from the CANTOS randomised controlled trial. *The Lancet* **391**, 319–328 (Jan. 2018).
259. Ridker, P. M., Everett, B. M., Pradhan, A., MacFadyen, J. G., Solomon, D. H., *et al.* Low-Dose Methotrexate for the Prevention of Atherosclerotic Events. *New England Journal of Medicine* **380**, 752–762 (Feb. 2019).
260. Ridker, P. M. Anti-inflammatory therapy for atherosclerosis: interpreting divergent results from the CANTOS and CIRT clinical trials. *Journal of Internal Medicine* **285**, 503–509 (2019).
261. Ridker, P. M. Anticytokine Agents. *Circulation research* **124**, 437–450 (2019).
262. LaRock, C. N., Todd, J., LaRock, D. L., Olson, J., O'Donoghue, A. J., *et al.* IL-1b is an innate immune sensor of microbial proteolysis. *Science Immunology* **1** (2016).
263. Smith, G. D. & Ebrahim, S. 'Mendelian randomization': Can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology* **32**, 1–22 (2003).
264. E., W. P. & Lock, R. H. Recent Progress in the Study of Variation, Heredity and Evolution. *Biometrika* **5**, 478 (1907).
265. Davies, N. M., Howe, L. J., Brumpton, B., Havdahl, A., Evans, D. M., *et al.* Within family Mendelian randomization studies. *Human Molecular Genetics* **28**, R170–R179 (2019).
266. Didelez, V. & Sheehan, N. Mendelian randomization as an instrumental variable approach to causal inference. *Statistical Methods in Medical Research* **16**, 309–330 (2007).
267. Davies, N. M., Holmes, M. V. & Davey Smith, G. Reading Mendelian randomisation studies: A guide, glossary, and checklist for clinicians. *BMJ (Online)* **362** (2018).

268. Schmidt, A. & Dudbridge, F. Mendelian randomization with Egger pleiotropy correction and weakly informative Bayesian priors. en. *Int J Epidemiol* **47**, 1217–1228. doi:[10.1093/ije/dyx254](https://doi.org/10.1093/ije/dyx254) (2018).
269. Holmes, M., Richardson, T. & Ference, B. Integrating genomics with biomarkers and therapeutic targets to invigorate cardiovascular drug development. en. *Nat Rev Cardiol* **18**, 435–453. doi:[10.1038/s41569-020-00493-1](https://doi.org/10.1038/s41569-020-00493-1). <https://doi.org/10.1038/s41569-020-00493-1> (2021).
270. King, E. A., Wade Davis, J. & Degner, J. F. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. *PLoS Genetics* **15** (2019).
271. Hughes, J. P., Rees, S. S., Kalindjian, S. B. & Philpott, K. L. Principles of early drug discovery. *British Journal of Pharmacology* **162**, 1239–1249 (2011).
272. Chong, M., Sjaarda, J., Pigeyre, M., Mohammadi-Shemirani, P., Lali, R., *et al.* Novel Drug Targets for Ischemic Stroke Identified Through Mendelian Randomization Analysis of the Blood Proteome. *Circulation* **140**, 819–830 (2019).
273. Gill, D., Georgakis, M. K., Koskeridis, F., Jiang, L., Feng, Q., *et al.* Use of Genetic Variants Related to Antihypertensive Drugs to Inform on Efficacy and Side Effects. *Circulation* **140**, 270–279 (2019).
274. Swerdlow, D. I. & Preiss, D. Genetic insights into statin-associated diabetes risk. *Current Opinion in Lipidology* **27**, 125–130 (2016).
275. Tanaka, T. & Kishimoto, T. The biology and medical implications of interleukin-6. *Cancer immunology research* **2**, 288–294 (2014).
276. Garbers, C. & Rose-John, S. Dissecting interleukin-6 classic- and trans-signaling in inflammation and cancer. *Methods in Molecular Biology* **1725**, 127–140 (2018).
277. Müller-Newen, G., Küster, A., Hemmann, U., Keul, R., Horsten, U., *et al.* Soluble IL-6 receptor potentiates the antagonistic activity of soluble gp130 on IL-6 responses. *Journal of immunology (Baltimore, Md. : 1950)* **161**, 6347–55 (1998).
278. Baran, P., Hansen, S., Waetzig, G. H., Akbarzadeh, M., Lamertz, L., *et al.* The balance of interleukin (IL)-6, IL-6soluble IL-6 receptor (sIL-6R), and IL-6sIL-6Rsgp130 complexes allows simultaneous classic and trans-signaling. *Journal of Biological Chemistry* **293**, 6762–6775 (2018).
279. Schett, G. Physiological effects of modulating the interleukin-6 axis. *Rheumatology (United Kingdom)* **57**, ii43–ii50 (2018).
280. Farahi, N., Paige, E., Balla, J., Prudence, E., Ferreira, R. C., *et al.* Neutrophil-mediated IL-6 receptor trans-signaling and the risk of chronic obstructive pulmonary disease and asthma. *Human Molecular Genetics* **26**, 1584–1596 (2017).
281. Garbers, C., Monhasery, N., Aparicio-Siegmund, S., Lokau, J., Baran, P., *et al.* The interleukin-6 receptor Asp358Ala single nucleotide polymorphism rs2228145 confers increased proteolytic conversion rates by ADAM proteases. *Biochimica et Biophysica Acta - Molecular Basis of Disease* **1842**, 1485–1494 (2014).
282. Jostock, T., Müllberg, J., Özbek, S., Atreya, R., Blinn, G., *et al.* Soluble gp130 is the natural inhibitor of soluble interleukin-6 receptor transsignaling responses. *European Journal of Biochemistry* **268**, 160–167 (2001).
283. Ference, B. A. Using Genetic Variants in the Targets of Lipid Lowering Therapies to Inform Drug Discovery and Development: Current and Future Treatment Options. *Clinical Pharmacology and Therapeutics* **105**, 568–581 (2019).

284. Nishimoto, N., Terao, K., Mima, T., Nakahara, H., Takagi, N., *et al.* Mechanisms and pathologic significances in increase in serum interleukin-6 (IL-6) and soluble IL-6 receptor after administration of an anti-IL-6 receptor antibody, tocilizumab, in patients with rheumatoid arthritis and Castleman disease. *Blood* **112**, 3959–3964 (2008).
285. Cai, T., Zhang, Y., Ho, Y. L., Link, N., Sun, J., *et al.* Association of interleukin 6 receptor variant with cardiovascular disease effects of interleukin 6 receptor blocking therapy: A phenome - Wide association study. *JAMA Cardiology* **3**, 849–857 (Sept. 2018).
286. Rosa, M., Chignon, A., Li, Z., Boulanger, M. C., Arsenault, B. J., *et al.* A Mendelian randomization study of IL6 signaling in cardiovascular diseases, immune-related disorders and longevity. *npj Genomic Medicine* **4**, 23 (2019).
287. Prins, B. P., Abbasi, A., Wong, A., Vaez, A., Nolte, I., *et al.* Investigating the Causal Relationship of C-Reactive Protein with 32 Complex Somatic and Psychiatric Outcomes: A Large-Scale Cross-Consortium Mendelian Randomization Study. *PLoS Medicine* **13** (2016).
288. Ligthart, S., Vaez, A., Vösa, U., Stathopoulou, M. G., de Vries, P. S., *et al.* Genome Analyses of >200,000 Individuals Identify 58 Loci for Chronic Inflammation and Highlight Pathways that Link Inflammation and Complex Disorders. *American Journal of Human Genetics* **103**, 691–706 (2018).
289. Keavney, B., Danesh, J., Parish, S., Palmer, A., Clark, S., *et al.* Fibrinogen and coronary heart disease: test of causality by 'Mendelian randomization'. *Int J Epidemiol* **35**, 935–943 (2006).
290. Sabater-Lleal, M., Huang, J., Chasman, D., Naitza, S., Dehghan, A., *et al.* Multiethnic meta-analysis of genome-wide association studies in >100 000 subjects identifies 23 fibrinogen-associated Loci but no strong evidence of a causal association between circulating fibrinogen and cardiovascular disease. *Circulation* **128**, 1310–1324 (2013).
291. Malik, R., Rannikmäe, K., Traylor, M., Georgakis, M. K., Sargurupremraj, M., *et al.* Genome-wide meta-analysis identifies 3 novel loci associated with stroke. *Annals of Neurology* **84**, 934–939 (2018).
292. Nelson, C. P., Goel, A., Butterworth, A. S., Kanoni, S., Webb, T. R., *et al.* Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nature Genetics* **49**, 1385–1391 (2017).
293. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., *et al.* UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Medicine* **12** (2015).
294. Woo, D., Falcone, G. J., Devan, W. J., Brown, W. M., Biffi, A., *et al.* Meta-analysis of genome-wide association studies identifies 1q22 as a susceptibility locus for intracerebral hemorrhage. *American Journal of Human Genetics* **94**, 511–521 (2014).
295. Yengo, L., Sidorenko, J., Kemper, K. E., Zheng, Z., Wood, A. R., *et al.* Meta-analysis of genome-wide association studies for height and body mass index in 700 000 individuals of European ancestry. *Human Molecular Genetics* **27**, 3641–3649 (2018).
296. NealeLab. UK Biobank GWAS Results. <http://www.nealelab.is/uk-biobank/>. (2019) (2019).

297. Nikpay, M., Goel, A., Won, H. H., Hall, L. M., Willenborg, C., *et al.* A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nature Genetics* **47**, 1121–1130 (2015).
298. Chen, Z., Lee, L., Chen, J., Collins, R., Wu, F., *et al.* Cohort profile: The Kadoorie Study of Chronic Disease in China (KSCDC). *International Journal of Epidemiology* **34**, 1243–1249 (2005).
299. Chen, Z., Chen, J., Collins, R., Guo, Y., Peto, R., *et al.* China Kadoorie Biobank of 0.5 million people: Survey methods, baseline characteristics and long-term follow-up. *International Journal of Epidemiology* **40**, 1652–1666 (2011).
300. Soininen, P., Kangas, A. J., Würtz, P., Suna, T. & Ala-Korpela, M. Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. *Circulation: Cardiovascular Genetics* **8**, 192–206 (2015).
301. Assarsson, E., Lundberg, M., Holmquist, G., Björkstén, J., Thorsen, S. B., *et al.* Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PLoS ONE* **9** (2014).
302. Kamath, S. & Lip, G. Y. Fibrinogen: Biochemistry, epidemiology and determinants. *QJM - Monthly Journal of the Association of Physicians* **96**, 711–729 (2003).
303. Ridker, P. M., Everett, B. M., Thuren, T., MacFadyen, J. G., Chang, W. H., *et al.* Antiinflammatory Therapy with Canakinumab for Atherosclerotic Disease. *New England Journal of Medicine* **377**, 1119–1131 (Sept. 2017).
304. Abd, T. T., Eapen, D. J., Bajpai, A., Goyal, A., Dollar, A., *et al.* The role of C-reactive protein as a risk predictor of coronary atherosclerosis: Implications from the JUPITER trial. *Current Atherosclerosis Reports* **13**, 154–161 (2011).
305. Agrawal, A., Hammond, D. J. & Singh, S. K. Atherosclerosis-related functions of C-reactive protein. *Cardiovascular & hematological disorders drug targets* **10**, 235–40 (2010).
306. Blake, G. J. & Ridker, P. M. C-reactive protein, subclinical atherosclerosis, and risk of cardiovascular events. *Arteriosclerosis, Thrombosis, and Vascular Biology* **22**, 1512–1513 (2002).
307. Green, D., Foiles, N., Chan, C., Schreiner, P. J. & Liu, K. Elevated fibrinogen levels and subsequent subclinical atherosclerosis: The CARDIA Study. *Atherosclerosis* **202**, 623–631 (2009).
308. Quaglia, L. A., Freitas, W. M., Soares, A. A., Santos, R. D., Nadruz, W., *et al.* C-reactive protein is independently associated with coronary atherosclerosis burden among octogenarians. *Aging Clinical and Experimental Research* **26**, 19–23 (2014).
309. Ridker, P. M. Targeting inflammatory pathways for the treatment of cardiovascular disease. *European Heart Journal* **35**, 540–543 (2014).
310. Zhang, Y., Zhu, C. G., Guo, Y. L., Xu, R. X., Li, S., *et al.* Higher fibrinogen level is independently linked with the presence and severity of new-onset coronary atherosclerosis among han Chinese population. *PLoS ONE* **9**, e113460 (2014).
311. Wakugawa, Y., Kiyohara, Y., Tanizaki, Y., Kubo, M., Ninomiya, T., *et al.* C-reactive protein and risk of first-ever ischemic and hemorrhagic stroke in a general Japanese population: The Hisayama study. *Stroke* **37**, 27–32 (2006).
312. Singh, G. M., Danaei, G., Farzadfar, F., Stevens, G. A., Woodward, M., *et al.* The age-specific quantitative effects of metabolic risk factors on cardiovascular diseases and diabetes: A pooled analysis. *PLoS ONE* **8** (2013).
313. Lenth, R. V. Least-squares means: The R package lsmeans. *Journal of Statistical Software* **69**, 33 (2016).

314. Viechtbauer, W. Conducting meta-analyses in R with the metafor. *Journal of Statistical Software* **36**, 1–48 (2010).
315. Plummer, M. Improved estimates of floating absolute risk. *Statistics in Medicine* **23**, 93–104 (2004).
316. Rosner, B., Willett, W. C. & Spiegelman, D. Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Statistics in Medicine* **8**, 1051–1069 (1989).
317. Danesh, J., Wheeler, J. G., Hirschfield, G. M., Eda, S., Eiriksdottir, G., *et al.* C-Reactive Protein and Other Circulating Markers of Inflammation in the Prediction of Coronary Heart Disease. *New England Journal of Medicine* **350**, 1387–1397 (2004).
318. Kaptoge, S., Di Angelantonio, E., Lowe, G., Pepys, M. B., Thompson, S. G., *et al.* C-reactive protein concentration and risk of coronary heart disease, stroke, and mortality: An individual participant meta-analysis. *The Lancet* **375**, 132–140 (2010).
319. Sacks, F. M. & Jensen, M. K. From high-density lipoprotein cholesterol to measurements of function: Prospects for the development of tests for high-density lipoprotein functionality in cardiovascular disease. *Arteriosclerosis, Thrombosis, and Vascular Biology* **38**, 487–499 (2018).
320. Levinson, S. S., Blevins, R. D., Smith, M. P., Rubinfire, M. & Maciejko, J. J. The relationship between apolipoprotein B-100 and low-density lipoprotein cholesterol after treatment with an HMG CoA reductase inhibitor. *American Journal of Clinical Pathology* **92**, 479–483 (1989).
321. Bittencourt, M. S., Fernandes, R. A., Ritt-Dias, R. M., Balagopal, B., Conceicao, R., *et al.* Longitudinal Changes in Physical Activity Levels Positively Impact High Sensitivity C-Reactive Protein Levels. *Journal of the American College of Cardiology* **71**, A1799 (2018).
322. Peralta, C. A., Katz, R., Sarnak, M. J., Ix, J., Fried, L. F., *et al.* Cystatin C identifies chronic kidney disease patients at higher risk for complications. *Journal of the American Society of Nephrology* **22**, 147–155 (2011).
323. Liu, X., Bragg, F., Yang, L., Kartsonaki, C., Guo, Y., *et al.* Smoking and smoking cessation in relation to risk of diabetes in Chinese men and women: a 9-year prospective study of 0·5 million people. *The Lancet Public Health* **3**, e167–e176 (2018).
324. Im, P. K., Millwood, I. Y., Guo, Y., Du, H., Chen, Y., *et al.* Patterns and trends of alcohol consumption in rural and urban areas of China: Findings from the China Kadoorie Biobank. *BMC Public Health* **19** (2019).
325. Stoltzfus, J. C. Logistic regression: A brief primer. *Academic Emergency Medicine* **18**, 1099–1104 (2011).
326. Iso, H., Cui, R., Date, C., Kikuchi, S. & Tamakoshi, A. C-reactive protein levels and risk of mortality from cardiovascular disease in Japanese: The JACC Study. *Atherosclerosis* **207**, 291–297 (2009).
327. Et al. Ridker, Paul M. Rosuvastatin to Prevent Vascular Events in Men and Women with Elevated CRP. *New England Journal of Medicine* (2008).
328. Zhao, D., Liu, J., Wang, M., Zhang, X. & Zhou, M. Epidemiology of cardiovascular disease in China: current features and implications. *Nature Reviews Cardiology* **16**, 203–212 (2019).

329. Kaptoge, S., White, I. R., Thompson, S. G., Wood, A. M., Lewington, S., *et al.* Associations of plasma fibrinogen levels with established cardiovascular disease risk factors, inflammatory markers, and other characteristics: Individual participant meta-analysis of 154,211 adults in 31 prospective studies. *American Journal of Epidemiology* **166**, 867–879 (2007).
330. Franceschi, C. & Campisi, J. Chronic inflammation (Inflammaging) and its potential contribution to age-associated diseases. *Journals of Gerontology - Series A Biological Sciences and Medical Sciences* **69**, S4–S9 (2014).
331. Calder, P. C., Bosco, N., Bourdet-Sicard, R., Capuron, L., Delzenne, N., *et al.* Health relevance of the modification of low grade inflammation in ageing (inflammageing) and the role of nutrition. *Ageing Research Reviews* **40**, 95–119 (2017).
332. Krobot, K., Hense, H. W., Cremer, P., Eberle, E. & Keil, U. Determinants of plasma fibrinogen: Relation to body weight, waist-to-hip ratio, smoking, alcohol, age, and sex: Results from the second MONICA Augsburg survey, 1989–1990. *Arteriosclerosis and Thrombosis* **12**, 780–788 (1992).
333. Prisco, D., Fedi, S., Brunelli, T., Cellal, A. P., Hagi, M. I., *et al.* Fibrinogen and Factor Vllag in Healthy Adolescents: The Floren-teen (Florence Teenager) Study. *Thrombosis and Haemostasis* **75**, 778–781 (1996).
334. Ahmadi-Abhari, S., Luben, R. N., Wareham, N. J. & Khaw, K. T. Distribution and determinants of C-reactive protein in the older adult population: European prospective investigation into cancer-norfolk study. *European Journal of Clinical Investigation* **43**, 899–911 (2013).
335. Zhao, Y., Wang, R., Ma, X., Yan, X., Zhang, Z., *et al.* Distribution of C-reactive protein and its association with cardiovascular risk factors in a population-based sample of Chinese. *Disease Markers* **28**, 333–342 (2010).
336. Wang, Z., Wang, X., Chen, Z., Zhang, L. & Zhu, M. Distribution of high-sensitivity C-reactive protein and its relationship with other cardiovascular risk factors in the middle-aged Chinese population. *International Journal of Environmental Research and Public Health* **13** (2016).
337. Ye, X., Yu, Z., Li, H., Franco, O. H., Liu, Y., *et al.* Distributions of C-Reactive Protein and its Association With Metabolic Syndrome in Middle-Aged and Older Chinese People. *Journal of the American College of Cardiology* **49**, 1798–1805 (2007).
338. Shah, T., Newcombe, P., Smeeth, L., Addo, J., Casas, J. P., *et al.* Ancestry as a determinant of mean population C-Reactive protein values: Implications for cardiovascular risk prediction. *Circulation: Cardiovascular Genetics* **3**, 436–444 (2010).
339. Hsia, S. H. Non-HDL cholesterol: Into the spotlight. *Diabetes Care* **26**, 240–242 (2003).
340. Fogari, R., Zoppi, A., Marasi, G., Vanasia, A. & Villa, G. Associations between plasma fibrinogen levels and cardiovascular risk factors in hypertensive men. *European Journal of Cardiovascular Prevention & Rehabilitation* **1**, 341–345 (1994).
341. Stec, J. J., Silbershatz, H., Tofler, G. H., Matheney, T. H., Sutherland, P., *et al.* Association of fibrinogen with cardiovascular risk factors and cardiovascular disease the Framingham Offspring Population. *Circulation* **102**, 1634–1638 (2000).

342. Robert, J., Lehner, M., Frank, S., Perisa, D., Von Eckardstein, A., *et al.* Interleukin 6 stimulates endothelial binding and transport of high-density lipoprotein through induction of endothelial lipase. *Arteriosclerosis, Thrombosis, and Vascular Biology* **33**, 2699–2706 (2013).
343. Zuliani, G., Volpato, S., Blè, A., Bandinelli, S., Corsi, A. M., *et al.* High interleukin-6 plasma levels are associated with low HDL-C levels in community-dwelling older adults: The InChianti study. *Atherosclerosis* **192**, 384–390 (2007).
344. Jin, W., Millar, J. S., Broedl, U., Glick, J. M. & Rader, D. J. Inhibition of endothelial lipase causes increased HDL cholesterol levels in vivo. *Journal of Clinical Investigation* **111**, 357–362 (2003).
345. Frank, P. G. & Marcel, Y. L. Apolipoprotein A-I: Structure-function relationships. *Journal of Lipid Research* **41**, 853–872 (2000).
346. Soeters, P. B., Wolfe, R. R. & Shenkin, A. Hypoalbuminemia: Pathogenesis and Clinical Significance. *Journal of Parenteral and Enteral Nutrition* **43**, 181–193 (2019).
347. Gruppen, E. G., Riphagen, I. J., Connelly, M. A., Otvos, J. D., Bakker, S. J., *et al.* GlycA, a pro-inflammatory glycoprotein biomarker, and incident cardiovascular disease: Relationship with C-reactive protein and renal function. *PLoS ONE* **10**, e0139057 (2015).
348. Keller, C., Katz, R., Cushman, M., Fried, L. F. & Shlipak, M. Association of kidney function with inflammatory and procoagulant markers in a diverse cohort: A cross-sectional analysis from the Multi-Ethnic Study of Atherosclerosis (MESA). *BMC Nephrology* **9** (2008).
349. Kerner, A., Avizohar, O., Sella, R., Bartha, P., Zinder, O., *et al.* Association between elevated liver enzymes and C-reactive protein: Possible hepatic contribution to systemic inflammation in the metabolic syndrome. *Arteriosclerosis, Thrombosis, and Vascular Biology* **25**, 193–197 (2005).
350. Colbert, L. H., Visser, M., Simonsick, E. M., Tracy, R. P., Newman, A. B., *et al.* Physical activity, exercise, and inflammatory markers in older adults: Findings from the health, aging and body composition study. *Journal of the American Geriatrics Society* **52**, 1098–1104 (2004).
351. Hunter, K. A., Garlick, P. J., Broom, L., Anderson, S. E. & McNurlan, M. A. Effects of smoking and abstention from smoking on fibrinogen synthesis in humans. *Clinical Science* **100**, 459–465 (2001).
352. Millwood, I. Y., Li, L., Smith, M., Guo, Y., Yang, L., *et al.* Alcohol consumption in 0.5 million people from 10 diverse regions of china: Prevalence, patterns and socio-demographic and health-related correlates. *International Journal of Epidemiology* (2013).
353. Kaptoge, S., White I, R., *et al.* Associations of Plasma Fibrinogen Levels with Established Cardiovascular Disease Risk Factors, Inflammatory Markers, and Other Characteristics: Individual Participant Meta-Analysis of 154,211 Adults in 31 Prospective Studies. *American Journal of Epidemiology* **166**, 867–879 (2007).
354. Nelson, C., Hamby, S. & Saleheen, D. Genetically determined height and coronary artery disease. *en. N Engl J Med* **372**, 1608–1618. doi:[10.1056/NEJMoa1404881](https://doi.org/10.1056/NEJMoa1404881) (2015).
355. Nordestgaard, B., Palmer, T., Benn, M., *et al.* The Effect of Elevated Body Mass Index on Ischemic Heart Disease Risk: Causal Estimates from a Mendelian

- Randomisation Approach. en. *PLoS Med* **9**, 5.
doi:<https://doi.org/10.1371/journal.pmed.1001212> (2012).
356. Linden, A. B., Clarke, R., Hopewell, J. C., *et al.* Genetic associations of adult height with risk of cardioembolic and other subtypes of ischaemic stroke: a Mendelian randomisation study in multiple ancestries. *medRxiv*.
doi:<https://doi.org/10.1101/2020.08.26.20174086> (2020).
357. Lacey, B., Lewington, S., Clarke, R., Kong, X. L., Chen, Y., *et al.* Age-specific association between blood pressure and vascular and non-vascular chronic diseases in 0·5 million adults in China: a prospective cohort study. *The Lancet Global Health* (2018).
358. Tsai, C. F., Anderson, N., Thomas, B. & Sudlow, C. L. Comparing risk factor profiles between intracerebral hemorrhage and ischemic stroke in Chinese and white populations: Systematic review and meta-analysis. *PLoS ONE* **11**, e0151743 (2016).
359. Elliott, P., Chambers, J. C., Zhang, W., Clarke, R., Hopewell, J. C., *et al.* Genetic loci associated with C-reactive protein levels and risk of coronary heart disease. *JAMA - Journal of the American Medical Association* **302**, 37–48 (2009).
360. Zacho, J., Tybjaerg-Hansen, A., Jensen, J. S., Grande, P., Sillesen, H., *et al.* Genetically Elevated C-Reactive Protein and Ischemic Vascular Disease. *New England Journal of Medicine* **359**, 1897–1908 (2008).
361. Schaper, F. & Rose-John, S. Interleukin-6: Biology, signaling and strategies of blockade. *Cytokine and Growth Factor Reviews* **26**, 475–487 (2015).
362. Yamasaki, K., Taga, T., Hirata, Y., Yawata, H., Kawanishi, Y., *et al.* Cloning and expression of the human interleukin-6 (BSF-2/IFN β 2) receptor. *Science* **241**, 825–828 (1988).
363. Taga, T., Hibi, M., Hirata, Y., Yamasaki, K., Yasukawa, K., *et al.* Interleukin-6 triggers the association of its receptor with a possible signal transducer, gp130. *Cell* **58**, 573–581 (1989).
364. Kishimoto, T. IL-6: From its discovery to clinical applications. *International Immunology* **22**, 347–352 (2010).
365. Oberg, H. H., Wesch, D., Grüssel, S. S., Rose-John, S. & Kabelitz, D. Differential expression of CD126 and CD130 mediates different STAT-3 phosphorylation in CD4+CD25- and CD25high regulatory T cells. *International Immunology* **18**, 555–563 (2006).
366. Riethmueller, S., Somasundaram, P., Ehlers, J. C., Hung, C. W., Flynn, C. M., *et al.* Proteolytic Origin of the Soluble Human IL-6R In Vivo and a Decisive Role of N-Glycosylation. *PLoS Biology* **15** (2017).
367. Rose-John, S. Il-6 trans-signaling via the soluble IL-6 receptor: Importance for the proinflammatory activities of IL-6. *International Journal of Biological Sciences* **8**, 1237–1247 (2012).
368. Burgess, S. & Thompson, S. G. Use of allele scores as instrumental variables for Mendelian randomization. *International Journal of Epidemiology* **42**, 1134–1144 (2013).
369. Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N. & Golani, I. Controlling the false discovery rate in behavior genetics research. *Behavioural Brain Research* **125**, 279–284 (2001).
370. Gauderman, W. J. Sample size requirements for association studies of gene-gene interaction. *American Journal of Epidemiology* **155**, 478–484 (2002).

371. Sun, B. B., Maranville, J. C., Peters, J. E., Stacey, D., Staley, J. R., *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).
372. Benner, C., Spencer, C. C., Havulinna, A. S., Salomaa, V., Ripatti, S., *et al.* FINEMAP: Efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493–1501 (2016).
373. Burgess, S. & Thompson, S. G. Avoiding bias from weak instruments in mendelian randomization studies. *International Journal of Epidemiology* **40**, 755–764 (2011).
374. Franceschini, N., Giambartolomei, C., de Vries, P. S., Finan, C., Bis, J. C., *et al.* GWAS and colocalization analyses implicate carotid intima-media thickness and carotid plaque loci in cardiovascular outcomes. *Nature Communications* **9** (2018).
375. Isobe, A., Sawada, K., Kinose, Y., Ohyagi-Hara, C., Nakatsuka, E., *et al.* Interleukin 6 receptor is an independent prognostic factor and a potential therapeutic target of ovarian cancer. *PLoS ONE* **10** (2015).
376. Wouters, M. C., Dijkgraaf, E. M., Kuijjer, M. L., Jordanova, E. S., Hollema, H., *et al.* Interleukin-6 receptor and its ligand interleukin-6 are opposite markers for survival and infiltration with mature myeloid cells in ovarian cancer. *OncoImmunology* **3**, 1–13 (2014).
377. Ferreira, R. C., Freitag, D. F., Cutler, A. J., Howson, J. M., Rainbow, D. B., *et al.* Functional IL6R 358Ala Allele Impairs Classical IL-6 Receptor Signaling and Influences Risk of Diverse Inflammatory Diseases. *PLoS Genetics* **9** (2013).
378. Peters, M., Jacobs, S., Ehlers, M., Vollmer, P., Müllberg, J., *et al.* The function of the soluble interleukin 6 (IL-6) receptor in vivo: Sensitization of human soluble IL-6 receptor transgenic mice towards IL-6 and prolongation of the plasma half-life of IL-6. *Journal of Experimental Medicine* **183**, 1399–1406 (1996).
379. Zhou, H., Wang, X., Ko, W. K. & Wong, A. O. Evidence for a novel intrapituitary autocrine/paracrine feedback loop regulating growth hormone synthesis and secretion in grass carp pituitary cells by functional interactions between gonadotrophs and somatotrophs. *Endocrinology* **145**, 5548–5559 (2004).
380. Fairfax, B. P., Humburg, P., Makino, S., Naranbhai, V., Wong, D., *et al.* Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* **343** (2014).
381. Ridker, P., Devalaraja, M., Baeres, F., Engelmann, M., *et al.* IL-6 inhibition with ziltivekimab in patients at high atherosclerotic risk (RESCUE): a double-blind, randomised, placebo-controlled, phase 2 trial. *Lancet* **29** (May 2021).
382. Levoe, A., Dam, J., Ayoub, M. A., Guillaume, J. L. & Jockers, R. Do orphan G-protein-coupled receptors have ligand-independent functions? New insights from receptor heterodimers. *EMBO Reports* **7**, 1094–1098 (2006).
383. HP Jr, A. *et al.* Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. en. *Stroke* **24(1):35-41**.
384. Sarwar, N., Butterworth, A. S., Freitag, D. F., Gregson, J., Willeit, P., *et al.* Interleukin-6 receptor pathways in coronary heart disease: A collaborative meta-analysis of 82 studies. *The Lancet* **379**, 1205–1213 (2012).
385. Cai, T., Zhang, Y., Ho, Y. L., Link, N., Sun, J., *et al.* Association of interleukin 6 receptor variant with cardiovascular disease effects of interleukin 6 receptor blocking therapy: A phenome - Wide association study. *JAMA Cardiology* **3**, 849–857 (2018).

386. Esmon, C. T. The impact of the inflammatory response on coagulation. *Thrombosis Research* **114**, 321–327 (2004).
387. Fuster, V., Moreno, P. R., Fayad, Z. A., Corti, R. & Badimon, J. J. Atherothrombosis and high-risk plaque: Part I: Evolving concepts. *Journal of the American College of Cardiology* **46**, 937–954 (2005).
388. Brumpton, B., Sanderson, E., Heilbron, K., Hartwig, F. P., Harrison, S., *et al.* Avoiding dynastic, assortative mating, and population stratification biases in Mendelian randomization through within-family analyses. *Nature Communications* **11** (2020).
389. Richardson, T. G., Sanderson, E., Palmer, T. M., Ala-Korpela, M., Ference, B. A., *et al.* Apolipoprotein B underlies the causal relationship of circulating blood lipids with coronary heart disease. *Apolipoprotein B underlies the causal relationship of circulating blood lipids with coronary heart disease*, 19004895 (2019).
390. Kawashiri, S. Y., Kawakami, A., Yamasaki, S., Imazato, T., Iwamoto, N., *et al.* Effects of the anti-interleukin-6 receptor antibody, tocilizumab, on serum lipid levels in patients with rheumatoid arthritis. *Rheumatology International* **31**, 451–456 (2011).
391. Davies, N. M., Holmes, M. V. & Davey Smith, G. Reading Mendelian randomisation studies: A guide, glossary, and checklist for clinicians. *BMJ (Online)* (2018).
392. Holmes, M. V., Ala-Korpela, M. & Smith, G. D. Mendelian randomization in cardiometabolic disease: Challenges in evaluating causality. *Nature Reviews Cardiology* **14**, 577–599 (2017).
393. Mokry, L. E., Ahmad, O., Forgetta, V., Thanassoulis, G. & Richards, J. B. Mendelian randomisation applied to drug development in cardiovascular disease: A review. *Journal of Medical Genetics* **52**, 71–79 (2015).
394. Zerbino, D. R., Johnson, N., Juetteman, T., Sheppard, D., Wilder, S. P., *et al.* Ensembl regulation resources. *Database* **2016** (2016).
395. Joshi, A. & Mayr, M. In Aptamers They Trust: Caveats of the SOMAscan Biomarker Discovery Platform from SomaLogic. *Circulation* **138**, 2482–2485 (2018).
396. Campbell, I. L., Ertz, M., Lim, S. L., Frausto, R., May, U., *et al.* Trans-signaling is a dominant mechanism for the pathogenic actions of interleukin-6 in the brain. *Journal of Neuroscience* **34**, 2503–2513 (Feb. 2014).
397. Garbers, C., Thaiss, W., Jones, G. W., Waetzig, G. H., Lorenzen, I., *et al.* Inhibition of classic signaling is a novel function of soluble glycoprotein 130 (sgp130), which is controlled by the ratio of interleukin 6 and soluble interleukin 6 receptor. *Journal of Biological Chemistry* **286**, 42959–42970 (2011).
398. Bowden, J., Smith, G. D. & Burgess, S. Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. *International Journal of Epidemiology* **44**, 512–525 (2015).
399. Hemani, G., Bowden, J. & Davey Smith, G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum Mol Genet* **27** (2018).
400. Schmidt, A. F., Swerdlow, D. I., Holmes, M. V., Patel, R. S., Fairhurst-Hunter, Z., *et al.* PCSK9 genetic variants and risk of type 2 diabetes: a mendelian randomisation study. *The Lancet Diabetes and Endocrinology* **5**, 97–105 (2017).
401. Swerdlow, D. I., Preiss, D., Kuchenbaecker, K. B., Holmes, M. V., Engmann, J. E., *et al.* HMG-coenzyme A reductase inhibition, type 2 diabetes, and bodyweight:

- Evidence from genetic analysis and randomised trials. *The Lancet* **385**, 351–361 (2015).
402. Sattar, N., Preiss, D. & Murray, H. M. Statins and risk of incident diabetes: A collaborative meta-analysis of randomised statin trials. *Revista Portuguesa de Cardiologia* **29**, 1077–1078 (2010).
403. Robson, J. C., Kiran, A., Maskell, J., Hutchings, A., Arden, N., *et al.* The relative risk of aortic aneurysm in patients with giant cell arteritis compared with the general population of the UK. *Annals of the Rheumatic Diseases* **74**, 129–135 (2015).
404. FDA approves first drug to specifically treat giant cell arteritis (accessed 5 Dec 2021). <https://www.fda.gov/news-events/press-announcements/fda-approves-first-drug-specifically-treat-giant-cell-arteritis> (2018).
405. Garbers, C., Heink, S., Korn, T. & Rose-John, S. Interleukin-6: Designing specific therapeutics for a complex cytokine. *Nature Reviews Drug Discovery* **17**, 395–412 (2018).
406. Nguyen, P. A., Born, D. A., Deaton, A. M., Nioi, P. & Ward, L. D. Phenotypes associated with genes encoding drug targets are predictive of clinical trial side effects. *Nature Communications* **10**, 1579 (2019).
407. Vainer, N., Dehlendorff, C. & Johansen, J. S. Systematic literature review of IL-6 as a biomarker or treatment target in patients with gastric, bile duct, pancreatic and colorectal cancer. *Oncotarget* **9**, 29820–29841 (2018).
408. Ashizawa, T., Okada, R., Suzuki, Y., Takagi, M., Yamazaki, T., *et al.* Clinical significance of interleukin-6 (IL-6) in the spread of gastric cancer: Role of IL-6 as a prognostic factor. *Gastric Cancer* **8**, 124–131 (2005).
409. Zhao, Z. F., Li, J. X., Ye, R., Wu, X., Gao, L. L., *et al.* Interleukin-6 as a potential molecular target in esophageal squamous cell carcinoma. *Oncology Letters* **11**, 925–932 (2016).
410. Waldner, M. J., Foersch, S. & Neurath, M. F. Interleukin-6 - A key regulator of colorectal cancer development. *International Journal of Biological Sciences* (2012).
411. Wong, V. W. S., Yu, J., Cheng, A. S. L., Wong, G. L. H., Chan, H. Y., *et al.* High serum interleukin-6 level predicts future hepatocellular carcinoma development in patients with chronic hepatitis B. *International Journal of Cancer* **124**, 2766–2770 (2009).
412. Brenner, D. R., Fanidi, A., Grankvist, K., Muller, D. C., Brennan, P., *et al.* Inflammatory Cytokines and Lung Cancer Risk in 3 Prospective Studies. *American Journal of Epidemiology* **185**, 86–95 (2017).
413. Kozłowski, L., Zakrzewska, I., Tokajuk, P. & Wojtukiewicz, M. Z. Concentration of interleukin-6 (IL-6), interleukin-8 (IL-8) and interleukin-10 (IL-10) in blood serum of breast cancer patients. *Roczniki Akademii Medycznej w Białymstoku (1995)* **48**, 82–84 (2003).
414. Rincon, M. & Irvin, C. G. Role of IL-6 in asthma and other inflammatory pulmonary diseases. *International Journal of Biological Sciences* **8**, 1281–1290 (2012).
415. King, P. T. Inflammation in chronic obstructive pulmonary disease and its role in cardiovascular disease and lung cancer. *Clinical and Translational Medicine* **4** (2015).
416. Bacci, M. R., Leme, R. C., Zing, N. P., Murad, N., Adami, F., *et al.* IL-6 and TNF- α serum levels are associated with early death in community-acquired

- pneumonia patients. *Brazilian Journal of Medical and Biological Research* **48**, 427–432 (2015).
417. Yokoyama, A., Kohno, N., Fujino, S., Hamada, H., Inoue, Y., *et al.* Circulating interleukin-6 levels in patients with bronchial asthma. *American Journal of Respiratory and Critical Care Medicine* **151**, 1354–1358 (1995).
418. Martinez, A. N., Mehra, S. & Kaushal, D. Role of interleukin 6 in innate immunity to mycobacterium tuberculosis infection. *Journal of Infectious Diseases* **207**, 1253–1261 (2013).
419. Kristiansen, O. P. & Mandrup-Poulsen, T. Interleukin-6 and diabetes: The good, the bad, or the indifferent? *Diabetes* **54** (2005).
420. Haapakoski, R., Mathieu, J., Ebmeier, K. P., Alenius, H. & Kivimäki, M. Cumulative meta-analysis of interleukins 6 and 1 β , tumour necrosis factor α and C-reactive protein in patients with major depressive disorder. *Brain, Behavior, and Immunity* **49**, 206–215 (2015).
421. Jain, S., Gautam, V. & Naseem, S. Acute-phase proteins: As diagnostic tool. *Journal of Pharmacy and Bioallied Sciences* **3**, 118–127 (2011).
422. Wannamethee, S. G., Shaper, A. G. & Perry, I. J. Serum creatinine concentration and risk of cardiovascular disease: A possible marker for increased risk of stroke. *Stroke* **28**, 557–563 (1997).
423. Keller, T., Messow, C. M., Lubos, E., Nicaud, V., Wild, P. S., *et al.* Cystatin C and cardiovascular mortality in patients with coronary artery disease and normal or mildly reduced kidney function: Results from the AtheroGene study. *European Heart Journal* **30**, 314–320 (2009).
424. Kim, Y. S., Guevara, J. P., Kim, K. M., Choi, H. K., Heitjan, D. F., *et al.* Hyperuricemia and coronary heart disease: A systematic review and meta-analysis. *Arthritis Care and Research* **62**, 170–180 (2010).
425. Van der Laan, S. W., Fall, T., Soumaré, A., Teumer, A., Sedaghat, S., *et al.* Cystatin C and Cardiovascular Disease: A Mendelian Randomization Study. *Journal of the American College of Cardiology* **68**, 934–945 (2016).
426. Kleber, M. E., Delgado, G., Grammer, T. B., Silbernagel, G., Huang, J., *et al.* Uric acid and cardiovascular events: A Mendelian randomization study. *Journal of the American Society of Nephrology* **26**, 2831–2838 (2015).
427. Efstathiadou, A., Gill, D., McGrane, F., Quinn, T. & Dawson, J. Genetically determined uric acid and the risk of cardiovascular and neurovascular diseases: A mendelian randomization study of outcomes investigated in randomized trials. *Journal of the American Heart Association* **8** (2019).
428. Hasegawa, S., Ichiyama, T., Sonaka, I., Ohsaki, A., Okada, S., *et al.* Cysteine, histidine and glycine exhibit anti-inflammatory effects in human coronary arterial endothelial cells. *Clinical and Experimental Immunology* **167**, 269–274 (2012).
429. Assarsson, E. *et al.* Homogenous 96-Plex PEA Immunoassay Exhibiting High Sensitivity, Specificity, and Excellent Scalability. *PLOS ONE* **9**, 1–11. <https://doi.org/10.1371/journal.pone.0095192> (Apr. 2014).
430. Guo, Y., Xu, F., Lu, T., Duan, Z. & Zhang, Z. Interleukin-6 signaling pathway in targeted therapy for cancer. *Cancer Treatment Reviews* **38**, 904–910 (2012).
431. Karakasheva, T. A., Lin, E. W., Tang, Q., Qiao, E., Waldron, T. J., *et al.* IL-6 mediates cross-talk between tumor cells and activated fibroblasts in the tumor microenvironment. *Cancer Research* **78**, 4957–4970 (2018).

432. Zhou, W., Nielsen, J. B., Fritsche, L. G., Dey, R., Gabrielsen, M. E., *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nature Genetics* **50**, 1335–1341 (2018).
433. Buniello, A., MacArthur, J. A., Cerezo, M., Harris, L. W., Hayhurst, J., *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research* **47**, D1005–D1012 (2019).
434. Tanikawa, C., Kamatani, Y., Takahashi, A., Momozawa, Y., Leveque, K., *et al.* GWAS identifies two novel colorectal cancer loci at 16q24.1 and 20q13.12. *Carcinogenesis* **39**, 652–660 (2018).
435. Tian, G., Mi, J., Wei, X., Zhao, D., Qiao, L., *et al.* Circulating interleukin-6 and cancer: A meta-analysis using Mendelian randomization. *Scientific Reports* **5** (2015).
436. Wei, J., Xiong, X. F., Lin, Y. H., Zheng, B. X. & Cheng, D. Y. Association between serum interleukin-6 concentrations and chronic obstructive pulmonary disease: A systematic review and meta-analysis. *PeerJ* **2015**, e1199 (2015).
437. Van Durme, Y. M., Lahousse, L., Verhamme, K. M., Stolk, L., Eijgelsheim, M., *et al.* Mendelian randomization study of interleukin-6 in chronic obstructive pulmonary disease. *Respiration* **82**, 530–538 (2011).
438. Gomes, K. B. IL-6 and type 1 diabetes mellitus: T cell responses and increase in IL-6 receptor surface expression. *eng. Annals of Translational Medicine* **5**, 16 (Jan. 2017).
439. Choy, E. H. & Calabrese, L. H. Neuroendocrine and neurophysiological effects of interleukin 6 in rheumatoid arthritis. *Rheumatology (United Kingdom)* **57**, 1885–1895 (2018).
440. Khandaker, G. M., Zammit, S., Burgess, S., Lewis, G. & Jones, P. B. Association between a functional interleukin 6 receptor genetic variant and risk of depression and psychosis in a population-based birth cohort. *Brain, Behavior, and Immunity* **69**, 264–272 (2018).
441. Khandaker, G. M., Oltean, B. P., Kaser, M., Dibben, C. R., Ramana, R., *et al.* Protocol for the insight study: A randomised controlled trial of singledose tocilizumab in patients with depression and low-grade inflammation. *BMJ Open* **8** (2018).
442. Jain, S., Gautam, V. & Naseem, S. Acute-phase proteins: As diagnostic tool. *Journal of Pharmacy and Bioallied Sciences* (2011).
443. Bradley, R. D., Fitzpatrick, A. L., Jacobs, D. R., Lee, D. H., Swords Jenny, N., *et al.* Associations between γ -glutamyltransferase (GGT) and biomarkers of atherosclerosis: The multi-ethnic study of atherosclerosis (MESA). *Atherosclerosis* **233**, 387–393 (2014).
444. Dinicolantonio, J. J., McCarty, M. F. & Okeefe, J. H. Role of dietary histidine in the prevention of obesity and metabolic syndrome. *Open Heart* **5** (2018).
445. Haynes, W. in *Encyclopedia of Systems Biology* (eds Dubitzky, W., Wolkenhauer, O., Cho, K.-H. & Yokota, H.) 78–78 (Springer New York, New York, NY, 2013). doi:[10.1007/978-1-4419-9863-7_1215](https://doi.org/10.1007/978-1-4419-9863-7_1215).
446. Sabater-Lleal, M., Huang, J., Chasman, D., Naitza, S., Dehghan, A., *et al.* Multiethnic meta-analysis of genome-wide association studies in >100 000 subjects identifies 23 fibrinogen-associated loci but no strong evidence of a causal

- association between circulating fibrinogen and cardiovascular disease. *Circulation* **128**, 1310–1324 (2013).
447. Mantovani, A., Locati, M., Vecchi, A., Sozzani, S. & Allavena, P. Decoy receptors: a strategy to regulate inflammatory cytokines and chemokines. en. *Trends Immunol* **22(6):328-36** (2001).
448. Schiff, M. H., Kremer, J. M., Jahreis, A., Vernon, E., Isaacs, J. D., *et al.* Integrated safety in tocilizumab clinical trials. *Arthritis Research and Therapy* **13** (2011).
449. Nishimoto, N., Sasai, M., Shima, Y., Nakagawa, M., Matsumoto, T., *et al.* Improvement in Castleman's disease by humanized anti-interleukin-6 receptor antibody therapy. *Blood* **95**, 56–61 (2000).
450. Nishimoto, N., Kanakura, Y., Aozasa, K., Johkoh, T., Nakamura, M., *et al.* Humanized anti-interleukin-6 receptor antibody treatment of multicentric Castleman disease. *Blood* **106**, 2627–2632 (2005).
451. Ridker, P. M., MacFadyen, J. G., Thuren, T. & Libby, P. Residual inflammatory risk associated with interleukin-18 and interleukin-6 after successful interleukin-1b inhibition with canakinumab: Further rationale for the development of targeted anti-cytokine therapies for the treatment of atherothrombosis. *European Heart Journal* **41**, 2153–2163 (2020).
452. Quach, H., Rotival, M., Pothlichet, J., Loh, Y. H. E., Dannemann, M., *et al.* Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. *Cell* **167**, 643–656.e17 (2016).
453. Aguet, F., Brown, A. A., Castel, S. E., Davis, J. R., He, Y., *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
454. Nédélec, Y., Sanz, J., Baharian, G., Szpiech, Z. A., Pacis, A., *et al.* Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. *Cell* **167**, 657–669.e21 (2016).
455. Alasoo, K., Rodrigues, J., Mukhopadhyay, S., Knights, A. J., Mann, A. L., *et al.* Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. *Nature Genetics* **50**, 424–431 (2018).
456. Grundberg, E., Small, K. S., Hedman, Å. K., Nica, A. C., Buil, A., *et al.* Mapping cis-and trans-regulatory effects across multiple tissues in twins. *Nature Genetics* **44**, 1084–1089 (2012).
457. Momozawa, Y., Dmitrieva, J., Théâtre, E., Deffontaine, V., Rahmouni, S., *et al.* IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes. *Nature Communications* **9** (2018).