Investigating genetic determinants of liver disease and its associations with cardiovascular diseases

Constantinos Athos Parisinos 17036338

This dissertation is submitted for the degree of Doctor of Philosophy, UCL

31 September 2022

Institute of Health Informatics

I dedicate this PhD to my friends and family that have helped me on this journey.

Declaration

'I, Constantinos Parisinos, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.'

Excerpts of this thesis have been published in the following academic publications. *joint first authors +corresponding author

Teo K, Abeyesekera KWM, Adams L, … [n =44], **Parisinos CA**, et al. Rs641738C>T near MBOAT7 is associated with liver fat, ALT and fibrosis in NAFLD: A meta-analysis. *J Hepatol* 2021; 74(1):20-30

Parisinos CA*+, Wilman HR*, Thomas EL, et al. Genome wide and Mendelian randomisation studies of liver MRI yield insights into the pathogenesis of steatohepatitis. *J Hepatol* 2020; 73(2):241-252

Wilman HR*, Parisinos CA*+, Atabaki Pasdar N, et al. Genetic studies of abdominal MRI data identify genes regulating hepcidin as major determinants of liver iron concentration. *J Hepatol* 2019; 71(3):594-602

Kuan V, Denaxas S, Gonzales-Izeuierdo A, [n=6], **Parisinos C**, et al. A chronological map of 308 physical and mental health conditions from 4 million individuals in the English National Health Service. *Lancet Digital Health* 2019; 1(2): E63-67

Thesis abstract

Background

Dramatic modifications in lifestyle have given rise to an epidemic in chronic liver diseases, predominantly driven by non-alcoholic fatty liver disease (NAFLD). The more severe NAFLD phenotypes are associated with elevated liver iron, inflammation (steatohepatitis), scarring and liver failure (fibrosis, cirrhosis), and possibly with cardiovascular diseases (CVDs); genetic and population studies of these phenotypes and their links to CVDs have been limited.

Aims

1) Investigate the genetic susceptibility underlying liver MRI phenotypes (iron and corrected T1 (cT1), a steatohepatitis proxy) and explore associations with other cardiometabolic traits.

2) Investigate whether liver fibrosis is an independent risk factor for CVDs.

Methods

We carried out genome-wide association studies (GWASs) of liver MRI phenotypes (iron (N = 8,289), and corrected T1 (a steatohepatitis proxy, N = 14,440)) in UK Biobank. We used genetics to investigate causality with other traits.

We calculated a FIB-4 score (a validated non-invasive scoring system that predicts liver fibrosis) in 44,956 individuals in the UK and investigated its association with the incidence of

five CVDs (ischaemic stroke, myocardial infarction, heart failure, peripheral arterial disease, atrial fibrillation (AF)).

Results

Three genetic variants known to influence hepcidin regulation (rs1800562 (C282Y) and rs1799945 (H63D) in *HFE,* rs855791 (V736A) in *TMPRSS6*) were associated with liver iron ($p < 5 \times 10^{-8}$). Mendelian randomisation provided evidence that central obesity causes higher liver iron.

Four variants (rs75935921 in *SLC30A10*, rs13107325 in *SLC39A8*, rs58542926 in *TM6SF2*, rs738409 in *PNPLA3*) were associated with elevated cT1 ($p < 5 \times 10^{-8}$). Insulin resistance, type 2 diabetes, fatty liver, and BMI were causally associated with elevated cT1 whilst favourable adiposity was protective.

In 44,956 individuals over a median of 5.4 years, adjusted models demonstrated strong associations of "suspected liver fibrosis" (FIB-4 ≥ 1.3) with cirrhosis (Hazard ratio (HR 13.64 [10.79 – 17.26], $p < 2 \times 10^{-16}$)) and hepatocellular carcinoma (HR 11.64 [5.15 – 26.31], $p = 3.5 \times 10^{-9}$), but no association with the incidence of most CVDs, albeit a modest increase in AF risk (HR 1.18 [1.01 – 1.37]), when compared to individuals with a FIB-4 < 1.3.

Conclusions

This thesis provides genetic evidence that mechanisms underlying higher liver iron content are likely systemic rather than organ specific. The association between two metal ion transporters and cT1 indicates a new mechanism in steatohepatitis. There is little evidence to suggest that liver fibrosis is an independent risk factor for most CVDs, except possibly a small increase risk

in incident AF risk. This thesis' findings can be used to investigate causality, generate hypotheses for drug development and inform health policies.

Acknowledgements

Many thanks to the employees and volunteers of UK Biobank, our industry collaborator Perspectum Diagnostics, the MHRA, NHS Digital, CPRD ISAC teams, IHI Data Lab and IMI Direct Consortium for their help with data acquisition and collaborative ethos without which this thesis would not be possible.

Many thanks to my supervisors Professor Harry Hemingway, Professor Riyaz Patel and Dr. Michalis Katsoulis for their support and invaluable mentorship. I would like to thank Professor Geraint Rees (Head of the UCL Wellcome Trust Clinical PhD Programme) and the Wellcome Trust for their support and funding.

I would like to acknowledge Dr. Roshni Joshi for many interesting conversations and walks in Regent's Park when times were challenging.

Most of all, I thank my family, friends, and partner Linsey for their patience, love, and unconditional support.

Impact Statement

This thesis has produced both academic and clinical impact. We have performed the first population based genetic studies on liver MRI phenotypes, a non-invasive and safe imaging modality that may be used to identify and monitor individuals with liver disease.

We have identified that elevated liver iron is predominantly driven by systemic mechanisms and is not organ specific. Managing individuals with elevated liver iron may be best through a multispecialty, multidisciplinary approach. We have validated existing and identified novel genetic variants that are associated with increased risk of steatohepatitis; the summary statistics from our research are publicly available via application to UK Biobank. These results can be interrogated and used for a plethora of scientific projects, including meta-analyses, drug development and further Mendelian Randomisation studies. In future, our findings may be used in prediction scores to help identify at risk individuals and target screening and surveillance accordingly. We have further used genetics to attribute causality to associations with liver disease such as central obesity and insulin resistance; these conclusions may in future inform public health policies.

We demonstrate that a validated non-invasive score for liver fibrosis as calculated from routinely collected blood tests in primary care is strongly associated with cirrhosis and hepatocellular carcinoma but not cardiovascular diseases, conclusions that may influence public health policies and screening / primary prevention strategies. We have designed and formulated electronic health record (EHR) phenotypes to extract liver disease outcomes from

primary and secondary care electronic health records. These phenotypes are publicly available for use by the academic and health communities.

Contents

List of Tables

List of Figures

List of Abbreviations and Acronyms

Abdominal aortic aneurysm - **AAA**

Alanine - Aminotransferase - **ALT**

Aspartate - Aminotransferase - **AST**

Atrial fibrillation - **AF**

Coronary heart disease - **CHD**

Cardiovascular disease(s) - **CVD(s)**

Clinical research using Linked Bespoke studies and Electronic Health Records - **CALIBER**

Clinical Practice Research Datalink - **CPRD**

Coronary artery calcification - **CAC**

Endoplasmic reticulum - **ER**

Enhanced Liver Fibrosis - **ELF**

Fibrosis-4 - **FIB-4**

Gamma - Glutamyltransferase - **GGT**

Glycokinase regulatory protein - **GCKR**

Heart failure – **HF**

Hazard Ratio **- HR**

Hepatocellular carcinoma – **HCC**

Human homeostatic iron regulator protein **- HFE**

Hospital Episode Statistics - **HES**

Ischaemic Stroke - **IS**

Liver blood tests - **LBTs**

Mitochondrial Amidoxime Reducing Component 1 - **MARC1**

Membrane bound O acyl-transferases 7 - **MBOAT7**

Myocardial Infarction - **MI**

Non-alcoholic fatty liver - **NAFL**

Non-alcoholic fatty liver disease - **NAFLD**

Nuclear magnetic resonance - **NMR**

NAFLD Fibrosis Score - **NFS**

Negative predictive value - **NPV**

Office for National Statistics - **ONS**

Peripheral arterial disease - **PAD**

Phenome wide association study - **PheWAS**

Patatin-like phospholipase domain-containing protein 3 - **PNPLA3**

Platelets - **PLT**

Positive predictive value - **PPV**

Randomised controlled trials - **RCTs**

Stable angina - **SA**

Solute Carrier Family 39 Member 8 - **SLC39A8**

Transient Ischaemic attack - **TIA**

Transmembrane protease, serine 6 - **TMPRSS6**

Transmembrane 6 superfamily 2 human gene - **TM6SF2**

Unstable Angina - **UA**

VACS - **Veteran Aging Cohort Study**

1 Introduction

1.1 Non-alcoholic fatty liver disease

1.1.1 Introduction

During the past century, dramatic modifications in lifestyle have altered the health priorities in most areas of the World, due to the growing incidence of non-communicable diseases. The new epidemic in chronic liver disease is related to non-alcoholic fatty liver disease (NAFLD). NAFLD is defined as a condition in which liver fat exceeds 5% of hepatocytes in the absence of secondary causes of lipid accumulation (e.g. certain medicines) or clinically significant alcohol consumption.[Citation error] It is a continuum of liver abnormalities, from non-alcoholic fatty liver (NAFL) to the more advanced state, non-alcoholic steatohepatitis (NASH), which in turn can lead to cirrhosis and liver cancer. An important paradox in the history of liver fat accumulation exists; despite the large proportion of adults affected by simple steatosis (fatty liver), only a relatively small proportion (2.4 - 12.8%) will experience significant liver disease or liver related death.[1,2] The global prevalence of NAFLD is currently estimated at 23-28%, with nearly 1 billion people affected.[1] NAFLD is emerging as the most common cause of chronic liver disease in the developed world, and is associated with an increased risk of mortality, with cardiovascular disease (CVD) being the most common cause; there is increasing evidence that NAFLD associates with increased risk for CVDs and mortality.[2,3]

It is important to identify which individuals are at risk of developing the more inflammatory phenotype, steatohepatitis, a condition characterised by lipotoxicity and histological necroinflammation, considered to be the main pathophysiological driver of liver fibrosis and subsequent disease progression.[4] Steatohepatitis and fibrosis affect approximately one in ten middle-aged adults, and can lead to cirrhosis, hepatocellular carcinoma (HCC) and death.[1]

Despite these alarming numbers, reducing disease burden through prevention has not been achieved. Pharmacotherapies have not yet been approved, since several clinical studies have fallen short of the required histological endpoints. The importance of chronic liver diseases and cirrhosis have been largely underestimated. For example, liver disease does not appear in the WHO list of non-communicable diseases that includes amongst others CVDs, cerebrovascular diseases, diabetes, and chronic respiratory diseases.[5] This lack of appropriate consideration is likely to be a contributory factor to the low awareness of chronic liver disease. The heterogeneity of the population with NAFLD with respect to its primary drivers and underlying pathophysiology represent an important challenge to the discovery of highly effective drug treatments; effective treatment requires that individuals should be targeted with precision, based on their individual phenotype and genetic background.

1.1.2 Diagnosis of NAFLD and NASH

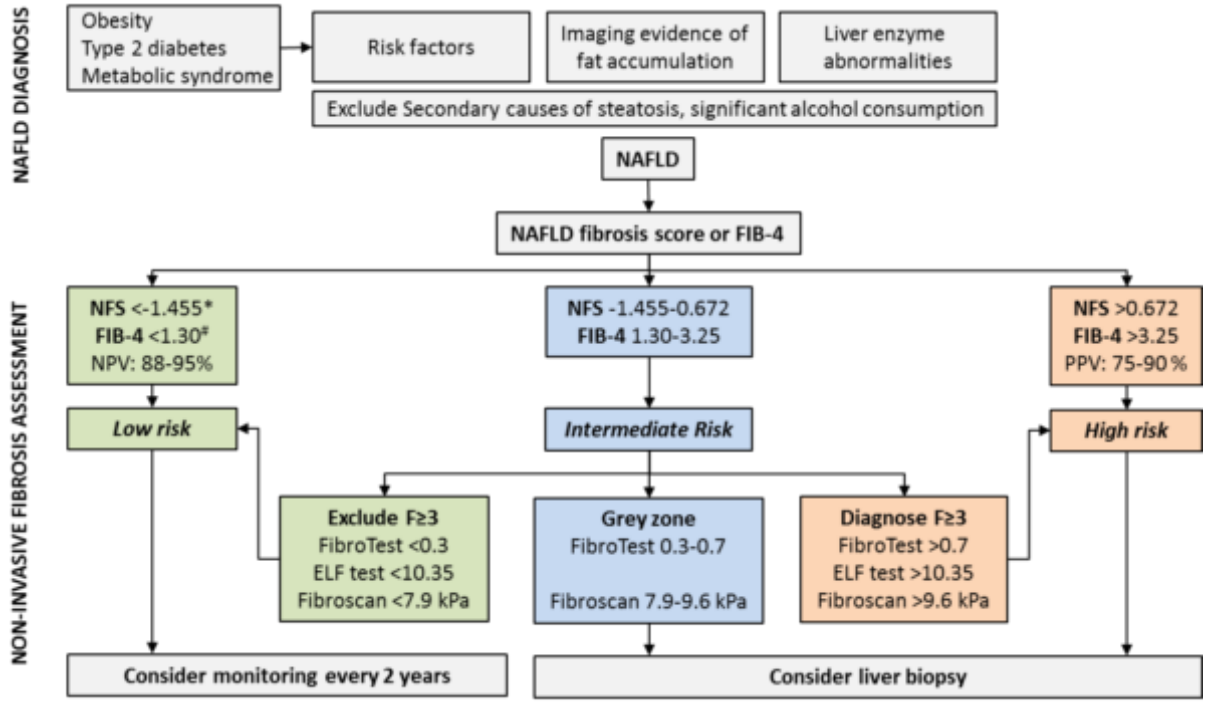
Hepatic steatosis can be routinely identified noninvasively through imaging such as ultrasound, computed tomography (CT) or magnetic resonance imaging (MRI). Ultrasound and CT have similar sensitivity, with the ability to detect steatosis when it comprises approximately 20% of liver mass, whilst MRI can detect as little as 5% steatosis.[6] The vast majority however of patients with NAFLD across the disease spectrum, including those with compensated cirrhosis, are asymptomatic and often have normal liver blood tests (LBTs), resulting in many patients progressing to cirrhosis or HCC undiagnosed.

A diagnosis of NASH is typically considered when aminotransferases are elevated or when abdominal imaging incidentally detects hepatic steatosis. Currently routinely available blood tests do not outperform ALT in identifying patients with NASH.[7] Clinical prediction

algorithms, such as the NAFLD fibrosis score and FIB-4 are based on a combination of routine liver tests, anthropometric data, and glycaemic index. Patients with intermediate risk may undergo additional secondary tests such as elastography (liver stiffness) techniques or the non-invasive enhanced liver fibrosis (ELF™) blood test, which have acceptable performance characteristics to identify advanced fibrosis and good predictive value. A promising, non-invasive measure of steatohepatitis and fibrosis severity is magnetic resonance imaging (MRI) based corrected T1 (cT1), a modality that is used as a proxy for fibroinflammatory disease in this thesis.[8–10] Higher cT1 values are associated with both histological liver inflammation (steatohepatitis, NASH) and fibrosis, although their relative contributions to the score are still unknown.[10,11] cT1 has already been used as a non-invasive outcome in randomised controlled trials for NASH[12] and is associated with liver disease outcomes.[9]

Figure 1.1. Proposed algorithm for diagnosis of NAFLD and non-invasive assessment of liver fibrosis (Figure taken from [13]**). NFS threshold < 0.12 for patients > 65 years. FIB-4 threshold < 2 for patients > 65 years old. Abbreviations: ELF – enhanced liver fibrosis, FIB-4 – Fibrosis 4,**

NAFLD – non-alcoholic fatty liver disease, NFS – NAFLD fibrosis score, NPV – negative predictive value, PPV – positive predictive value.



1.1.2.1 Liver blood tests

LBTs are inexpensive, routinely performed investigations in both primary and secondary care for a broad spectrum of clinical indications (

Table 1.1.1.).

Table 1.1.1. The 10 most commonly recorded reasons for why the LBTs were undertaken by the PCP in a large prospective primary care cohort study in Birmingham; other reasons accounted for 20.9%.[14]

Documented reason	Percentage
Diabetes review	18
Non-specific routine blood tests	15.2
Hypertensive disease review	11.4

Gastrointestinal symptoms (excluding liver specific)	10
Generalised fatigue or tiredness	6.2
Cardiovascular disease review	4.7
Medication review (non-specific)	4.5
Hyperlipidaemia disease review	3.8
Neurological symptoms (inc. confusion)	2.7
Musculoskeletal symptoms (i.e. joint pain)	2.4
Other	20.9

These tests often produce an abnormal result (as defined by the reference ranges of the individual laboratories), the clinical significance of which is frequently unclear. A recent, large (n = 1118) prospective analysis of incidental abnormal LBTs demonstrated that NAFLD (48%) and alcohol excess (46%) accounted for the majority of incidental LBTs abnormalities for which a cause could be found.

1.1.2.2 What LBTs are typically elevated in NAFLD, and what do they tell us?

ALT and **AST** are enzymes present in liver cells (hepatocytes) and are released into the blood stream in response to hepatocyte injury or death (hepatitis). Elevation in either of these enzymes has historically been considered to be the most common abnormality seen on LBT profiles,[7] however a recent large community study showed that **GGT** elevation may be more common.[14] Both enzymes are present in many differing types of tissue, but ALT is considered more liver-specific since it is present in low concentrations in non-hepatic tissue; for this reason, ALT is performed more commonly than AST in a first LBT panel. Although ALT is considered a more specific indicator of liver disease, AST levels may be a more sensitive indicator of liver injury in conditions such as alcohol related liver disease.[15]

GGT is abundant in the liver and present in the kidney, intestine, prostate and pancreas but not in bone; GGT is commonly elevated as a result of obesity, excess alcohol consumption, or maybe induced by certain drugs. It is further elevated alongside ALP in cholestatic liver disease, such as primary sclerosing cholangitis, common bile duct obstruction, and drug induced cholestasis. In a recent community study, GGT was the most common LBT abnormality in the NAFLD cohort, accounting for 75% of abnormal liver blood tests.[14]

1.1.2.3 What constitutes a standard LBT panel?

The recommendation from the most recent national guidelines on the management of abnormal LBTs suggest that the initial panel for potential liver disease should include bilirubin, albumin, alanine aminotransferase (ALT), alkaline phosphatase (ALP) and g-glutamyltransferase (GGT), accompanied with a full blood count if not already performed within the previous 12 months.[7]

In adults, clues of the level and severity of fibrosis can be gleaned from the use of non-invasive scores such as the AST:ALT ratio or FIB-4 index.[16] In this thesis, we will be focussing on a calculated FIB-4 index from routinely collected blood tests in primary care for one of our studies, as a proxy for "suspected fibrosis".

1.1.3 Clinical disease progression, and risk factors for NAFLD and NASH

NAFLD has variable rates of progression among individuals and is likely to reflect the interactions between the environment, metabolism, microbiome, genetic and epigenetic factors. Overall, the majority of individuals have no or mild fibrosis (defined as stages F0, F1, F2), whilst approximately 20% of patients rapidly progress to advanced fibrosis (defined as

stage 3 and stage 4 fibrosis).[17] Methods to identify these rapid progressors, and predict the individuals that may develop significant liver disease and subsequent HCC remain elusive.[18]

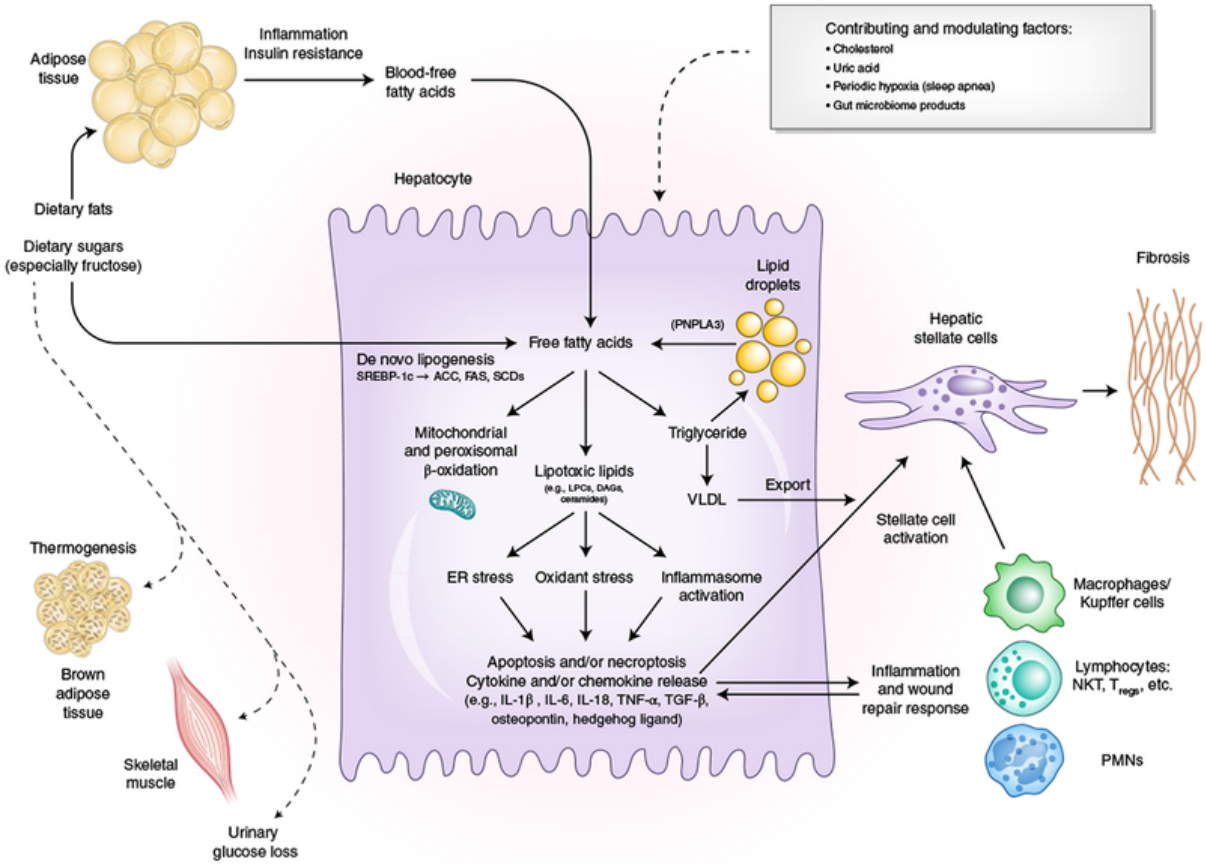
The presence of the metabolic syndrome, a condition which is typically defined by increased waist circumference, hypertension, dyslipidaemia, and insulin resistance, is considered the strongest risk factor for NAFLD and NASH. Up to 75% of individuals with type 2 diabetes have NAFLD, with this population enriched with NASH and advanced fibrosis compared to non-diabetics with NAFLD. Fifty per cent of patients with hypertension have NAFLD, a risk factor that is also associated with higher risk of fibrosis progression.[19,20]

Several genetic risk factors have been identified that associate with risk of NASH. Amongst these include variants in genes such as *PNPLA3*,[21,22] *TM6SF2*,[23,24] *HSD17B13*,[25] *MBOAT7*,[26] *MARC1*,[27] and will be discussed in further detail later in the introduction. Drug development strategies targeting these loci are already underway.[28]

1.1.4 Pathogenesis

Detailed discussion of NAFLD and NASH pathogenesis is beyond the scope of this thesis, however a summary of the main drivers is useful. The liver's capacity to handle energy substrates (carbohydrates and fatty acids) is overwhelmed, leading to toxic lipids.[29–33] These lipids induce hepatocellular stress, inflammation, injury, necrosis, with subsequent fibrosis and genomic instability that predispose to cirrhosis and HCC. Delineating the sources and fate of fatty acids in hepatocytes is essential for understanding NASH pathogenesis. When fatty acids are either a) supplied in excess or b) their disposal from the hepatocyte is impaired, they may trigger lipotoxicity which in turn leads to endoplasmic reticulum (ER) stress and hepatocellular injury (Figure 1.2.).

Figure 1.2. Summary of NASH pathogenesis (Figure taken from [6]). Free fatty acids from the lipolysis of triglycerides in adipose tissue are delivered through blood to the liver, or develop de-novo when hepatocytes breakdown fructose to fatty acids. Fatty acids are disposed from the liver through mitochondrial beta-oxidation or re-esterification to form triglyceride, which is subsequently exported to the blood as VLDL or stored within the hepatocyte in lipid droplets. Lipid droplet triglyceride undergoes lipolysis to release fatty acids back into the hepatocyte free fatty acid pool. PNPLA3 participates in this process, and a SNP in *PNPLA3* is robustly associated with NASH progression. When the disposal of fatty acids through beta-oxidation of triglyceride formation is overwhelmed, fatty acids provide a substrate for the formation of lipotoxic species that lead to ER stress, oxidative stress and inflammasome activation.



1.1.5 Liver iron and dysmetabolic iron overload syndrome (DIOS)

Excess liver iron is the direct cause of liver disease in those with genetic conditions such as hereditary haemochromatosis and thalassaemia,[34,35] and is associated with increased severity and progression of liver diseases including cirrhosis and HCC in individuals with non-

alcoholic fatty liver disease (NAFLD).[36] Observational associations have been described between excess liver iron content, high ferritin, and several metabolic diseases such as high blood pressure, obesity, polycystic ovarian syndrome and type 2 diabetes - a condition recognised as dysmetabolic iron overload syndrome (DIOS) which affects up to 5-10% of the general population.[37,38]. At least half of patients with DIOS have NAFLD, which is not surprising since DIOS shares multiple risk factors with NAFLD, as highlighted above. It further possibly shares other common pathogenic mechanisms, including dysregulation of fatty acid metabolism and insulin resistance. DIOS corresponds to a mild increase in both liver and body iron stores associated with various components of the metabolic syndrome, in the absence of any identifiable cause of iron excess. Previous studies have demonstrated that the presence and hepatic iron is associated with a higher risk of advanced hepatic fibrosis in individuals with NAFLD, when compared to the absence of siderosis.[39,40] Underlying mechanisms once again seem to involve insulin resistance and oxidative stress. Iron is a potent catalyst for oxidative stress via the Fenton reaction, and can directly cause lipid peroxidation which in turn may activate hepatic stellate cells, a major contributor in NAFLD fibrogenesis.[41] Gene expression of adiponectin, which exerts a protective role against insulin resistance, is reduced by iron.[42] MRI can accurately measure liver iron;[43] studying this phenotype in large cohorts with linked genetic and clinical data, such as the UK Biobank,[44] may help identify mechanisms underlying hepatic iron deposition and potentially identify drug targets that may prevent liver disease progression. Current management strategies for individuals with DIOS predominantly rely on similar behavioural changes to those suggested in patients with NAFLD (e.g. exercise, weight loss), with medical treatments such as venesection not currently supported by RCTs.[38,45]

1.2 Magnetic resonance imaging in NAFLD and DIOS

The burden of NAFLD has driven a rapid increase in the number of clinical trials evaluating NAFLD severity and pharmacotherapies. Liver biopsy is the current gold standard measurement for both clinical diagnoses, and as an endpoint in clinical trials; a method that is expensive, invasive, and suffers from high discordance rate among pathologists. This has driven the need to identify alternative, non-invasive endpoints. There is an ongoing unmet clinical need for identifying individuals with the more severe end of the NAFLD spectrum, namely NASH and fibrosis. A promising, non-invasive measure of steatohepatitis and fibrosis severity is magnetic resonance imaging (MRI) based corrected T1 (cT1).[8–10] T1 relaxation time reflects extracellular fluid which is characteristic of fibrosis and inflammation. The presence of iron, which can be determined from T2* maps, has an opposing effect. Combining T2* and T1 values can correct for this opposing effect, from which cT1 (in milliseconds) is derived. Higher cT1 values are associated with both histological liver inflammation and fibrosis, although their relative contributions to the score are still unknown.[10,11] cT1 has already been used as a non-invasive outcome in randomised controlled trials (RCTs) for non-alcoholic steatohepatitis (NASH)[12] and is associated with liver disease outcomes.[9]

Measuring liver iron has traditionally been difficult. As mentioned earlier, liver biopsy is an invasive procedure and therefore unsuitable for population research studies. An alternative is magnetic resonance imaging (MRI); a non-invasive, quick, robust and validated method for quantifying liver iron content.[43] There is excellent inverse correlation between MRI signal and hepatic iron concentration, allowing for the detection of hepatic iron excess with a 84-91% sensitivity and a 80-100% specificity.[43,46,47] It further has a role for identifying patients alcohol related iron overload or DIOS, and those who require further genetic testing

of other haemochromatosis genes (*TFR2, SLC40A1, HAMP, HJV*) following exclusion of C282Y homozygosity in individuals with increased iron stores; its importance in the above diagnostic work up has been highlighted in EASL guidelines for *HFE* haemochromatosis.[48]

The availability of genetic and clinical data, as well as MRI scans of liver in the UK Biobank cohort has provided an unparalleled opportunity to study the genetics of liver iron content and steatohepatitis in the general population. This thesis will explore the genetic determinants of both MRI derived hepatic iron content and cT1 in an unselected population.

1.3 Genetics of NAFLD and liver iron

1.3.1 Genetics of NAFLD

Understanding the underlying genetic susceptibility of steatohepatitis and fibrosis may allow new insights on the main pathophysiological mechanisms contributing to chronic liver disease and help identify potential new drug targets. Genetic studies have so far been limited due to the phenotyping challenge. Liver biopsy is an invasive procedure with associated risks (thus limiting power in GWAS, despite recent excellent efforts to form a centralised histologically characterised cohort),[24] significant sampling error[49] and marked interobserver variance.[50] Routinely available LBTs such as aminotransferases, despite being useful in the identification of important liver disease susceptibility loci, are overall poor predictors of liver disease severity.[51,52]

1.3.2 Heritability of NAFLD

Like other complex traits, the phenotypic manifestations and severity of NAFLD are the outcome of gene - environments interactions. Heritability estimates range from 20 – 70%, with an estimated shared genetic effect between steatosis and fibrosis of 75%.[21,53–57] Similar heritability ranges have been described for other related metabolic traits such as BMI, lipid levels, type 2 diabetes mellitus, blood pressure and other cardiovascular diseases.[58]

1.3.3 Genetic contributions to NAFLD

Since the release of the human reference genome in 2005, genome wide association studies (GWAS) have become the default methodology to determine genotype-phenotype correlations; tests for associations are performed between single nucleotide polymorphisms (SNPs) and a single trait. Because of these large scale, hypotheses free studies, our understanding of the genetic mechanisms that are associated with NAFLD and its progression has increased over the last few years. At least five variants in different genes have been robustly associated with the susceptibility to and progression of NAFLD and are discussed further below.

1.3.3.1 PNPLA3

A common single nucleotide polymorphism (SNP) in the *PNPLA3* gene (rs738409 c.444 C > G p.I148M, inherited in ~ 17% of the population) has been reported, on several occasions, to be associated with susceptibility to steatosis, steatohepatitis, hepatic inflammation, and fibrosis. This association was first reported by Romeo *et al* in 2008.[59] *PNPLA3* rs738409 was significantly associated with increased hepatic triglyceride content and an increase in hepatic inflammation. The observation that the rs738409 variant mediates disease progression has since been repeatedly validated in multiple gene studies.[22,60] The clinical importance of *PNPLA3* has further been highlighted by GWAS both with CT imaging[61] (N = 7,176), but

also histologically characterised cohorts (1,483 European NAFLD cases and 17,781 genetically matched controls) associating *PNPLA3* not only with simple steatosis but also with disease severity.[24] Further to its association with NAFLD, rs738409 has been reported to be a susceptibility variant for HCC.[62] This significant association has been replicated on a background of NAFLD, as well as alcoholic liver disease and viral hepatitis.[63]

PNPLA3 is an enzyme implicated in lipid regulation, and there has been a lot of work focusing on the functional characterisation of rs738409. The C > G mutation causes an isoleucine to methionine change at the amino acid residue level (I148M). In 2010, it was reported in the literature that this change in amino acid sequence hindered the protein's ability to act as a lipase by disturbing the configuration of the active site. As a result, levels of triglycerides start to accumulate in the liver. Human studies on rs738409 have shown that *PNPLA3* has a role in the regulation of very low-density lipoprotein (VLDL) secretion. The SNP reduces the amount of VLDL that is secreted from the liver. As a result of this impaired lipid secretion, lipids accumulate in hepatocytes, leading to the increased steatosis observed in G allele carriers. Consistent with a reduction in circulating VLDL, large genetic studies have suggested that rs738409 was not significantly associated (and may even be nominally protective) with coronary heart disease (CHD).[64]

1.3.3.2 TM6SF2

The search for genetic modifiers of chronic liver disease progression has since been extended, with an exome wide association study identifying *TM6SF2* (rs58542926, T allele, inherited in ~ 7% of the population) as a modifier of hepatic triglyceride content and serum lipoprotein levels as well as a prognostic marker for advanced fibrosis and cirrhosis.[23] This SNP has been replicated in multiple studies, including the recent GWAS by Anstee *et al* in a

histologically characterised NAFLD cohort,[24] and a meta-analysis of all-cause cirrhosis (12,361 all-cause cirrhosis cases and 790,095 controls from eight cohorts).[65] Although the exact function of TM6SF2 is poorly understood, it seems to regulate cholesterol synthesis and the secretion of lipoproteins from hepatocytes.[66,67] Functional studies have indicated that homozygous carriers of rs5854296 have an increased risk of hepatic steatosis and progression to advanced fibrosis but exhibit lower levels of circulating VLDL and triglycerides; this reduction most likely contributes to the decrease in CHD risk seen in large GWAS.[68]

1.3.3.3 Other genes

A 2011 GWAS identified glucokinase regulatory protein (*GCKR*) as a modifier of NAFLD, by controlling de novo lipogenesis and regulating the influx of glucose into hepatocytes,[69] and replicated once again in the recent GWAS by Anstee *et al*.[24] The causal variant seems to be a common loss-of-function mutation (rs1260326), which codes the P446L protein variant and leads to impaired glucokinase activity; as a consequence, there is an increase in glucose uptake in the liver. As a result, levels of malonyl – CoA increase, which promotes hepatic steatosis by acting as a substrate for lipogenesis.[70] Further associations between *GCKR* and NAFLD have been established whereby *GCKR* variants were significantly associated with fibrosis in patients with NAFLD.[70]

More recently, the *MBOAT7* minor (T) allele of rs641738 has been associated with higher hepatic triglyceride content and progression to fibrosis.[26] The *MBOAT7* product is a lysophosphatidylinositol acyltransferase, has a role in inflammatory lipid pathways, such as transferring fatty acids between phospholipids and lysophospholipids and regulating free arachidonic acid in cells. A recent large meta-analysis (to which this thesis has contributed) that included 42 studies and more than one million individuals found that rs641738 > T was

positively associated with liver fat, ALT, fibrosis and HCC, and negatively associated with serum triglycerides.[71]

Variants in *HSD17B13* have recently been reported to be associated with decreased risk cirrhosis and NAFLD, in a large GWAS with exome sequence data and electronic health records from 46,544 participants,[25] as well as a histologically characterised cohort respectively.[24] Although the mechanism is yet unknown, its effect on retinol dehydrogenase and subsequently retinoic acid may be key, in line with recent evidence that all-trans retinoic acids are found at significantly decreased levels in human livers with NAFLD.[72]

A missense variant in Mitochondrial Amidoxime Reducing Component 1 gene (*MARC1* pA165T) was recently shown to protect from all-cause cirrhosis, in a GWAS meta-analysis of 12,361 all-cause cirrhosis cases and 790,095 controls from eight cohorts.[65] It also associated with lower hepatic fat on computed tomography (CT) imaging, lower ALT, lower cholesterol and LDL cholesterol, suggesting that deficiency in MARC1 enzyme may both lower blood cholesterol levels and protect against cirrhosis, making it a very attractive drug target.

1.3.4 Genetics of liver iron

In the most recent GWAS meta-analysis (in 131,471 - 246,139 individuals depending on the biomarker studied) of iron related blood biomarkers (ferritin, serum iron, TIBC, TS), 62 independent variants at 56 loci were identified.[73] However, there is little known about the genetic background of liver iron content. Genetic studies have been limited to gene candidate, family, or small case-control studies; no studies have been performed in unselected populations; research has been limited predominantly to individuals who portray a hereditary haemochromatosis (HH) phenotype. The associations between excess liver iron and hepatic

and non-hepatic diseases necessitate exploration of underlying pathophysiological mechanisms. Furthermore, it is unknown whether iron accumulation is a systemic disorder involving multiple organs or whether there are mechanisms specific to the liver. Previous genome-wide association studies (GWAS) have focussed on peripheral biochemical markers of iron status that do not correlate well with liver iron.[74]

1.3.5 Mendelian Randomisation

Another challenging question is which metabolic traits are a cause (or consequence) of steatohepatitis and iron overload, since treating causal factors can help prevent liver disease. Observational associations between steatohepatitis and other features of the metabolic syndrome might occur because they share common risk factors, rather than one causing the other. Mendelian randomisation (MR) is an established epidemiological approach that uses genetic studies to provide insight on causality.[75] MR uses genetic variants associated with an exposure (e.g. BMI, LDL cholesterol, insulin resistance) to assess their causal effect on an outcome of interest (e.g. cT1, steatohepatitis). Genetic markers of a risk factor are largely independent of confounders that may otherwise cause bias since genetic variants are randomly allocated before birth. Furthermore, the non-modifiable nature of genetic variants provides an analogy to randomised trials, in which exposure is allocated randomly and is non-modifiable by subsequent disease.[76] In this thesis, will be using MR to improve our understanding of cause and effect in the context of liver disease. We will also carry our related phenome-wide studies (PheWAS) to explore associations across all diseases.

1.3.5.1 Principle of Mendelian Randomisation

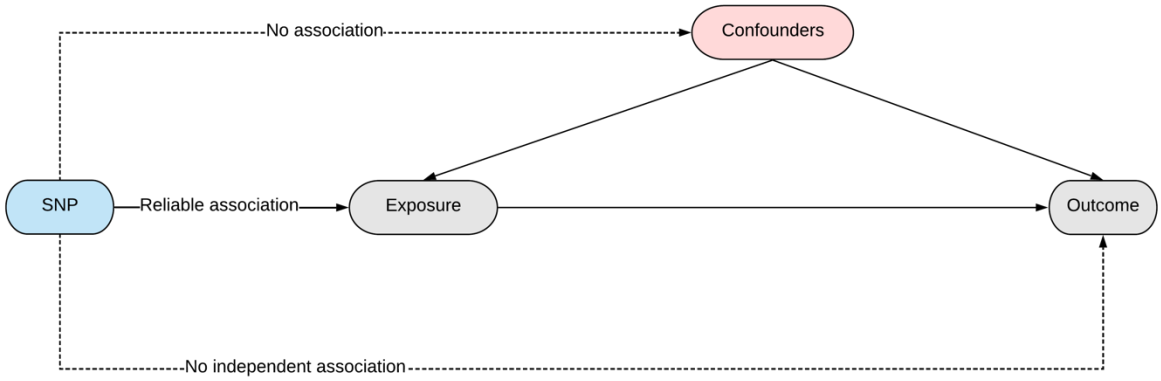
Observational epidemiologic studies measure the association between an exposure and an outcome; however they are vulnerable to confounding, reverse causation, and other forms of

bias; they do not provide evidence that an observed association is causal. MR uses genetic variants associated with an exposure of interest (e.g. BMI, cholesterol, blood pressure) to assess its causal effect on an outcome of interest (e.g. steatohepatitis). In the classic MR paradigm, genetic associations are free from confounding since they are assigned randomly at conception from parents to offspring (according to Mendel's second law) and reverse causation is precluded since the sequence of the germline is not modifiable by disease. MR can be thought of as analogous to a randomised controlled trial (RCT) that uses naturally randomised genetic variation rather than randomised allocation to a drug or treatment, as the 'intervention'.[76]

Genetic polymorphisms that are associated with an exposure of interest are used as an instrument to randomly allocate study participants to higher or lower levels of the exposure under study. Because allocation to genetic variation in levels of the exposure is random, this study design should be less susceptible to confounding. In addition, the allocation of the polymorphism occurs at conception so this study design should not be vulnerable to reverse causation. The results of a Mendelian randomisation study can be interpreted as follows: If a polymorphism (or a collection of polymorphisms, a genetic instrument) is associated with an exposure and the outcome of interest, then the observed association between the exposure and outcome is likely to be causal. If not, then the observed association between the exposure and outcome is likely to be an artefact of confounding, reverse causation, or other study bias.

Figure 1.3. The Mendelian randomization (MR) model: the causal role of an exposure (e.g. BMI) on a disease (e.g. Liver cT1 as a proxy for steatohepatitis) is being examined. A single SNP or collection of genetic variants (a genetic instrument) is formulated and shown to be robustly associated with the exposure (continuous arrow) but not with measured or unmeasured confounders (dotted arrow). The genetic variant is also associated with the disease only through its effects on the exposure and not directly (dotted arrow). The model rests on three assumptions: (i) the genetic instrument is associated with the exposure or biomarker of interest (ii) the genetic instrument must not associate with confounders that are either known or unknown; (iii) the

outcome is associated with the genetic instrument only through the effect of the exposure, and is in all other respects independent.



In this thesis, we used Mendelian randomisation to evaluate the causal association between multiple metabolic traits and diseases (e.g. insulin resistance, obesity, coronary artery disease) previously observationally associated with our outcome of interest (steatohepatitis).

MR also has limitations and results need to be interpreted alongside other evidence in the field in the spirit of triangulation of evidence.[77] Despite presumed random allocation of genetic polymorphisms according to Mendel's law of independent assortment, this study design is still vulnerable to confounding e.g. by population structure or pleiotropy. Confounding by population structure can be addressed by performing studies within ethnically homogeneous study populations (as in this study). Confounding by pleiotropy can be addressed by selecting polymorphisms that are only associated with the exposure of interest, but not with other exposures that are known to be causally associated with the outcome under study, as well as using statistical methods (e.g. MR Egger) as this thesis has done.

To create genetic instruments for possible causes of steatohepatitis, we constructed genetic scores for 24 predominantly metabolic traits. We combined multiple independently inherited polymorphisms to create genetic instruments. These genetic scores are instruments that reflect the combined effect of the polymorphisms included on the exposure of interest. As a result, each score has a much larger effect than any individual polymorphism included in the score. Genetic instruments were constructed by using the effect sizes of independent, genome-wide significant genetic variants ($R^2 < 0.1$) associated with a particular exposure from previous GWASs. Using genetic instruments from a separate population compared to where the outcome of interest was measured results in less bias and more power, a method known as two-sample Mendelian randomisation.[78] We investigated the potential causal associations between 24 predominantly metabolic traits on cT1 using two-sample Mendelian randomisation analysis.

1.3.5.2 An example of Mendelian Randomisation in alcohol and blood pressure

MR is increasingly being used because it can overcome a major limitation of evidence from observational studies: unmeasured confounding. It has previously been used to investigate the effects of alcohol consumption on blood pressure with a view to understanding the overall relationship of alcohol with risk of coronary heart disease.[79] One source of evidence is the association between alcohol and blood pressure in observational studies. This association may be a poor indicator of the causal effects of alcohol if there are other factors, "confounders", that influence both alcohol intake and blood pressure. Many epidemiological methods attempt to correct for, or minimise, observed differences in confounders between study participants. These methods can give useful evidence about causal relations if we measure enough confounders so that, after adjustment or matching, study participants who consume different

amounts of alcohol are otherwise comparable. But this assumption is unverifiable; if it does not hold, then findings from observational studies will be biased estimates of causal effects.

People who consume more alcohol may also have other risk factors for cardiovascular disease, such as smoking more heavily than those with lower alcohol consumption. The confounding factor (smoking) induces a positive association between the risk factor (alcohol) and an outcome (blood pressure); interpreting this as causal would be misleading. Measuring a confounder does not perfectly characterise it, so measurement error leads to residual confounding, even after apparent statistical adjustment. Reverse causality is a form of confounding that is difficult to account for. It arises if the outcome or preclinical aspects of the disease that lead to the outcome affect the risk factor. People with symptoms of cardiovascular disease, for example, may consume less alcohol than those without symptoms. This would lead to a negative association between a risk factor (alcohol) and an outcome (cardiovascular disease); interpreting this as being because alcohol consumption decreases the risk of cardiovascular disease would be misleading.

Mendelian randomisation uses genetic variants, which are fixed at conception, to support causal inferences about the effects of modifiable risk factors, which can overcome some types of confounding. In the case of alcohol and blood pressure, a variant in the *ALDH2* gene (specifically the minor A allele of rs671, rather than the wild type or major allele G) found in east Asian populations slows the metabolism of acetaldehyde, which causes a flush response and other adverse responses to alcohol consumption. In a study of 4057 people selected from the general population, 170 of 1919 men carried two copies of the A allele and drank an average of 1.1 g of alcohol a day, whereas those with no copies drank 23.7 g. If men with one or more copies of the A allele have lower blood pressure, then this implies that lower alcohol

consumption decreases blood pressure. The study found an odds ratio of 2.42 p =4.8×10^{-6}) for hypertension comparing GG with AA homozygotes and an odds ratio of 1.72 (95% CI 1.17–2.52, p =0.006) comparing heterozygotes (surrogate for moderate drinkers) with AA homozygotes. Systolic blood pressure was 7.44 mmHg (p = 1.1×10^{-12}) greater among GG than among AA homozygotes, and 4.24 mmHg (p = 0.00005) greater among heterozygotes than among AA homozygotes.[79] These findings support the hypothesis that alcohol intake has a marked effect on blood pressure and the risk of hypertension.

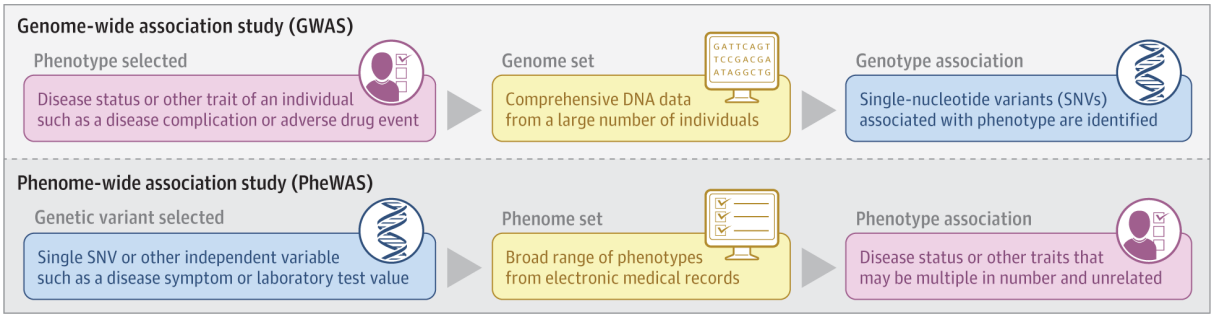
1.3.6 Phenome Wide Association Studies (PheWAS)

Genome-wide association studies (GWAS) have made clear that single-nucleotide polymorphisms (SNPs) that occur at multiple locations across the genome can be associated with a specific condition or trait, also known as a phenotype. Phenome-wide association studies (PheWAS) invert the idea of a GWAS by searching for phenotypes associated with specific SNPs across the range of thousands of human phenotypes, or the "phenome". Analogous to GWAS, PheWAS have shown that specific genetic variations may be associated with multiple conditions and traits.[80]

PheWAS has been developed during the last decade as a research tool. One application is to better understand genetic contributions to human disease, and to begin to identify shared mechanisms across diseases. The approach provides validation for important biologic findings and may have an important role in drug development and drug repurposing. In our GWAS of MRI liver iron, we identified SNPs that are associated with systemic iron metabolism. To gain further insights into possible other, non-liver conditions associated with iron overload, we used established online platforms[81] and publicly available GWASs on thousands of different

phenotypes to investigate how our identified liver iron SNPs are associated with extrahepatic conditions.

Figure 1.4. Genome-Wide Association Study (GWAS) Compared With Phenome-Wide Association Study (PheWAS). A GWAS (top) contrasted with a PheWAS (bottom). A GWAS starts with families or populations in which individuals have been assigned affected or unaffected status for a disease or other trait, such as a complication of a disease or an adverse outcome during drug treatment, and searches for associated genetic variants. A PheWAS starts with a genetic variant and searches across a set of curated human phenotypes (the "phenome") to identify associated phenotypes. The "input function" for the PheWAS can be a single genetic variant or sets of variants or other traits. Figure from [80]



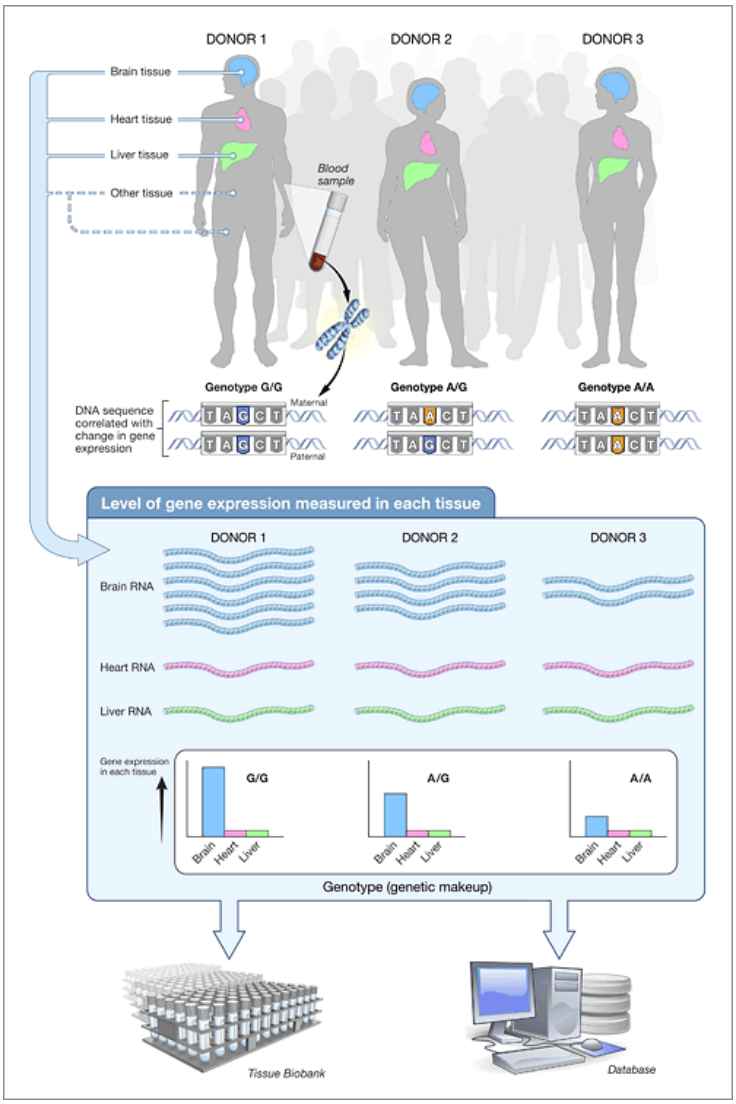
1.3.7 Gene-set and tissue expression enrichment analysis

DNA expresses RNA which in turn encodes for protein. Some human genes and subsequently genetic variants affect the amount of RNA produced, a crucial step in development and maintaining a healthy individual. However, some of these changes only occur in a small number of tissues within the body. The Genotype-Tissue Expression (GTEx) project has been expanded over time and in version 8, Aguet et al. present a deep characterisation of genetic associations and RNA expression in 838 individuals and 53 tissues (Figure 1.5).[82] This large study was able to characterise the details underlying many aspects of gene expression and provides a resource with which to better understand the fundamental molecular mechanisms of how genetic variants affect gene regulation and complex traits in humans. Certain genes and variants preferentially express RNA in certain tissues and analysing such results may allow further mechanistic insights into biological pathways. Tools incorporating GTEx results have

been developed to identify biological processes and tissue specificity that are potentially associated ("enriched") with the relevant genetic variation detected from analysing the input files from a particular GWAS.

The effect of the genetic variation from a GWAS on the RNA expression in 53 distinct tissues and known biological pathways based on the GTEx v6 RNA-seq data is analysed, by grouping certain genetic variants into known "gene sets" that have similar RNA expression. The effect of the genetic variation is mathematically tested and if significant, that tissue (Table 7.1.8) or biological pathway (Table 7.1.9) is "enriched" and may provide a valuable mechanistic insight. An example would be researchers working in muscular dystrophy, where their genetic studies on rare variants may identify genes with enriched expression in muscle and pathways important in glycoprotein biology – those variants should be prioritised as drug targets due to biological plausability.

Figure 1.5. The Genotype-Tissue Expression (GTEx) project aims to provide to the scientific community a resource with which to study human gene expression and regulation and its relationship to genetic variation. This project analyses multiple human tissues from donors who are also densely genotyped, to assess the effect of genetic variation within their genomes on RNA expression. By analyzing global RNA expression within individual tissues and treating the expression levels of genes as quantitative traits, variations in gene expression that are highly correlated with genetic variation can be identified as expression quantitative trait loci, or eQTLs (Figure from [84]**). From our GWAS, we will compare our results to the GTEX reference panel and investigate whether our prioritised genetic variants affect RNA expression within certain tissues types, molecular pathways and diseases, to gain further insights into biology.**



1.4 Population studies in NAFLD and cardiovascular disease

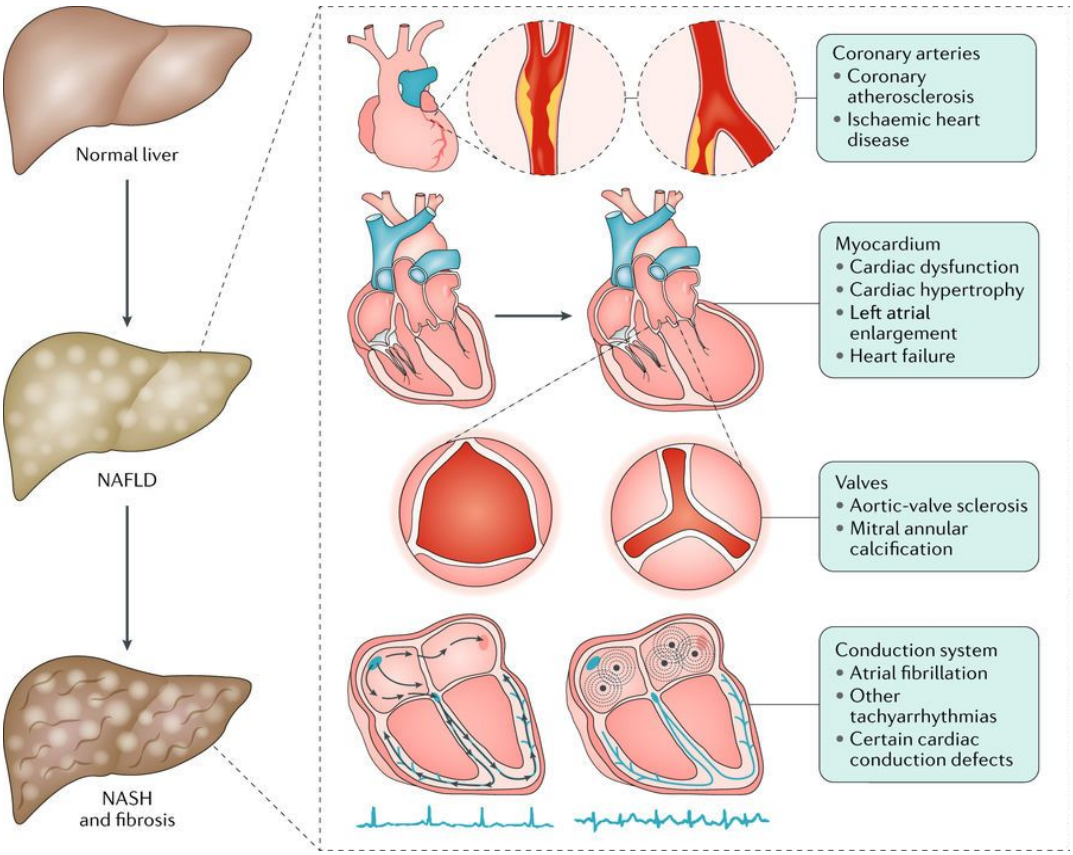
1.4.1 NAFLD and CVDs

The liver is a key organ for the regulation of metabolic homeostasis, mediating the interaction of the external environment including dietary intake and gut derived microbial signals. NAFLD also demonstrates close associations with cardiometabolic risk factors which make up the metabolic syndrome, including abdominal obesity, dyslipidaemia, central obesity, insulin

resistance high blood pressure.[85] Hence, it is not surprising that NAFLD is associated and may be a determinant of extrahepatic diseases including those of the cardiovascular system, diabetes risk and cancer.[3,4,85–89] NAFLD may play a causative role in the genesis of CVDs (Figure 1.6.), the primary cause of mortality in this cohort; accumulation of hepatic fat may result in an increase of hepatic insulin resistance, as well as increasing glucose levels, leading to an atherogenic lipid profile. Hepatic lipid accumulation is directly proportional to the severity of each component of the metabolic syndrome.[4] However this causal hypothesis has not been robustly demonstrated; In epidemiological studies, issues with confounding and reverse causation prevent inferences to causation; identifying cause-effect associations between NAFLD and CVDs is key, as intervening to treat a potential causal exposure might help prevent the disease of interest, whereas should the association be found to be non-causal, important health resources may be directed elsewhere.

NAFLD has been inadequately studied in a population setting in relation to distinct CVD endpoints (CVDs), with studies primarily focusing on CVD mortality, myocardial infarction (MI) and stroke. Studies focusing on other CVDs, such as atrial fibrillation (AF), heart failure (HF) and peripheral arterial disease (PAD) are limited, despite such conditions being associated with significant morbidity. Detection, treatment, and primary and secondary prevention strategies for these conditions differ from those for MI. Furthermore, the validation of NAFLD as a significant prognostic risk factor for CVD may have direct relevance for screening and surveillance strategies; addition of existing diseases such as rheumatoid arthritis and diabetes to improve cardiovascular risk prediction models for MI and stroke has already occurred in risk scores.[90] Addressing CVD risk in patients with NAFLD is also the aspect of the disease most amenable to medical management, improving long-term outcomes.

Figure 1.6. Possible adverse effects of NAFLD on coronary arteries and other anatomical structures of the heart (Figure from [3]**).**



1.4.2 NAFLD and increase of incident CVDs

Several studies support an association between NAFLD and incident CVD. An updated meta-analysis incorporated a total of 16 observational studies with 34043 adults and captured nearly 2600 CVD outcomes over a median 6.9-year follow up. This meta-analysis concluded that the presence of NAFLD (diagnosed by imaging or histology) conferred an OR of 1.64 [1.26 - 2.13] for fatal and non-fatal incident CVD events, a risk that appeared to increase further with greater severity of NAFLD (defined either by presence of fatty liver on imaging plus elevated GGT, high NFS, high FDG uptake on PET or increasing fibrosis stage on liver biopsy, OR 1.94 [1.78 - 3.75]).[91] Limitations do exist such as failure to adjust for confounding, NAFLD defined

differently between studies, selective reporting of studies and different CVDs reported, all of which are later pulled into one outcome. Other prospective studies in histologically proven NAFLD suggested that the risk of incident CVDs paralleled the underlying severity and fibrosis stage.[92,93] Others recently demonstrated an independent link between NAFLD and incidence of MI, even in primary care populations.[94,95]

Recent cohort studies, predominantly in Asia, reported significant associations between ultrasound diagnosed NAFLD and the progression of subclinical coronary or carotid atherosclerosis independent of multiple CVD risk factors.[96,97] The risk of subclinical carotid/ coronary atherosclerosis progression was also higher among patients with NAFLD with increased non-invasive markers of advanced fibrosis at baseline (NFS, FIB-4 score or elevated GGT levels). Additionally, the regression of NAFLD on ultrasound over time was associated with a decreased risk of subclinical atherosclerosis development.

Overall, the available observational evidence seems to demonstrate an association between NAFLD and CVD and supports the view that NAFLD may be independently associated with higher CVD risk; this relationship may be modified by a range of factors including fibrosis severity and genetic variation. However, recent large population studies have cast doubt on whether this is really the case. In a cohort of 285 US adults with biopsy proven NAFLD without pre-existing CVD, Henson *et al* found that advanced fibrosis, but no other histological features of NAFLD, were associated with increased CVD incidence over a median of 5.2 years.[98] A recent large, international population based case control study (120,795 individuals with NAFLD, each case matched to up to 100 controls) failed to find any significant associations between a recorded diagnosis of NAFLD and risk of developing MI and stroke, after adjustment for traditional risk factors, using electronic health records from four large European

primary healthcare databases; however, the lack of association may be due to misclassification bias of NAFLD cases.[99] In a recent, prospective study with a median follow up of 4 years of outcomes in 1773 adults with biopsy proven NAFLD, cardiac events were similar across all fibrosis stages.[100]

It remains unclear whether NAFLD and NAFLD severity is causal of CVDs or useful as an independent, prognostic factor in CVD risk scores. Despite multiple cohort studies investigating associations between NAFLD and CVD, most focus on CVD mortality, CHD, MI and stroke; very few studies have investigated and directly compared the relationship between NAFLD with incidence of other specific CVDs (e.g. HF, AF, PAD).

This thesis will aim to address this question, by calculating a FIB-4 index using routinely collected blood tests and demographic data in primary care, adopting strict exclusion criteria, adjusting for confounders, and linking with incident CVD outcomes in primary, secondary and mortality data in England. A continuous variable such as FIB-4 in an estimated 50,000 individuals with no CVD history will give us unparalleled power to investigate risk of independent CVDs, the ability to examine shapes of association, whilst also providing a proxy of fibrosis severity.

1.4.3 LBTs, FIB-4 and CVD risk

Liver enzymes such as GGT, ALT, AST, have been used as markers of hepatic dysfunction and non-alcoholic fatty liver disease (NAFLD) and have attracted attention as emerging risk factors for CVDs. Fraser *et al* reported a meta-analysis exploring the association of GGT but not ALT with incident coronary heart disease (CHD), stroke and combined outcomes.[101] Other meta-analyses have been negative, or found ALT being inversely associated with CHD

and positively associated with stroke.[102] These variations in findings of previous studies suggest differential associations of specific liver enzymes with CVD and mortality. Furthermore, these discordant results may reflect small sample size, insufficient follow up period, minimal corrections of confounders such as common medications (e.g. statins), differences in age, gender and ethnicity, and the grouping of all CVDs in one outcome; In a key review investigating the relationships between liver enzymes, NAFLD, and incident CVD, Ghouri *et al.* concluded that biochemical and imaging markers of NAFLD are insufficient to identify patients at high risk of CVD.[103] MR studies failed to show a causal association between elevated liver enzymes and risk of ischaemic heart disease.[104]

In this thesis, we will concentrate on the FIB-4 score, as a marker of suspected fibrosis and its association with risk of incident CVDs. Studies have so far been limited to relatively small numbers, single CVDs or composite endpoints, predominantly male participants (e.g. Veteran Aging Study), or within disease cohorts (e.g. in patients with existing AF) (Table 1.4.1.).

Table 1.4.1. Studies exploring FIB-4 score and incidence of CVDs. VACS = Veteran Aging Cohort Study, CAC = coronary artery calcification, IS = ischaemic stroke, MI = myocardial infarction, AF = atrial fibrillation, PAD = peripheral arterial disease).

Reference	FIB-4 cut-offs	Population	Outcome	Conclusion
Kaku A, *et al* 2017[105]	1.45, 3.25	VACS, N = 96,373	Incident HF	HR 1.17 [0.96-1.27], 1.65 [1.56-1.92]
Saito Y, *et al* 2020[106]	Tertiles	Patients with AF, N = 3067	CVD	HR 1.72 [1.31-2.25]
Lee J, *et al* 2020[107]	1.3	Asymptomatic adults with CAC score, N = 1173	CAC progression	HR 1.7 [1.12 - 2.58]
Akuta, N, *et al* 2021[108]	2.67	Histologically proven NAFLD, N = 477	Incident CVDs	HR 2.73 [1.21-6.14]

This thesis	1.3, FIB-4 as continuous variable, quintiles	Unselected, England, no history of CVDs, N = 49,946	IS, MI, AF, HF, PAD, Composite	Chapter 5

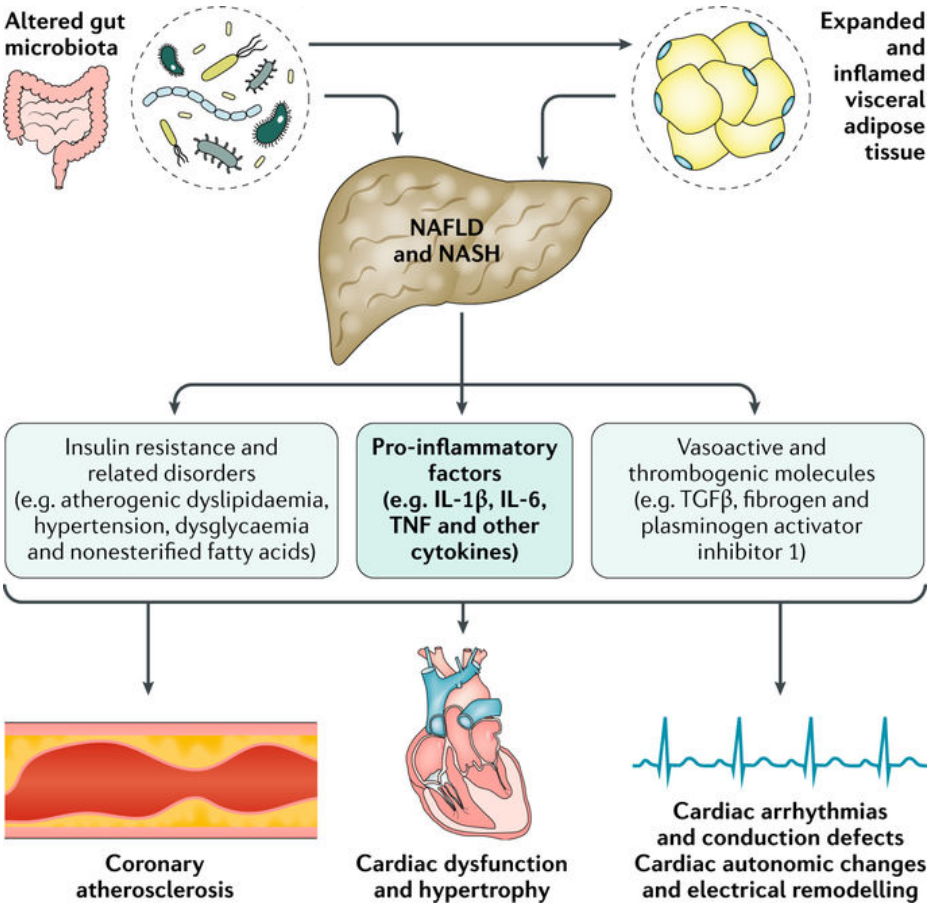
1.4.4 Putative mechanisms linking NAFLD to CVDs

The pathophysiology behind the association of NAFLD with CVDs is poorly understood; these conditions share similar risk factors, such as high blood pressure, insulin resistance, central obesity, and dyslipidaemia, but questions remain as to whether there is an independent link. A detailed review of all putative mechanisms is beyond the scope of this thesis, but we highlight the main concepts below.

Low-grade systemic inflammation, leading to higher levels of cytokines such as interleukin-6 (IL-6) and tumour-necrosis factor (TNF) characterises metabolic disorders such as NAFLD and diabetes and could crucially contribute to the development of NAFLD-related extrahepatic complications and CVDs.[3] The underlying mechanisms responsible for this systemic inflammatory reaction include lipotoxicity, oxidative stress, ER stress and alterations of the gut microbiota (Figure 1.7.).

Figure 1.7. In NAFLD, systemic inflammation may be generated by interactions between diet, microbiome, genetics, adipose tissue and the liver. Visceral adipose tissue and the liver are major cytokine producers in NAFLD. Low grade systemic inflammation plays a crucial part in the pathophysiology of cardiomyopathy, cardiac arrhythmias and atherosclerosis. Pro-inflammatory cytokines can induce myocardial remodelling, dysfunction of calcium homeostasis and abnormalities of specific connexin-formed channels that are associated with changes in

myocardial fibre continuity and development of arrhythmias. TGF-b, transforming growth factor beta (Figure taken from [89]**).**



The presence and generation of leukocytes and pro-inflammatory cytokines as observed in NAFLD and NASH might affect distal tissues, such as vasculature and the heart, and contribute to associated pathologies, such as atherogenesis, HF, and cardiac arrhythmias.[109] To date, it is not known whether proinflammatory pathways in ectopic fat directly affect cardiac function and atherosclerosis development. Liver fat has also been linked to plasma inflammatory markers in the Framingham Study.[110] There is emerging genetic and RCT data that inflammation contributes to CVDs, independent of lipid levels.[111] Large genetic studies demonstrate that genetically impaired interleukin-6 receptor (IL6R) signaling (a key

inflammatory pathway) reduced the risk of CHD, AAA, and AF.[112–114] The strongest evidence yet however comes through a recent RCT, demonstrating that anti-inflammatory therapy targeting the interleukin-1b innate immune pathway with monoclonal antibody canakinumab led to a significant lower rate of recurrent cardiovascular events than placebo, independent of lipid level lowering (CANTOS trial).[115]

There is increasing evidence that gut microbiota controls metabolic functions, with early animal studies demonstrating a crucial role in the development of adipose tissue and NAFLD pathogenesis.[116] A microbiome signature for cirrhosis due to NAFLD has already been described,[117] with significant changes in the gut microbiome observed in NAFLD-associated coronary heart disease.[118] It is still unknown however which gut microbiota need to be modified, both in type and quantity, in order to potentially benefit the liver and/ or CVD risk in NAFLD.

Although it is uncertain whether NAFLD causes certain CVDs, the fact that NAFLD reflects a low-grade inflammatory state and is strongly associated with cardiac pathologies make such a connection likely. Therefore, lowering the chronic inflammatory burden in NAFLD might represent an effective intervention to reduce CVD risk.

1.5 Introduction summary

In metabolic liver disease, both iron overload and steatohepatitis are common, and are associated with increased risk of chronic liver disease, HCC, and many other metabolic and non-metabolic traits and diseases. Due to the lack of large-scale genetic data linked to accurate non-invasive imaging and clinical datasets, investigating underlying genetic determinants has been limited, despite valuable insights from large GWAS studies on LBTs. NAFLD has also

been linked with CVDs, however literature on whether this link is causal or independent is still conflicted.

In this thesis we will carry out GWAS of liver MRI phenotypes iron and cT1, to gain insight in aetiopathogenesis and susceptibility. We will use Mendelian randomisation to gain insight in the directional relationships of these phenotypes with metabolic and non-metabolic diseases. Finally, we will use population wide EHRs from primary, secondary care and mortality data to investigate whether there is a link between suspected liver fibrosis (calculated by FIB-4) and 5 CVDs.

1.6 Key Introduction Highlights

- Liver disease is common and increasing worldwide, mirroring the rise of the obesity epidemic.
- Population studies have so far been limited due to the risks of diagnostic modalities such as liver biopsy and CT.
- MRI is a safe, non-invasive method for quantifying measures associated with chronic liver disease such as liver iron, fat, and cT1, and is now available at scale due to large populations studies such as UK Biobank. MRI imaging linked with genetic data allow us to investigate the genetic background of such phenotypes, possibly identifying new mechanistic insights into the pathophysiology of liver disease.
- Liver disease is linked with multiple metabolic and extrahepatic traits and diseases; however, it is unknown whether these associations are causal. Novel methodologies

such as Mendelian randomisation can provide insights into causality, paving the way for future interventions.

- NAFLD and suspected secondary liver fibrosis have been linked with increased cardiovascular disease risk, however literature is still divided, and multiple CVD outcomes have rarely been investigated individually in the same cohort. Linked EHRs alongside non-invasive scores for liver fibrosis such as FIB-4 may allow us to examine whether it is an independent risk factor on a population scale.

2 Datasets used

2.1 UK Biobank

UK Biobank (UKB) is a national and international health resource and cohort study open to all bona-fide health researchers.[44] It follows the health and well-being of 500,000 volunteer participants aged 40-69 when recruited in 2006-2010 and provides health information data to approved researchers in the UK and overseas. Health information includes baseline assessment, linkage to a wide range of EHRs including primary care, secondary care, cancer, and death registries. Genotyping has been undertaken on all 500,000 participants. A further major study to scan (whole body MRI) 100,000 participants is also underway, as well as a one-off blood biochemistry level for all participants. The UKB dataset provides an invaluable resource for agnostic research, including genetic and phenome-wide studies. The availability of MRI liver scans would be a novel source for detailed studies in the cause and effect of liver diseases. In this thesis, we used UK Biobank data to perform GWAS studies on MRI liver iron and cT1 measures. Details for the application and methods for these studies can be found at sub-headings 3.6 and 4.6 respectively.

2.2 Publicly available genetic datasets

The availability of summary statistics from large genome wide association studies (GWAS) allow us to investigate the effect of SNPs associated with NAFLD with many traits and diseases. We used curated genetic datasets available publicly on well-maintained academic online platforms (Phenoscanner, MR-Base, MR-PheWAS),[81,119] and further downloaded and extract summary statistics from individual genetic studies of diseases of interest. These datasets allow us to characterise the relationship and shared pathways between NAFLD and

certain metabolites (e.g. VLDL cholesterol), traits (e.g. blood pressure, body mass index), and diseases (e.g. CHD, AF, IS) of interest.

2.3 CALIBER

CALIBER is a unique research platform consisting of research ready variables extracted from linked EHR from primary care (CPRD) including demographics, blood tests, prescription data, disease codes; all coded inpatient and outpatient hospital records (HES), social deprivation information and cause specific mortality data (ONS) in England.[120] The resource consists of data up to 2016 encompassing more than 10 million adults with 400,000,000 person year follow up. In this thesis, we used the CALIBER platform to investigate the associations between calculated FIB-4 and incidence of cardiovascular diseases in approximately 50,000 individuals. Details on the application, phenotype derivation and methods for this study can be found at sub-heading 5.7.

2.3.1 CALIBER data resource

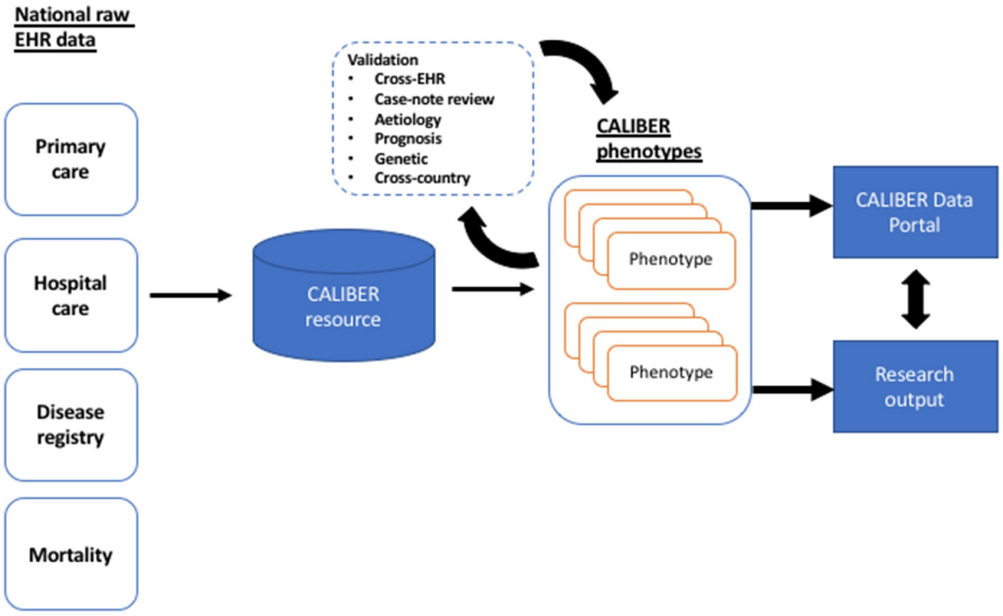
The CALIBER platform[120] is currently built around 4 national EHR data sources (Figure 2.1) linked using NHS number (unique 10-digit identifier assigned at birth or first interaction), gender, postcode, and date of birth; 96% of patients with a valid NHS number successfully linked.[121]

The baseline cohort is composed of a national primary care EHR database, the Clinical Practice Research Datalink (CPRD).[120] Primary care has used computerised health records since 2000 and general practices use one of several EHR systems. CPRD contains longitudinal

primary care data (extracted from the Vision and Egton Medical Information Systems clinical information systems) on diagnoses, symptoms, drug prescriptions, vaccinations, blood tests, and risk factors irrespective of disease status and hospitalisation. The CPRD uses Read[122] terms (112 806 terms; subset of the International Health Terminology Standards Development Organization SNOMED-CT [Systematized Nomenclature of Medicine Clinical Terms])[122] to record information. Prescriptions are recorded using Gemscript (a commercial derivative of the NHS Dictionary of Medicines and Devices (72 664 entries). The CPRD contains >10 billion rows of data from >15 million patients (from all the contributing primary care practices, irrespective of consent to linkage) shown to be representative in terms of age, sex, mortality, and ethnicity and of high validity.[123–126]

Hospital Episode Statistics (HES) (https://digital.nhs.uk/) contains administrative data on diagnoses and procedures generated during hospital interactions.[127] Diagnoses are recorded using the ICD-10 and procedures using the Office of Population Censuses and Surveys Classification of Surgical Operations and Procedures, Fourth Revision (10 713 terms, like Current Procedural Terminology). Up to 20 primary and secondary discharge diagnoses are recorded per finished consultant episode. The Myocardial Ischaemia National Audit Project (MINAP) is a national disease and quality improvement registry capturing all acute coronary syndrome events across England. MINAP contains diagnostic, severity and treatment information using 120 structured data fields, however this resource was not used in this thesis.[128] The Office for National Statistics (ONS) contains socioeconomic deprivation using the Index of Multiple Deprivation and physician-certified cause-specific mortality (underlying and up to 14 secondary causes using International Classification of Diseases–Ninth Revision [ICD-9] or ICD-10).[129]

Figure 2.1. The CALIBER platform (https://www.caliberresearch.org) links national structured electronic health records (EHRs) across primary care, secondary care, and mortality for research (Figure from [130]**).**



2.3.2 Ethical approval for CALIBER and the project in this thesis

The CPRD has broad ethical approval for purely observational research using pseudonymized linked primary or secondary care data for supporting medical purposes that are in the interests of patients and the wider public. Linkages were performed by NHS Digital, the statutory body in England responsible for providing core healthcare information technology and curating many of the national datasets. Our specific study was approved by the Medicines and Healthcare Products Regulatory Agency Independent Scientific Advisory Committee in 2017 (protocol reference 17062 RAR). CPRD is jointly sponsored by the Medicines and Healthcare products Regulatory Agency (MHRA) and the National Institute for Healthcare Research (NIHR), as part of the Department of Health and Social Care.

3 Genetic studies of MRI liver iron content identify susceptibility loci and yield insights into its link with other diseases.

3.1 Abstract

3.1.1 Background & Aims

Excess liver iron content is relatively common, however its genetic background and link to hepatic and extrahepatic disease risk is unknown. We aimed to identify genetic variants that confer susceptibility to higher liver iron content and use genetics as a tool to understand how it is associated to other traits and diseases.

3.1.2 Methods

First, we performed a GWAS in 8,289 individuals in the UK Biobank study with MRI quantified liver iron and validated our findings in an independent cohort (N = 1,513 from IMI DIRECT). Second, we used Mendelian randomisation to test the causal effects of 25 predominantly metabolic traits on liver iron content. Third, we tested phenome-wide associations for genetic variants associated with liver iron content using 770 anthropometric traits and diseases in UK Biobank and publicly available GWAS.

3.1.3 Results

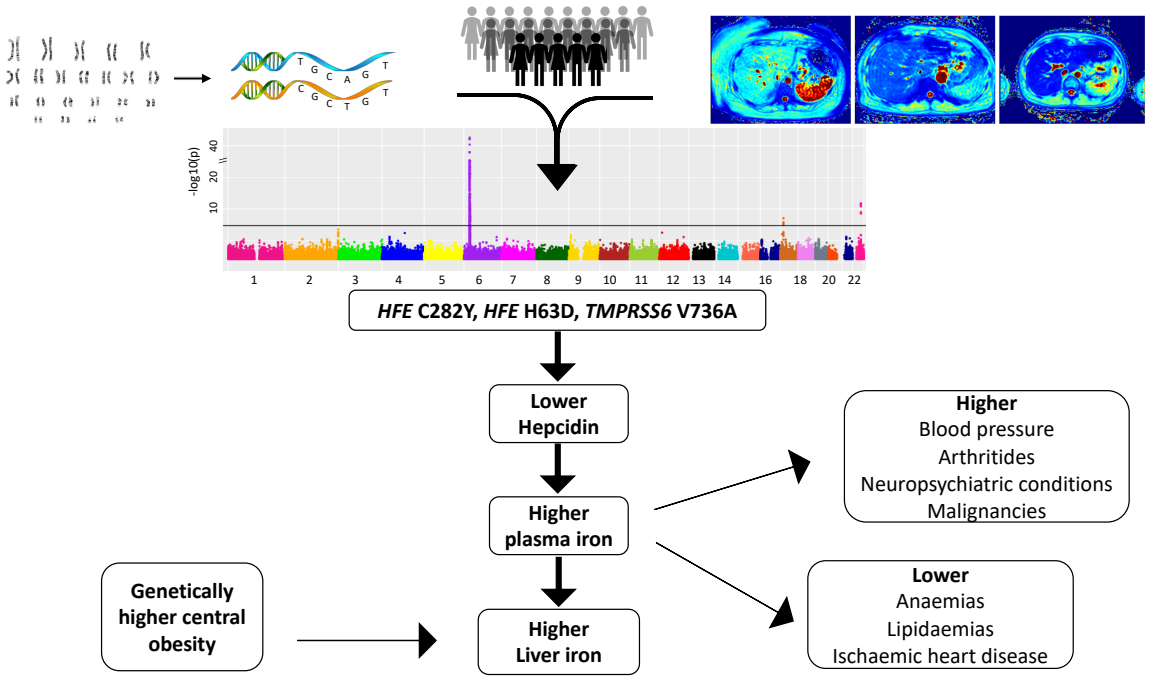
We identified three independent genetic variants (rs1800562 (C282Y) and rs1799945 (H63D) in *HFE,* rs855791 (V736A) in *TMPRSS6*) associated with liver iron content that reached the GWAS significance threshold ($p < 5 \times 10^{-8}$). The two *HFE* variants account for ~85% of all cases of hereditary haemochromatosis, whilst *TMPRSS6* is implicated in hepcidin regulation and affects circulating iron levels. Mendelian randomisation analysis provided evidence that higher central obesity plays a causal role in increased liver iron content. Phenome-wide association analysis demonstrated shared pathogenic mechanisms for elevated liver iron, high blood pressure, cirrhosis, malignancies, neuropsychiatric and rheumatological conditions,

while also highlighting an inverse association with anaemias, lipidaemias and ischaemic heart disease.

3.1.4 Conclusion

Our study provides genetic evidence that mechanisms underlying higher liver iron content are likely systemic rather than organ specific, that central obesity is a likely cause of higher liver iron, and that liver iron shares common aetiology with multiple metabolic and non-metabolic diseases. A multi-specialty, multidisciplinary approach to the management of individuals with higher liver iron content may be a more appropriate strategy compared to focussing attention purely in the liver clinic.

3.2 Visual Abstract



3.3 Lay summary

Excess liver iron content is common and linked to diseases such as diabetes, high blood pressure, heart and liver disease, however little is known about underlying genetic risk. We find that three genetic variants are linked to increased risk of developing excess liver iron content. We show that the same genetic variants are linked to higher risk of many diseases, but they may also be associated with some health advantages. Finally, we use genetic variants as a tool to show that central obesity is causally associated with increased liver iron content, pointing to weight loss as a potential intervention in certain individuals.

3.4 Highlights

- Variants in *HFE* and *TMPRSS6* are associated with higher liver iron.
- There is genetic evidence that higher central obesity causes higher liver iron.
- Liver iron variants are not organ specific and associate with multiple diseases.

3.5 Introduction

Liver disease constitutes the third most common cause of premature death in the UK, and its prevalence is substantially higher compared to other countries in Western Europe.[1,131,132] Excess liver iron is associated with increased severity and progression of liver diseases including cirrhosis and HCC in individuals with NAFLD, [40,133,134] and is the direct cause of liver disease in those with hereditary haemochromatosis and thalassaemia.[34,35] Observational associations have been described between excess liver iron content and several metabolic diseases such as high blood pressure, obesity, polycystic ovarian syndrome and type 2 diabetes - a condition recognised as dysmetabolic iron overload syndrome (DIOS) which affects up to 5-10% of the general population.[37,38]

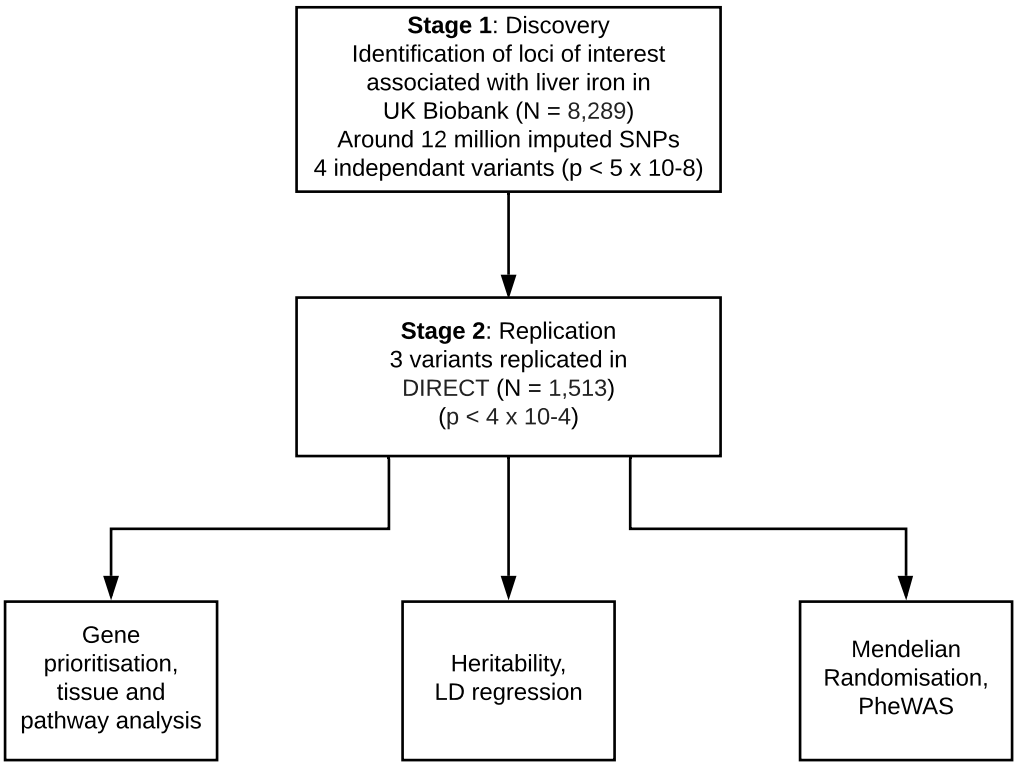
The associations between excess liver iron and hepatic and non-hepatic diseases necessitate exploration of underlying pathophysiological mechanisms. There is little known about the genetic background of liver iron content. Studies of hereditary haemochromatosis patients with autosomal recessive mutations show they have higher liver iron, measured from biopsies, when compared to controls, however no studies have been performed in unselected populations. Furthermore, it is unknown whether iron accumulation is a systemic disorder involving multiple organs or whether there are mechanisms specific to the liver. Previous GWAS have focussed on peripheral biochemical markers of iron status that do not correlate well with liver iron.[74]

Measuring liver iron has traditionally been difficult. Liver biopsy, the “gold standard” for assessment of liver iron, is an invasive procedure and therefore unsuitable for population research studies. An alternative is magnetic resonance imaging (MRI); a non-invasive, quick,

robust and validated method for quantifying liver iron content.[43] The availability of genetic and clinical data, as well as MRI scans of liver in the UK Biobank cohort has provided an unparalleled opportunity to study the genetics of liver iron content in the general population.

The aim of this study was to (i) identify genetic variants specifically associated with liver iron content, (ii) investigate which metabolic traits and diseases might cause higher liver iron content, and (iii) characterise the traits/diseases associated with liver iron content susceptibility variants. To facilitate this, we performed the first GWAS of MRI-determined liver iron content in 8,289 UK Biobank participants and replicated our findings in an independent cohort of 1,513 participants of European ancestry from the IMI DIRECT study (Figure 3.1.).[44,135]

Figure 3.1. Study design. GWAS on liver iron content was performed in UK Biobank (N = 8,289) and replicated in IMI DIRECT (N = 1,513).



3.6 Methods

3.6.1 UK Biobank participants

UK Biobank consists of over 500,000 individuals aged 37–73 years (99.5% were between 40 and 69 years of age) who were recruited between 2006 and 2010 from across the U.K.[44] This research has been conducted using the data obtained via UK Biobank Access Application number 9914. UK Biobank field numbers used for this analysis can be found in Table 7.1.1. We used data from the first subset of UK Biobank participants invited for multiparametric MRI imaging between 2014 and 2016.[136] After image analysis and quality control steps (see below), liver iron was available for 8,674 individuals who also had genetic data. We based our study on 8,289 individuals of white European descent as defined by principal component (PC) analysis. Briefly, we first generated PCs in the 1000 Genomes cohort using high-confidence SNPs to obtain their individual loadings. We then used these loadings to project all the UK Biobank samples into the same PC space, and individuals were clustered using PCs 1–4.

3.6.2 Genetic Data

Protocols for the participant genotyping, data collection, and quality control have previously been described in detail.[44] Briefly, participants were genotyped using one of two purpose-designed arrays (UK BiLEVE Axiom Array (N=50,520) and UK Biobank Axiom Array (N=438,692)) with 95% marker overlap. We excluded individuals who were identified by the UK Biobank as outliers based on either genotyping missingness rate or heterogeneity, or whose sex inferred from the genotypes did not match their self-reported sex. We removed individuals with a missingness >5% across variants which passed our quality control procedure. We used the latest release which included imputed data using two reference panels: a combined UK10K and 1000 Genomes panel and the Haplotype Reference Consortium (HRC) panel. We limited

our analysis to genetic variants with a minimum minor allele frequency (MAF) > 1% and imputation quality score (INFO) > 0.3.

3.6.3 Imaging protocol and analysis

The imaging protocol and analysis of liver iron content in UK Biobank participants has previously been published, and the measures were provided by our collaborator Dr. Henry Wilman and Perspectum Diagnostics.[136] Briefly, participants were scanned at the UK Biobank centre in Cheadle (UK) using a Siemens 1.5T Magnetom Aera. A single-breath-hold MRI sequence was acquired as a single transverse slice captured through the centre of the liver, superior to the porta hepatis. This sequence forms part of the UK Biobank abdominal imaging protocol. The data was analysed using the Liver*MultiScan*™ software by a team of trained analysts, blinded to any subject variables. Analysts selected three 15mm diameter circular regions of interests, to cover a representative sample of the liver parenchyma, avoiding vessels, bile ducts and other organs. The repeatability and reproducibility of the image analysis was high.[136]

3.6.4 Genome-wide association analysis

We performed the association tests using 2 different software: (1) GEMMA version 0.96 as our main analysis using all individuals of genetically defined Europeans (N=8,289),[137] and (2) PLINK version 1.9 as our sensitivity analysis using unrelated white British individuals (defined in UK Biobank field 22006, N = 6,758).[138] For this analysis, I used automated pipelines optimised for GWAS analysis by the senior author of the published paper, Dr Hanieh Yaghootkar and the Frayling lab, University of Exeter.

GEMMA applies a linear mixed-model (LMM) to adjust for the effects of population structure and relatedness. Therefore, we increased our power by including all related individuals of European descent. The relatedness matrix was computed using common (MAF>5%) genotyped variants that passed quality control. Prior to association testing, liver iron was first log-transformed and then adjusted for age, sex and study centre and in our sensitivity analysis additionally for BMI or alcohol consumption. We then inverse normal transformed the values. At runtime, we included genotyping array (as a categorical variable for UKBileve array, UKB Axiom array interim release and UKB Axiom array full release) as a covariate.

We used PLINK to perform a sensitivity analysis. Prior to association testing, liver iron was adjusted for age, sex, BMI and genotyping array, and then we quantile normalised the resulting values. At runtime, we included the first 10 genetic PCs (UK Biobank field 22009) as covariates to control for confounding by population stratification. We used Quanto (http://biostats.usc.edu/Quanto.html) to calculate our discovery GWAS power in 8,289 individuals from UK Biobank at different allele frequencies and effect sizes at α level 5×10^{-8} assuming an additive effect model (Figure 7.1.).

3.6.5 LD Score regression and cross-trait genetic correlation analysis

We used LD Hub to conduct linkage disequilibrium (LD) score regression and heritability analysis. LD Hub is a centralized database of summary level GWAS for > 100 diseases/traits from publicly available resources/consortia and uses a web interface that automates LD score regression, heritability and cross-trait genetic correlation analysis pipeline.[139] We ran heritability analysis as well as genetic correlation analysis across 448 potentially relevant traits.

SNP-based heritability (h^2_{SNP}) is the proportion of total variation in liver iron content due to the additive genetic variation between individuals in our study population.

3.6.6 Gene-set and tissue expression enrichment analysis

We performed gene-set and tissue expression analyses using MAGMA.[140] Lead variants were assigned to a minimum P of $5x10^{-8}$. We used the default settings provided by the software. We chose 1000 Genomes Phase 3 as the reference panel population. The minimum minor allele frequency was set to 0.01. We used a maximum allowed distance of 250k between LD blocks for variants to be included in the same locus.

For gene-set enrichment analyses, positional mapping was used with variants assigned to a gene if they were within the gene start and end points (by setting the distance either side to 0kb). Only protein-coding genes were included in the mapping process. Tested gene-sets include BioCarta, REACTOME, KEGG and GO. Bonferroni correction was used to adjust for the number of gene-sets tested. Analysis of differentially expressed genes was based on GTEx v6 RNA-seq data.[84]

3.6.7 Replication analysis

Associations reaching $P < 5x10^{-8}$ were followed up in the IMI DIRECT cohort, who provided us with the relevant variant results from their GWAS. A summary of their GWAS methodology is summarised below. IMI DIRECT includes 1,513 participants who had both liver iron and GWAS data to replicate our findings. The IMI-DIRECT consortium is a collaboration among investigators from a range of European academic institutions and pharmaceutical companies.[135] Liver iron was measured using a T2*-based multiecho MRI technique.[141] DNA extraction was carried out using Maxwell 16 Blood DNA purification kits and a Maxwell

16 semi-automated nucleic acid purification system (Promega). Genotyping was conducted using the Illumina HumanCore array (HCE24 v1.0) and genotypes were called using Illumina's GenCall algorithm. A total of 517,958 markers passed quality control procedures. We took autosomal variants with MAF>1% that passed quality control and constructed axes of genetic variation using PC analysis implemented in the GCTA software to identify ethnic outliers defined as non-European ancestry using the 1000 Genomes samples as reference. We identified six individuals as ethnic outliers. We performed the association tests in 3 models: (i) non-diabetic participants (N=1,010), (ii) diabetic participants (N=503), and (iii) combined (N=1,513). We took residuals from a model of liver iron and age, sex, BMI, 10 PCs and Centres and then inverse-normal transformed the values.

3.6.8 Sensitivity Analyses

We performed 4 sensitivity analyses. First, to assess whether there is any sex-specific association, we carried out GWASs in men and women separately. Second, to test whether relatedness was responsible for any of the individual variant associations and replicate GEMMA's results, we ran a GWAS using only unrelated European individuals in UK Biobank and a different GWAS software tool (PLINK version 1.9). Third, we adjusted models for alcohol intake frequency (field 1558; categories treated as ordinal scale – "Never"=0 to "Daily or almost daily"=5) measured at baseline that may have had an impact on liver iron. Individuals responding "Do not know" or "Prefer not to answer" for "Alcohol intake frequency" were excluded from this sensitivity analysis. Fourth, to investigate the potential for collider bias resulting from conditioning liver iron on BMI, we performed a GWAS of liver iron without adjustment for BMI.

3.6.9 Mendelian randomisation

Multiple traits have shown association with liver iron in observational studies, including BMI, lipids and NAFLD.[36,38,142,143] We therefore investigated the causal effects of 29 predominantly metabolic traits using two-sample Mendelian randomisation analysis.[78] Mendelian randomisation is a method that uses genetic variants associated with the exposure (e.g. metabolic traits) to infer causal relationships between an exposure and an outcome (e.g. liver iron content). This method relies on a simple principle; if a modifiable exposure is causal for a disease, then the genetic variants associated with that exposure will also be associated with disease risk. Since genetic variants are inherited at birth, Mendelian randomisation experiments are free from confounding and biases that are seen in observational studies.

Following correction for multiple testing, associations with a false discovery rate (FDR) <5% were considered statistically significant. We used the inverse variance weighted approach (IVW) as our main analysis, and MR-Egger and penalised weighted median as sensitivity analyses in the event of unidentified pleiotropy of our genetic instruments. Genetic instruments for the 28 metabolic traits as an exposure were constructed by developing risk scores using only genome-wide significant SNPs that were not in linkage disequilibrium (R^2 <0.1).[144–146] Analysis was performed using MR Base, an automated web tool and R package which supports Mendelian randomisation and phenome wide studies.[81]

3.6.10 Phenome-wide association study (PheWAS)

We used the SNPs associated with liver iron content and carried out a PheWAS using publicly available summary statistics from GWASs on predefined ICD10 disease codes, anthropometric traits, and self-reported conditions previously carried out in 452,264 UK Biobank participants

of European ancestry,[147] as well as publicly available, curated summary statistics from previous GWAS (Table 7.1.2., Table 7.1.3., Table 7.1.4.).[81] Associations with a false discovery rate (FDR) < 5% were considered statistically significant. A description of how ICD codes were grouped to represent a clinical phenotype are provided elsewhere.[147] Analysis was performed using MR Base, an automated web tool and R package which supports Mendelian randomisation and phenome wide studies.[81]

3.7 Results

3.7.1 The characteristics of liver iron content cohort.

The median liver iron content in UK Biobank was 1.28 mg/g (interquartile range (IQR):1.16-1.44 mg/g) in men (N=3,928) and 1.23 mg/g (IQR: 1.13-1.38) in women (N=4,361) (Table 3.7.1, Figure 7.2). Among men 6.5% and among women 3.4% had an elevated liver iron content, above the commonly accepted 1.8mg/g threshold.[148] In IMI DIRECT cohort, the median liver iron content was 1.3 (1.2-1.5) in both men (N=1,101) and women (N=412). BMI, waist circumference and diabetes prevalence were lower in the liver iron cohort (N=8,289) than the remainder of UK Biobank (N=402,071) (Table 7.1.5.). Although invitation was not based on any medical information, MRI exclusion criteria (e.g. metal or electrical implants, surgery in six weeks prior to appointment, severe hearing or breathing problems) may have also contributed to a slightly healthier cohort. The Townsend deprivation index was on average lower in this study cohort. This may be related to MRI imaging participants being biased towards those who live close to the imaging center (Cheadle) where all the liver iron cohort were imaged.

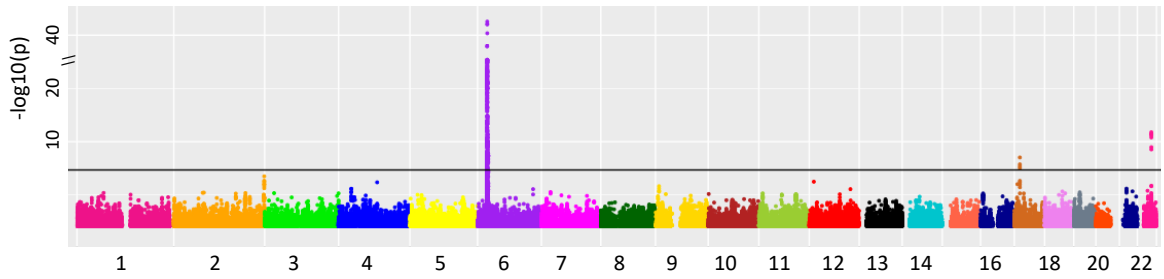
Table 3.7.1. Characteristics of UK Biobank and IMI DIRECT study participants.

Characteristics	**UK Biobank liver iron cohort**		**DIRECT**	
	Men	**Women**	**Men**	**Women**
No	3,928	4,361	1,101	412
Age (IQR) (years)	57 (51,62)	56 (49,61)	62 (56,66)	62 (57,67)
Liver iron (IQR) (mg/g)	1.28 (1.16,1.44)	1.23 (1.13,1.38)	1.3 (1.2, 1.5)	1.3 (1.2, 1.5)

Waist Circumference (IQR) (cm)	94 (88, 101)	80 (74, 89)	101 (95,109)	97 (88,108)
Townsend deprivation index (IQR)	-2.72 (-3.95, -0.76)	-2.63 (-3.87,-0.80)	NA	NA
Self-reported diabetes (%)	134 (3.7%)	88 (2.2%)	287 (26%)	216 (52.4%)
BMI (IQR) (kg/m2)	26.49 (24.3, 29)	25.08 (22.59,28.35)	27.8 (25.8, 30.5)	28.7 (25.8,33.2)
No consuming alcohol daily (%)	1,088 (27.7%)	825 (18.9%)	129(15.6%)	25(10.7%)

***N = Number, IQR = interquartile range**

Figure 3.2. Manhattan plot illustrating genetic variants (~30 million imputed SNPs) associated with liver iron in UK Biobank. The x-axis is the chromosomal position and y axis is -log(P) for the association with each variant. Black line indicates genome-wide significance level (5 x 10^{-8}).



3.7.2 There are three genetic variants associated with liver iron content.

We performed a GWAS of MRI derived measures of liver iron using 8,289 individuals of European ancestry from UK Biobank (Figure 3.2., Figure 7.3.). We estimated to have more than 80% power in our discovery set to detect variants with MAF ≥ 5% and effect size ≥ 0.2 standard deviation (SD) on liver iron content (Figure 7.1.). We detected no evidence for inflation of test statistics (λ_{GC} = 1.016). Our discovery GWAS identified 4 independent variants at P < $5x10^{-8}$ (Table 3.7.2.). Two independent variants lie within *HFE*: C282Y (rs1800562; 0.41 SD increase in liver iron per allele; p = 5.2 x 10^{-42}) and H63D (rs1799945; 0.17 SD; p = 8.2 x 10^{-15}). The third variant lies in *TMPRSS6*; V736A (rs855791; 0.11 SD; p = $1.3x10^{-11}$). The fourth variant, rs149275125, lies between *HS3ST3B1* and *PMP22* (0.41 SD; p = $3x10^{-9}$). This variant is a rare variant (MAF 1%) and has not previously been reported to be associated with any other traits.

Table 3.7.2. Genome-wide significant independent variants associated with MRI liver iron content in UK Biobank (P < $5x10^{-8}$) and validation in IMI DIRECT.

					UK Biobank (N = 8,289)				**DIRECT (N = 1,513)**			
SNP	**Gene**	**Chr**	**EA**	**OA**	**EAF**	**BETA**	**SE**	**P**	**EAF**	**BETA**	**SE**	**P**
rs1800562	*HFE*	6	A	G	0.08	0.41	0.03	$5.2x10^{-42}$	0.04	0.35	0.08	$05x10^{-5}$
rs1799945	*HFE*	6	G	C	0.15	0.16	0.02	$8.2x10^{-15}$	0.12	0.19	0.05	$2x10^{-4}$
rs855791	*TMPRSS6*	22	G	A	0.56	0.11	0.02	$1.3x10^{-11}$	0.59	0.12	0.04	$4x10^{-4}$
rs149275125	*HS3ST3B1----[]----PMP22*	17	C	T	0.98	0.41	0.07	$3.1x10^{-9}$	0.99	-.22	0.2	0.27

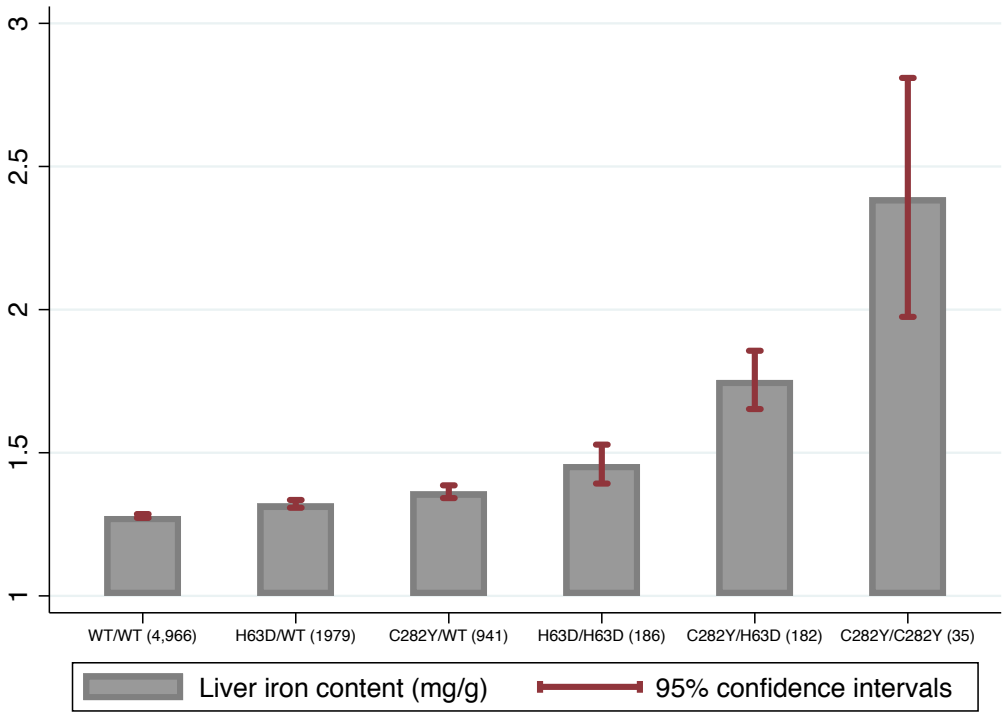
***Chr = chromosome, EA = effect allele, OA = other allele, EAF = effect allele frequency, Beta = per allele effect on liver iron (SD), SE = standard error, P = p- value.**

In 1,513 IMI DIRECT participants, we replicated all three common variants at $p < 4x10^{-4}$ with a consistent direction of effect and similar effect size (Table 7.1.6.). The rare variant did not associate with liver iron in IMI DIRECT and the direction of effect was opposite to our discovery dataset. We focused all other analyses on the three replicated variants.

Both *HFE* and *TMPRSS6* produce proteins that form part of the signalling pathway regulating hepcidin production, the key hormone responsible for iron balance in the body. C282Y homozygotes and C282Y/H63D compound heterozygotes account for ~85% of cases of hereditary haemochromatosis.[34] In UK Biobank, 35 individuals (0.4%) were C282Y homozygotes and had highest levels of liver iron (mean: 2.39 mg/g (+-1.2)); 182 (2.2%) were C282Y/H63D compound heterozygotes (1.75 mg/g (+-0.7)); 186 (2.2%) were H63D homozygotes (1.46 mg/g (+-0.47)); 2920 (35%) were either C282Y or H63D heterozygotes (1.34 mg/g (+-0.32)) and 4966 (60%) did not have any of the variants and had the lowest liver iron (1.28 mg/g (+-0.24)) (Figure 3.3.).

Figure 3.3. Liver iron content per genotype group. X-axis are the 6 genotypes groups based on the number of C282Y and H63D they carry. Y-axis is the mean of liver iron (mg/g) per category.

Error bars indicate 95% confidence intervals. Numbers in brackets are the number of individuals per genotype category.



Correlation between the effect sizes and p-values in the two separate GWASs carried out in GEMMA and PLINK showed strong agreement (Figure 7.4, Figure 7.5.). We did not detect any sex-specific variants and the magnitude of effect was similar between men and women (Table 7.1.6., Figure 7.6.). The sensitivity analysis adjusting for alcohol consumption and BMI did not identify any additional signal and did not change the effect size (Table 7.1.6.). Our pathway analysis demonstrates overlap between liver iron gene-sets and pathways involved in autism and schizophrenia (Figure 7.7.). Nearby genes were visualised with locuszoom plots (Figure 7.8., Figure 7.9.).

3.7.3 Liver iron content is heritable and has a high genetic correlation with blood levels of iron biomarkers.

We estimated the SNP-based heritability (h^2_{SNP}) of liver iron to be 7%. This is similar to heritability estimated for conditions and traits such as coronary artery disease (7%)[149] and eczema (7%),[150] but lower than heritability estimated for body fat % (10%)[151] and transferrin (16%).[74]

To identify genetic overlap between liver iron content and other diseases and traits, we performed LD score regression analyses against a range of available traits and diseases with GWAS summary statistics (448 traits/diseases, Table 7.1.7.). The most genetically correlated traits were transferrin (r_G=-0.78, *P*=0.04) and ferritin (r_G=1.24, *P*=0.05) with nominal significant correlation. Joint disorders (r_G=-1.17, *P*=0.50), hypertrophic cardiomyopathy (r_G=-1.11, *P*=0.35), type 2 diabetes (r_G=0.44, *P*=0.17), chronic kidney disease (r_G=0.57, *P*=0.48), tinnitus (r_G=0.65, *P*=0.17), polyuria (r_G=0.75, *P*=0.36), and gout (r_G=0.90, *P*=0.19) were highly correlated (r_G>0.4) but did not reach a nominal significance threshold (*p*>0.05). Metabolic traits including fasting insulin (r_G=0.17, *P*=0.53), HOMA-IR (r_G=0.37, *P*=0.48), fasting glucose (r_G=0.01, *P*=0.97) and coronary artery disease (r_G=-0.01, *P*=0.97) were not genetically correlated with liver iron content.

3.7.4 Gene-set enrichment analysis did not identify any enriched tissue or pathways

We used MAGMA implemented as part of the FUMA GWAS platform to assess tissue enrichment of genes at associated loci. We did not find any tissue enrichment, but differentially expressed gene sets were enriched in blood vessels, lung, and adipose tissue, although they did not reach a significant threshold following adjustment for multiple testing (Table 7.1.8.). None

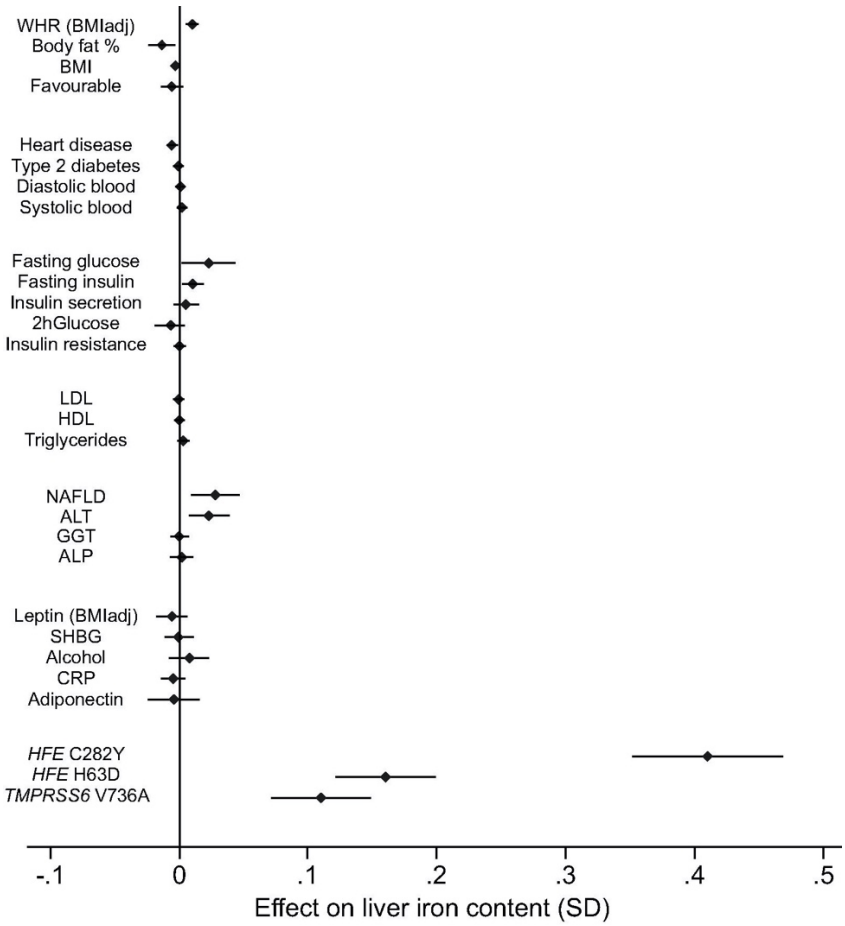
of the pathways reached our FDR significance threshold (Table 7.1.9.). This could be due to the limited power in our GWAS analysis.

3.7.5 Mendelian randomisation analysis provides evidence for a causal link between central obesity and liver iron content.

We examined the potential causal effect of 25 metabolic traits and diseases (Figure 3.4.), Table 7.1.10.) on liver iron content. Following correction for multiple testing (FDR<5%), we found evidence of a causative effect of central obesity, as measured by higher waist-to-hip ratio (adjusted for BMI), on elevated liver iron content (IVW *P*=0.003) (Table 7.1.9., Figure 7.10.). There was suggestive evidence that higher fasting glucose (IVW *P*=0.03), higher NAFLD (IVW *P*=0.04) and higher alanine transaminase (IVW *P*=0.05) were causally associated with higher liver iron content but none of these associations reached our multiple testing threshold of being statistically significant. As a positive control, elevated transferrin saturation levels (IVW *P*=0.0007), blood iron content (IVW *P*=0.01) and ferritin levels (IVW P=0.01) were causally associated with higher liver iron content. All the results were robust to a range of Mendelian randomisation sensitivity analyses.

Figure 3.4. Mendelian randomisation investigating the effect of 25 predominantly metabolic traits and diseases on liver iron content (standard deviation (SD)). We used two-sample Mendelian randomisation analysis to investigate the causal effects of 28 predominantly metabolic traits on liver iron. The X-axis list 28 exposures and Y-axis shows the results from the inverse variance

weighted approach (IVW) as our main analysis. The error bars indicate 95% confidence intervals (for full results, including sensitivity analyses, please see Table 7.1.10.).

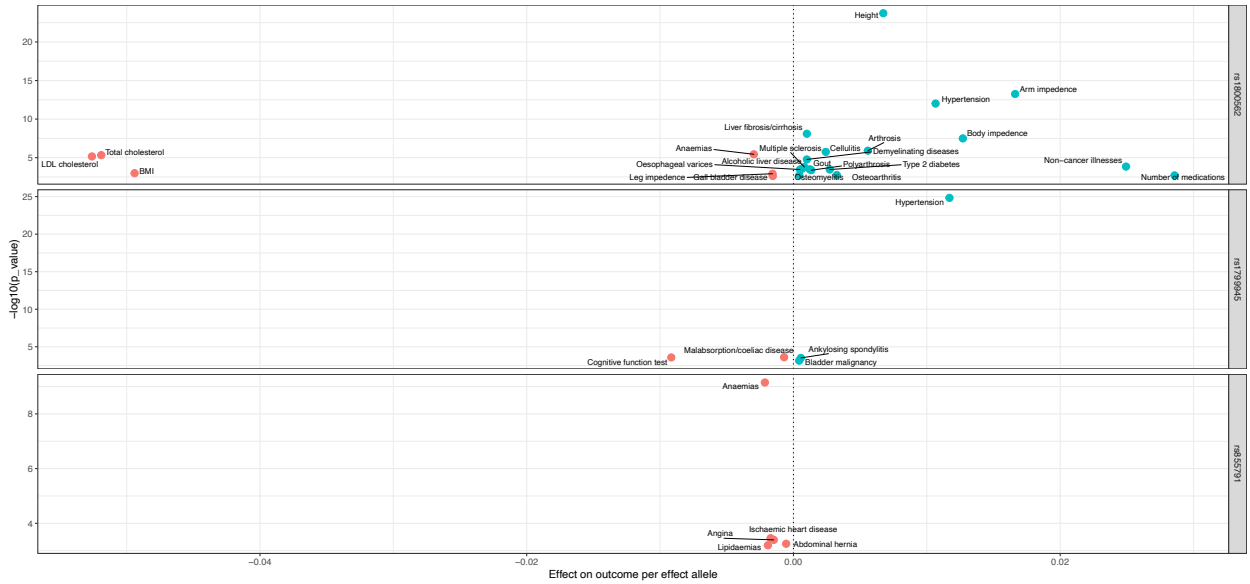


3.7.6 PheWAS identifies novel associations of liver iron variants with traits and diseases.

The three liver iron content variants had been previously reported to be associated with multiple haematological parameters, glycated haemoglobin, lipid and bilirubin levels, as well as blood pressure traits (Table 7.1.11.). We performed a hypothesis-free PheWAS to investigate the association of these three variants with other traits and diseases using predefined ICD10 codes, self-reported conditions and traits from UK Biobank and publicly available GWAS.

HFE C282Y was associated with higher liver fibrosis/cirrhosis, higher risk of type 2 diabetes, hypertension, alcoholic liver disease, arthrosis, chronic and degenerative neurological problem including multiple sclerosis, arthritis and higher height but lower total cholesterol, lower LDL-C and lower BMI (FDR<5%, Figure 3.5., Table 7.1.2.). *HFE* H63D was associated with higher risk of hypertension, ankylosing spondylitis and bladder malignancy but lower risk of malabsorption or coeliac disease and lower cognitive ability (FDR<5%, Figure 3.5., Table 7.1.3.). The liver iron increasing allele at *TMPRSS6* was associated with lower risk of ischaemic heart disease, angina pectoris, and lipidaemias (FDR<5%, Figure 3.5., Table 7.1.4.).

Figure 3.5. Illustration of prioritised associations following phenome-wide association studies (PheWAS) of rs1800562, rs1799945 and rs855791 and significant traits from UK Biobank and publicly available summary statistics. Blue indicates a positive association and red an inverse association, following correction for multiple testing (False discovery rate < 5%). Continuous traits betas were scaled to per SD where appropriate for better visualisation. Effect on disease risk is given in log(odds ratio).



3.8 Discussion

We performed the first GWAS of multi-parametric MRI determined liver iron content in an unselected population. The identification of loci implicated in increased iron absorption (*HFE* and *TMPRSS6*) provides genetic validation of the utility of MRI for the non-invasive assessment of liver iron content.

The three independent variants in *HFE* and *TMPRSS6* have previously been reported to be associated with circulating iron traits including transferrin saturation, blood iron, ferritin, and transferrin levels.[74] Both *HFE* and *TMPRSS6* play a major role in iron homeostasis by modulating the expression of hepcidin production by the liver.[Citation error] Hepcidin inhibits iron transport and absorption from the gut into the circulation by binding to the main iron transport channel expressed on the surface on duodenal enterocytes, ferroportin.[152] *TMPRSS6* encodes matriptase 2 (MTP-2), a liver serine protease which inhibits hepcidin leading to higher iron absorption and bioavailability. In vitro studies have shown that major allele at rs888571 inhibits hepcidin more effectively than missense variant rs888571(V736A).[153] *HFE* is a positive upstream regulator of hepcidin transcription. Missense variants C282Y and H63D in *HFE* result in lower hepcidin responsiveness to iron, leading to relative or absolute hepcidin deficiency and subsequent increase in iron absorption and bioavailability.[154]

Elevated liver iron is observationally associated with multiple metabolic traits and diseases in a common condition described as DIOS.[38] Our Mendelian randomisation analysis supports a causal role for higher central obesity on excess liver iron, providing further evidence for DIOS. Other traits such as fasting glucose, NAFLD, and alanine transaminase showed

suggestive causal associations. Animal studies have suggested a putative mechanism through defective iron handling and subsequent iron overload due to an inflammatory shift and cytokine secretion by activated macrophages that accumulate around adipocytes in obesity.[155,156] The underlying mechanism, however, is still unclear, and is likely to involve a complex interplay between diet and genetic factors as well as cross-talk between liver and visceral adipose tissue.[38]

Our GWAS study identified variants that are likely to regulate iron stores systemically and are not specific to the liver. We, therefore, were not able to examine the causal role of higher liver iron content *per se* on other diseases and traits using Mendelian randomisation. To investigate, however, the phenotypic architecture and shared pathological mechanisms of higher liver iron content with other traits and diseases, we carried out a PheWAS of the 3 genome-wide significant variants against all available disease outcomes and traits from UK Biobank[147] and publicly available genetic summary statistics.[81]

Our PheWAS indicates *HFE* C282Y is associated with arthrosis, coxarthrosis, osteoarthritis, and gout, and *HFE* H63D is associated with ankylosing spondylitis and has a suggestive association with dorsalgia. The association between *HFE* C282Y and higher risk of cellulitis, abscesses, furuncles and curbuncles, subcutaneous infections as well as osteomyelitis provides further evidence that genetically elevated iron levels are associated with higher infection risk. Some infectious disease agents are more virulent in an environment with excess iron. There is also evidence that iron overload compromises the ability of phagocytes to kill microorganisms.[157]

Several observational studies have demonstrated iron overload is correlated with carcinogenesis.[158] An important pathogenic mechanism may be oxidative stress and the catalytic activity of iron in the formation of hydroxyl radicals. Iron may also suppress host defence cell activity and promote cancer cell proliferation. A recent meta-analysis found *HFE* C282Y to be significantly associated with higher risk of breast cancer, colorectal cancer, hepatocellular carcinoma, and total cancer risk.[159] We found additional evidence of associations with extra-hepatic malignancies; *HFE* H63D was associated with higher risk of bladder cancer (OR=1.0004 per effect allele, p=4.7x10^{-6}, FDR 0.03) and *HFE* C282Y with higher risk of renal cancer (OR=1.0005 per effect allele, p=0.004, FDR 0.05). Despite very small effects , these findings provide genetic evidence for shared mechanisms between higher liver iron and extra-hepatic cancers.

The association between *HFE* C282Y and neurological conditions such as multiple sclerosis and epilepsy is consistent with the role of iron in many important processes in the central nervous system (CNS), including oxygen transportation, oxidative phosphorylation, myelin production, and the synthesis and metabolism of neurotransmitters. In a recent GWAS of brain MRI scans, *HFE* C282Y variant was associated with iron accumulation in certain parts of the brain.[160] Observational studies show that iron accumulation in the brain is associated with multiple sclerosis, parkinsonism and Alzheimer's disease.[161,162] Individuals with hereditary haemochromatosis frequently develop psychological symptoms, including extreme fatigue, irritability and depression.[163] Our pathway analysis demonstrates overlap between liver iron gene-sets and pathways involved in autism and schizophrenia. *HFE* H63D was associated with a reduction in reaction time in specific cognitive function tests, providing further evidence that iron accumulation may cause premature, and indeed preventable,

cognitive decline. This data highlights the need for early identification of individuals at risk of excess iron accumulation in the brain.

The association between *HFE* C282Y and H63D variants and hypertension is consistent with previous findings.[164] Possible mechanisms include increased vascular tone secondary to the generation of reactive oxygen species and oxidative stress,[165] or excess iron accumulation in renal arterioles leading to activation of the renin-angiotensin aldosterone system. We further validated the known association between *HFE* C282Y and type 2 diabetes which could be, at least partly, due to iron accumulation in the pancreas.

The liver iron increasing allele at *TMPRSS6* rs855791 was associated with lower LDL-C, lower risk of angina and ischaemic heart disease. Similar observations have been previously reported in *HFE* C282Y homozygotes.[166] A Mendelian randomisation study has recently reported that elevated circulating iron may have a causal (protective) effect on coronary artery disease.[167] The underlying mechanism is unclear. It is possible that part of this effect is driven through effects on haematological parameters, or lower LDL-C. In conditions where excess iron stores are treated (e.g. hereditary haemochromatosis), further research is needed on whether LDL-C levels subsequently increase, and whether the risk of coronary artery disease can be kept low with statins and other preventive interventions.[166]

This study is limited in that the UK Biobank MRI cohort is not a completely unbiased sample of the population. The UK Biobank MRI cohort is slightly more healthy, wealthy, and well educated compared to the whole cohort of 40-69 year olds in the UK.[168] The population studied in this work has a slightly lower average BMI and waist circumference than the UK Biobank population as a whole, while the average Townsend Deprivation Index is slightly

lower in this study cohort than the remainder of the UK Biobank. Larger GWAS studies (e.g. on completion of the full 100,000 UK Biobank imaging cohort) may elucidate further susceptibility loci. Ongoing development and validation of MRI scores that may allow accurate determination of the level of inflammation and fibrosis, may lead to further genetic studies focussing on the more severe spectrum of liver disease.

3.9 Conclusion

We report a large GWAS for MRI liver iron content and identify 3 susceptibility loci previously implicated in circulating iron traits. We provide genetic validation for multi-parametric MRI as a novel, non-invasive and radiation free imaging modality for liver iron content. Our genetic study suggests that higher liver iron content may be caused in part by higher central adiposity. The deposition of excess iron in the liver seems to share common mechanisms with circulating iron accumulation, which eventually results in widespread damage to parenchymal tissue, leading to several pathologies through a common mechanism.

3.10 Data availability

Full data have been returned to UK Biobank and are publicly available via application (amsportal.ukbiobank.ac.uk).

4 Genome wide and Mendelian randomisation analysis of magnetic resonance imaging of the liver yield insights into the pathogenesis of steatohepatitis.

4.1 Abstract

4.1.1 Background & Aims

A non-invasive method to grade the severity of steatohepatitis and liver fibrosis is magnetic resonance imaging (MRI) based corrected T1 (cT1). We aimed to identify genetic variants influencing liver cT1 and use genetics to understand mechanisms underlying liver fibroinflammatory disease and its link with other metabolic traits and diseases.

4.1.2 Methods

First, we performed a genome-wide association study (GWAS) in 14,440 Europeans in UK Biobank with liver cT1 measures. Second, we explored the effects of the cT1 variants on liver blood tests, and a range of metabolic traits and diseases. Third, we used Mendelian randomisation to test the causal effects of 24 predominantly metabolic traits on liver cT1 measures.

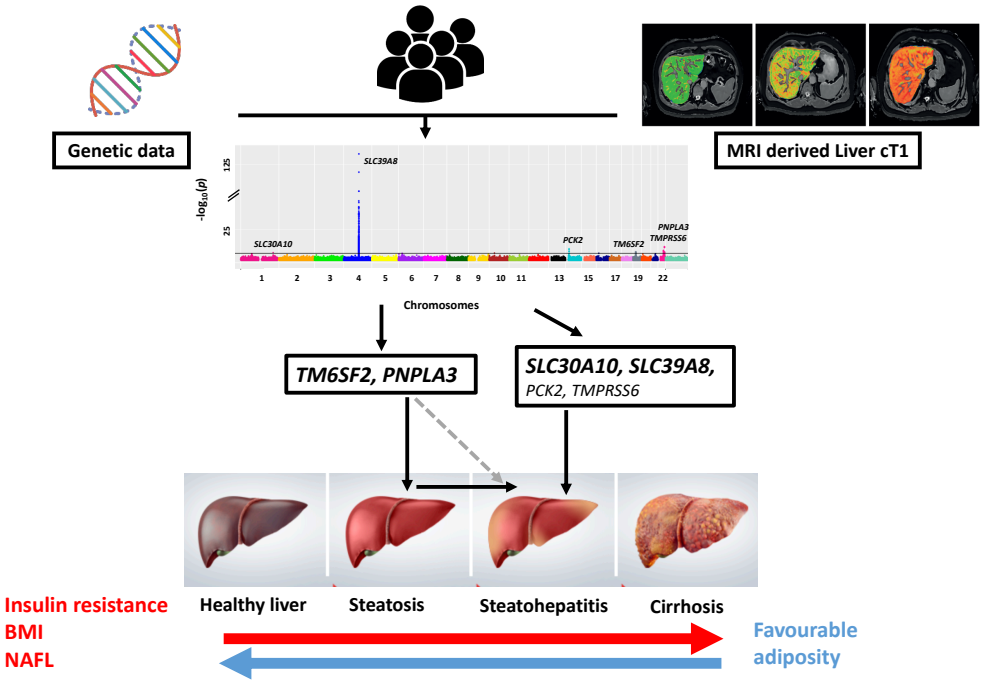
4.1.3 Results

We identified six independent genetic variants associated with liver cT1 that reached GWAS significance threshold ($p < 5x10^{-8}$). Four of the variants (rs75935921 in *SLC30A10*, rs13107325 in *SLC39A8*, rs58542926 in *TM6SF2*, rs738409 in *PNPLA3*) were also associated with elevated transaminases and had variable effects on liver fat and other metabolic traits. Insulin resistance, type 2 diabetes, non-alcoholic fatty liver, and BMI were causally associated with elevated cT1 whilst favourable adiposity (instrumented by variants associated with higher adiposity but lower risk of cardiometabolic disease and lower liver fat) was found to be protective.

4.1.4 Conclusion

The association between two metal ion transporters and cT1 indicates an important new mechanism in steatohepatitis. Future studies are needed to determine whether interventions targeting the identified transporters might prevent liver disease in at risk individuals.

4.2 Visual Abstract



4.3 Lay summary

We estimated levels of liver inflammation and scarring based on magnetic resonance imaging of 14,440 UK Biobank participants. We performed a genetic study and identified variations in six genes associated with levels of liver inflammation and scarring. Participants with variations in four of these genes also had higher levels of markers of liver cell injury in blood samples, further validating their role in liver health. Two identified genes are involved in the transport of metal ions in our body. Further investigation of these variations may lead to better detection, assessment, and treatment of liver inflammation and scarring.

4.4 Highlights

- Variants in metal ion transporters and NAFLD genes are associated with liver MRI-derived cT1, a steatohepatitis and fibrosis proxy.
- cT1 is highly heritable, and is correlated with BMI, NAFLD and VLDL, and inversely correlated with HDL.
- Insulin resistance, NAFLD and higher BMI are genetically linked to higher liver cT1, whilst favourable adiposity is linked to lower cT1.

4.5 Introduction

Non-alcoholic and alcoholic fatty liver diseases are common in an era of a global obesity epidemic and concerning alcohol use.[169,170] They affect up to a third of the adult population worldwide and account for the vast majority of chronic liver diseases.[171] However, an important paradox in the history of liver fat accumulation exists; despite the large proportion of adults affected by simple steatosis (fatty liver), only a relatively small proportion (2.4 - 12.8%) will experience significant liver disease or liver related death.[18]

It is important to identify which individuals are at risk of developing the more inflammatory phenotype, steatohepatitis, which is a condition characterised by lipotoxicity and histological necroinflammation and is considered to be the main pathophysiological driver of liver fibrosis and subsequent disease progression.[172] Steatohepatitis and fibrosis affect approximately one in ten middle-aged adults, and can lead to cirrhosis, HCC and death.[173]

A promising, non-invasive measure of steatohepatitis and fibrosis severity is magnetic resonance imaging (MRI) based corrected T1 (cT1) (Figure 3.2).[8–10] T1 relaxation time reflects extracellular fluid which is characteristic of fibrosis and inflammation. The presence of iron, which can be determined from T2* maps, has an opposing effect. Combining T2* and T1 values can correct for this opposing effect, from which cT1 (in milliseconds) is derived. Higher cT1 values are associated with both histological liver inflammation and fibrosis, although their relative contributions to the score are still unknown.[10,11] cT1 has already been used as a non-invasive outcome in RCTs for non-alcoholic steatohepatitis (NASH)[12] and is associated with liver disease outcomes.[9]

Understanding the underlying genetic susceptibility of steatohepatitis and fibrosis may allow new insights on the main pathophysiological mechanisms contributing to chronic liver disease and help identify potential new drug targets. Genetic studies have so far been limited due to the phenotyping challenge. Liver biopsy is an invasive procedure with associated risks, significant sampling error and marked interobserver variance,[50] while routinely available LBTs such as aminotransferases, despite being useful in the identification of important liver disease susceptibility loci, are overall poor predictors of liver disease severity.[51,52]

Another challenging question is which metabolic traits cause steatohepatitis since treating causal factors can help prevent liver disease. Observational associations between steatohepatitis and other features of the metabolic syndrome might occur because they share common risk factors, rather than one causing the other. Mendelian randomisation (MR) is an established epidemiological approach that uses genetic studies to provide insight on causality.[75] MR uses genetic variants associated with an exposure (e.g. BMI, LDL cholesterol, insulin resistance) to assess their causal effect on an outcome of interest (e.g. cT1, steatohepatitis). Genetic markers of a risk factor are largely independent of confounders that may otherwise cause bias since genetic variants are randomly allocated before birth. Furthermore, the non-modifiable nature of genetic variants provides an analogy to randomised trials, in which exposure is allocated randomly and is non-modifiable by subsequent disease.[76]

In this study, we aimed to (i) identify genetic variants influencing liver cT1 (ii) identify the effect of liver cT1 variants on other metabolic traits, (iii) investigate which metabolic traits are genetically correlated with cT1 measures and (iv) use MR to investigate whether 24 metabolic traits and conditions are causally associated with cT1. We performed the first genome-wide

association study (GWAS) on MRI liver cT1 in 14,440 European individuals from UK Biobank. Finally, to investigate whether there are shared variants between liver cT1 and liver fat, we carried out a GWAS on MRI determined liver proton density fat fraction (PDFF) in the same cohort.

4.6 Methods

4.6.1 UK Biobank participants

UK Biobank is a prospective cohort study that consists of over 500,000 individuals aged 37–73 years (99.5% were between 40 and 69 years of age) who were recruited between 2006 and 2010 from across the U.K.[44] This research has been conducted using the data obtained via UK Biobank Access Application number 9914 and 31037. The UK Biobank has approval from the North West Multi-Centre Research Ethics Committee (ref: 11/NW/0382) and obtained written informed consent from all participants prior to the study.

4.6.2 Imaging protocol and analysis

Invitation to the UK Biobank imaging study is based only on proximity to one of the main imaging sites. Participants were invited and scanned at the UK Biobank Imaging Centre in Cheadle (UK) using a Siemens 1.5T Magnetom Aera as previously described.[174,175] Medical conditions were not taken into account except from those which would exclude the participant from being able to have an MRI (e.g. if they had an implanted defibrillator or metal implant).

Characterisation of cT1 in the UK Biobank cohort, alongside normal values and inter and intra-reader variability have previously been published.[174] Briefly, two sequences were used to acquire data: a Shortened Modified Look Locker Inversion (ShMOLLI) to quantify liver T1, and a multiecho-spoiled gradient-echo, to quantify liver iron and fat (PDFF). In both cases, data was acquired as a single transverse slice captured through the centre of the liver superior to the porta hepatis. Acquisition was performed in end-expiration breath-hold and without the aid of any contrast agent injection. The slice-based methodology has previously been shown to correlate well with histology and predict liver outcomes.[8,10]

The MRI sequence is part of the Liver*MultiScan*© protocol from Perspectum Diagnostics (UK) which forms part of the UK Biobank abdominal imaging protocol.[174,176,177] The data was analysed by a team of trained analysts blinded to any participant variables, using Liver*MultiScan*© Discover 4.0 software. This software creates T2*, cT1 and PDFF maps from the image data, and produces an automated delineation of the liver excluding its major vessels within the image slice using a deep learning approach which has previously been published;[178] The median value from this delineation on the T2* map is converted to an iron value,[179] which is used with the ShMOLLI data to derive the cT1 map.[180] All values reported in this work are the median, for each metric, of all usable voxels in the liver within the image slice. T1 relaxation time reflects extracellular fluid and is characteristic of fibrosis and inflammation. The presence of iron, which can be determined from T2* maps, has an opposing effect on the T1, and algorithms have been formed to correct for the resulting bias.[10] All processed data are available through application to UK Biobank. Figure 3.2**.** illustrates the 3 MRI scans with different levels of cT1 in 3 participants. The results were provided by our collaborator Dr. Henry Wilman and Perspectum Diagnostics.

From an initial collection of 20,386 imaging sessions (each of a unique individual), 691 did not have all necessary image data, 1354 were run with an early flawed protocol, 1717 did not correctly trigger the sequence, 126 had more than half of their liver excluded due to poor model fitting and motion artefacts, leaving 16498 for human quality control.

From these a further 959 were removed through a combination of fat/water swaps, erroneous overcorrection of iron, misplacement of the image slice, segmentation failure, field artefacts, and cysts within the image slice preventing reasonable quantification of parenchyma leaving 15539 participants.

4.6.3 Genetic Data

Protocols for the participant genotyping, data collection, and quality control have previously been described in detail.[44] Briefly, participants were genotyped using one of two purpose-designed arrays (UK BiLEVE Axiom Array (n= 50,520) and UK Biobank Axiom Array (n = 438,692)) with 95% marker overlap. We excluded individuals who were identified by UK Biobank as outliers based on either genotyping missingness rate or heterogeneity, or whose sex inferred from the genotypes did not match their self-reported sex. We removed individuals with a missingness > 5% across variants which passed our quality control procedure. We used the latest release which included imputed data using two reference panels: a combined UK10K and 1000 Genomes panel and the Haplotype Reference Consortium (HRC) panel. We limited our analysis to genetic variants with a minimum minor allele frequency (MAF) > 1% and imputation quality score (INFO) > 0.3.

To define “white European” ancestry participants, we first used data from 1000 genomes samples to generate ancestry informative principal components (PCs). We then used these PCs

in UK Biobank participants and employed K-means clustering to identify samples clustered with the three main 1000 genomes populations (European, African, and South Asian). Those clustered with the 1000 genomes' "European" cluster were classified as having European ancestry.

In total, after image analysis and quality control steps, liver cT1 and PDFF measures were available for 14,440 white European individuals who also had genetic data available and were classified as white European.

4.6.4 Genome-wide association analysis

We used BOLT-LMM v2.3.4 to conduct a linear mixed model GWAS which accounts for population structure and relatedness. We increased our power by including all related individuals of European descent (n = 14,440). The relatedness matrix was computed using common (MAF>5%) genotyped variants that passed quality control in all 106 batches and were present on both genotyping arrays. Prior to association testing, liver cT1 and PDFF were inverse-normal transformed. We used age, sex, centre and genotyping array as covariates in the model. For this analysis, we used an automated pipeline optimised for GWAS analysis by the senior author of the published paper, Dr Hanieh Yaghootkar and the Frayling lab, University of Exeter.

4.6.5 Sensitivity Analyses

We performed 6 sensitivity analyses (Table 7.2.1.). We carried out GWASs and adjusted for (i) BMI and (ii) alcohol units consumed. We derived alcohol units per day variable in UK Biobank as previously suggested.[181] In summary, we considered 1.5 units for a glass of wine, 2.8 units for a pint of beer or cider and 1.5 units for other alcoholic drinks. We calculated one

unit per week for individuals reported drinking alcohol at least once a week and one unit per month for individuals reporting less frequent drinking. We further adjusted for (iii) MRI determined liver fat and (iv) liver iron to rule out the confounding effects of these two traits in our image processing pipeline. Finally, we carried out GWASs in (v) males and (vi) females separately to detect sex-specific associations.

4.6.6 Association of cT1 variants with liver biomarkers and metabolic traits and diseases.

To further understand the role of each cT1 variants in the pathophysiology of liver disease, and as a positive control, we tested the association between each variant and liver biomarkers in UK Biobank white European participants. The liver biomarkers include liver enzymes (ALT, AST, GGT, ALP in up to 378,821 individuals), MRI derived liver PDFF (n = 14,440), and MRI derived liver iron (to understand if the correction of T1 measures for liver iron content has caused any bias; n = 14,440). The protocols for the derivation of MRI PDFF and liver iron have previously been published.[176,177] To validate the associations with transaminases in a non-UK Biobank dataset, we looked up the effects of cT1 variants in an existing GWAS of ALT and AST levels in up to 61,089 individuals.[182]

To understand the effect of cT1 variants on cardiometabolic traits and diseases, we tested their associations with 15 predominantly metabolic traits including BMI, HDL-cholesterol (HDL), LDL-cholesterol (LDL), triglycerides, systolic blood pressure, diastolic blood pressure, type 2 diabetes, and coronary artery disease in up to n = 379,308 white European UK Biobank participants.

4.6.7 LD Score regression and cross-trait genetic correlation analysis

We used LD Hub to conduct linkage disequilibrium (LD) score regression and heritability analysis. LD Hub is a centralized database of summary level GWAS for > 500 diseases and traits from publicly available resources/consortia and uses a web interface that automates LD score regression, heritability and cross-trait genetic correlation analysis pipeline.[139] We ran heritability analysis as well as genetic correlation analysis across 120 potentially relevant traits. SNP-based heritability (h^2_{SNP}) is the proportion of total variation in liver cT1 measures due to the additive genetic variation between individuals in our study population.

4.6.8 Liver cirrhosis variants

To investigate the effect of liver cirrhosis variants on cT1 measures, and also as a positive control, we used variants associated with all-cause cirrhosis including rs2642438 (in or near *MARC1*), rs72613567 (*HSD17B13*), rs58542926 (*TM6SF2*), rs738409 (*PNPLA3*), rs1800562 (*HFE*), and rs28929474 (*SERPINA*).[27]

4.6.9 Mendelian randomisation

We investigated the potential causal associations between 24 predominantly metabolic traits on cT1 using two-sample MR analysis.[78] We used the inverse variance weighted approach (IVW) as our main analysis, and MR-Egger and penalised weighted median as sensitivity analyses in order to detect unidentified pleiotropy of our genetic instruments. Genetic instruments were constructed by using the independent genome-wide significant genetic variants ($R^2 < 0.1$) of the exposure of interest from previous GWASs. For more information on MR and genetic instrument selection please see Supplementary Material 7.2.1.2. Analysis was performed using MR Base, a web tool and R package.[81]

4.7 Results

4.7.1 The characteristics of liver cT1 cohort.

In our discovery cohort, median age was 57 years (interquartile range (IQR) 50 - 62) for males and 55 years (IQR 48 - 60) for females. The median liver cT1 was 694 milliseconds (ms; IQR: 662 - 730) in males and 676ms (IQR: 647 - 710) in females (Figure 7.11.). 5.3% of males (299 / 5,595) and 2.6% of females (169 / 6,455) had values above 800ms, a threshold that has been set in current clinical trials as a cut-off for steatohepatitis,[183] and is under evaluation by the FDA and European Medicines Agency as a diagnostic enrichment biomarker for non-alcoholic steatohepatitis. Baseline characteristics were comparable to the rest of the UK Biobank cohort who did not participate in the imaging study except BMI, waist circumference and diabetes prevalence which were lower in both males and females in the liver cT1 cohort compared to the rest of UK Biobank (Table 4.7.1.). Although invitation was not based on any medical information, MRI exclusion criteria (e.g. metal or electrical implants, surgery in six weeks prior to appointment, severe hearing or breathing problems) as well as the imaging site location (Cheadle, UK) may have contributed to a slightly healthier cohort.[177]

Table 4.7.1. Characteristics of UK Biobank participants in the imaging subset and the subset of participants who were not part of the imaging study.

	UK Biobank imaging subset		**UK Biobank non-imaging subset**	
Characteristics	**Men**	**Women**	**Men**	**Women**
No (%)	7,142	8,396	229,134	273,402
Age				
(IQR) (years)	57 (50;62)	55 (48;60)	58 (50;64)	57 (50;63)

Waist Circumference (IQR) (cm)*	94 (87;100)	79 (73;87)	96 (89;103)	83 (75;92)
Townsend deprivation index (IQR)	-2.78 (-3.98;0.82)	-2.66 (-3.90; -0.69)	-2.12 (-3.65;0.63)	-2.14 (-3.63;0.49)
Self-reported diabetes (%)*	245 (3.43%)	116 (1.38%)	15,950 (7.0%)	9,794 (3.6%)
Liver cT1 (IQR) (ms)	694 (662;730)	676 (647;710)	NA	NA
BMI (IQR) (kg/m2)*	26.6 (24.5;28.8)	25 (22.9;28)	27.3 (25;30.1)	26.1 (23.5;30)

*** BMI (Mann Whitney U test, $p = 1 \times 10^{-80}$), waist circumference (Mann Whitney U test, $p = 1 \times 10^{-100}$), diabetes prevalence (Pearson's chi squared test, p = 1 x 10-27) were lower in the imaging subset compared to the rest of UK Biobank. Levels of significance for all tests: ($p < 0.05$).**

4.7.2 Genetic variants in six loci show association with liver cT1.

In our GWAS of liver cT1 in individuals of European ancestry variants in six independent loci (Table 4.7.2.) reached genome wide significance. Genomic inflation was low ($\lambda_{GC} = 1.006$, Figure 7.12.). We observed the strongest association with a missense variant, rs13107325, located in an exon of *SLC39A8* (Figure 4.1.). The minor allele (T; allele frequency 7%) of rs13107325 was associated with 0.54 standard deviation (SD) increase in cT1 ($p = 1.2 \times 10^{-133}$). The mean cT1 was 692ms in individuals with no risk allele, 727ms in heterozygotes, and 772ms in risk allele homozygotes (Figure 7.13).

Table 4.7.2. The association between six independent genetic variants and liver cT1. A linear mixed model was used for genetic associations (levels of significance: $p < 5x10^{-8}$).

SNP	CHR	BP	EA	OA	EAF	Gene	Amino acid change	BETA	SE	P-value	Variance explained
rs759359281	1	220100497	C	CA	0.06	*SLC30A10*		0.137	0.026	2.8×10^{-8}	0.23
rs13107325	4	103188709	T	C	0.07	*SLC39A8*	A391T	0.544	0.022	1.2×10^{-133}	3.95
rs111723834	14	24572932	A	G	0.02	*PCK2*	R561Q	0.291	0.046	3.0×10^{-11}	0.27
rs58542926	19	19379549	T	C	0.07	*TM6SF2*	E167K	0.124	0.022	1.4×10^{-8}	0.22
rs4820268*	22	37469591	G	A	0.46	*TMPRSS6*		0.066	0.012	1.6×10^{-9}	0.20
rs738409	22	44324727	G	C	0.21	*PNPLA3*	I148M	0.095	0.014	9.6×10^{-13}	0.29

The coordinates and SNP IDs are in build 37. Beta represents effects in standard deviation (SD). SE = Standard error; CHR: = Chromosome ; BP = Base pair ; EA = Effect allele ; OA = Other Allele ; EAF = Effect allele frequency. * rs4820268 is in LD (r2=0.77) with rs855791 (V736A).

Other independent variants included an intronic variant (rs759359281-CA > C) in *SLC30A10* ($p = 2.8 \times 10^{-8}$), a missense variant (rs111723834-G > A) in *PCK2* ($p = 3.0 \times 10^{-11}$), a missense variant (rs4820268-A > G) in *TMPRSS6* ($p = 1.6 \times 10^{-9}$), and two known cirrhosis variants (rs58542926-A > G) in *TM6SF2* ($p = 1.4 \times 10^{-8}$) and (rs738409-C > G) in *PNPLA3* ($p = 9.6 \times 10^{-13}$). The six variants together explained 5.38% of variation in cT1 measures in white European UK Biobank participants with *SLC39A8* variant explaining most of this variation (3.95%) (Table 4.7.2.). We estimated the SNP-based heritability (h^2_{SNP}) of liver cT1 to be 20%.

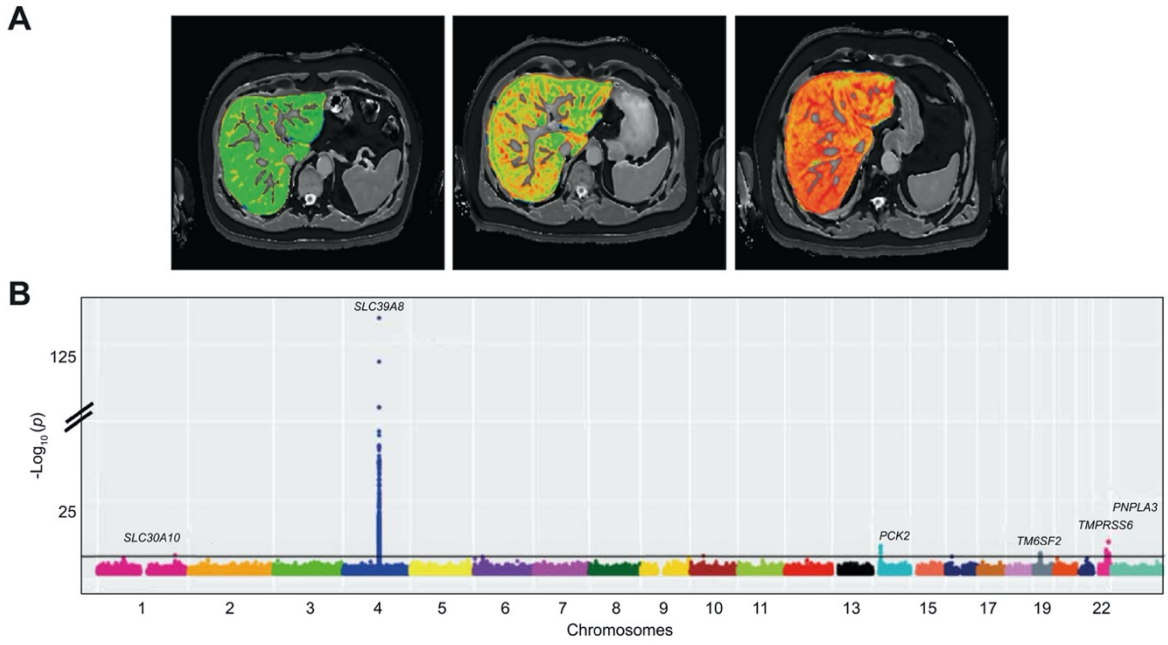
This is higher than the heritability estimated for conditions and traits such as coronary artery disease (7%),[149] eczema (7%),[150] body fat % (10%)[151] and transferrin (16%), but lower than non-alcoholic fatty liver disease (NAFLD) (22-34%).[21]

We did not detect any sex-specific associations and the effects were similar between men and women (Table 7.2.1.). Sensitivity analyses that further controlled for alcohol units intake and BMI did not identify any additional signals and did not significantly change the effect size (Table 7.2.1.). Sensitivity analyses that controlled for liver PDFF removed the effects of rs58542926 in *TM6SF2* and rs738409 in *PNPLA3*, suggesting that the effects of these variants on cT1 measures are mediated through liver fat accumulation (Table 7.2.1.). The cT1 increasing allele (G) at *TMPRSS6*-rs4820268 is associated with lower plasma iron levels and lower liver iron.[177] The effect of this variant on cT1 may be due to its effect on liver iron concentration since iron has an opposing effect to T1 relaxation time. However, sensitivity analyses that controlled for liver iron only slightly attenuated its effect on cT1 (from beta = 0.066, $p = 2 \times 10^{-9}$ to beta = 0.054, $p = 7 \times 10^{-7}$) suggesting that other mechanisms are involved and that this is a true signal.

4.7.3 Genetic variants in four loci show association with liver MRI determined PDFF.

In our GWAS of liver PDFF in 14,440 individuals of European ancestry missense variants in four independent loci reached genome wide significance (rs1260326-C > T in *GCKR*, $p = 3.9 \times 10^{-8}$, rs58542926-C > T in *TM6SF2*, $p = 6.3 \times 10^{-37}$, rs429358-C > T in *APOE*, $p = 5.6 \times 10^{-11}$, rs738409-C > G in *PNPLA3*, $p = 5.4 \times 10^{-66}$ (Table 7.2.2., Figure 7.14.). Genomic inflation was low ($\lambda_{GC} = 1.04$). Two of the four variants (rs738409 in *PNPLA3,* rs58542926 in *TM6SF2*) were shared between PDFF and cT1 in our GWASs.

Figure 4.1. GWAS of Liver cT1 in UK Biobank. 1A. Liver MRI scans of cT1. Three selected cases of liver MRI scans showing, from left to right, progressively elevated cT1 values (671ms, 777ms, 917ms). 1B. Manhattan plot illustrating GWAS of liver cT1 measurements in 14,440 UK Biobank individuals (~12 million imputed variants). The x-axis is the chromosomal position and y-axis is the significance of association for each variant in log10(p-values). Grey line indicates genome-wide significance level. For the GWAS, a linear mixed model was used. Levels of significance: $p < 5x10^{-8}$.

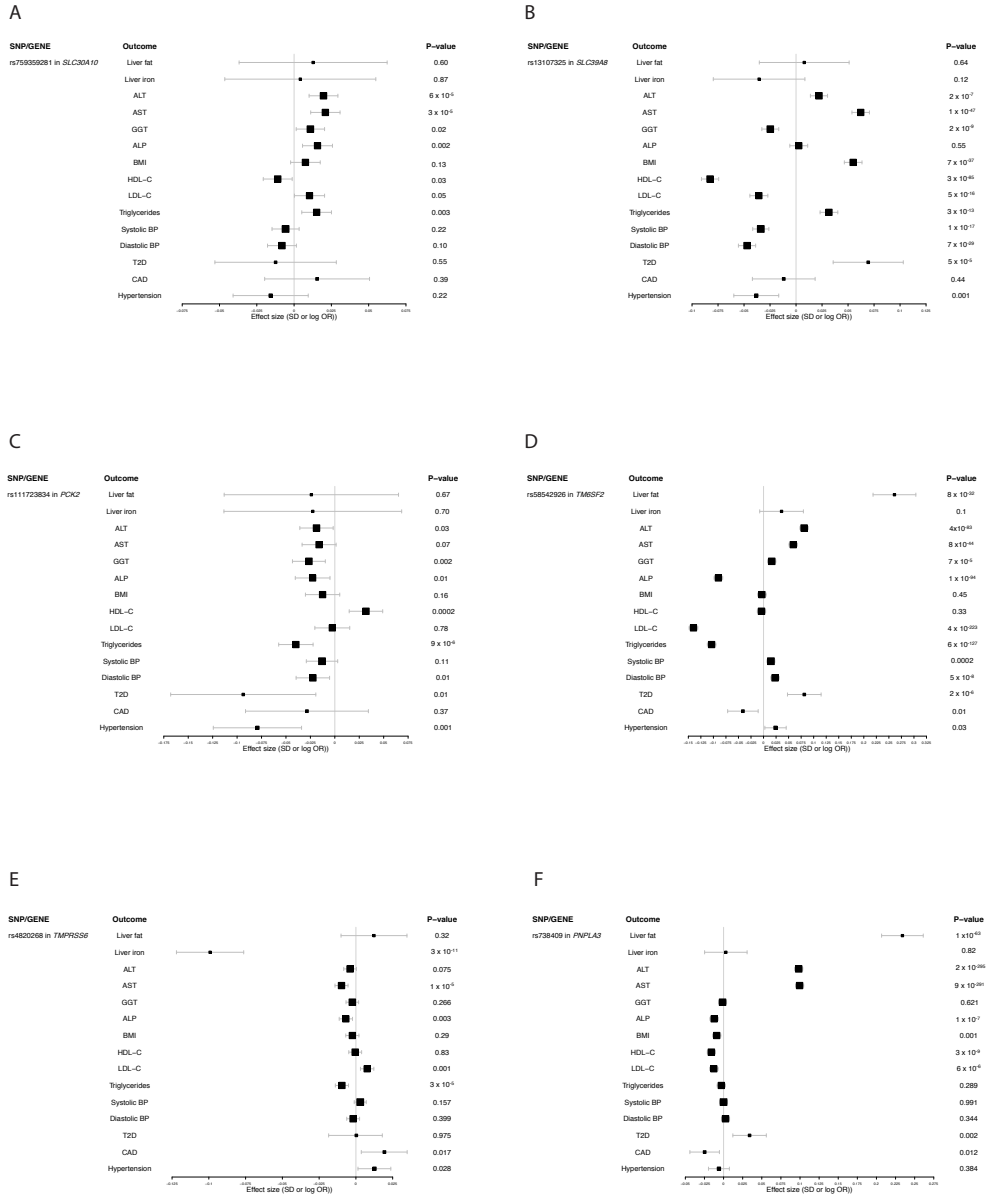


4.7.4 Four of the cT1 variants are associated with higher levels of aminotransferases and demonstrate variable effects on metabolic traits and diseases.

To validate these variants and further understand their role in other metabolic traits and diseases, we investigated their association with liver blood tests, MRI determined liver iron and liver PDFF, lipids, blood pressure, BMI and cardiometabolic disease outcomes **(**Figure 4.2., Table 7.2.3.). cT1-increasing alleles at four variants (in *SLC30A10, SLC39A8*, *TM6SF2,* and *PNPLA3*) were associated with higher ALT and AST (all with p-values < 2 x 10^{-5}) and higher risk of type 2 diabetes (all with p < 0.002, except the *SLC30A10* variant). None of cT1 variants were associated with cardiovascular disease risk, whilst their effects on other metabolic traits including lipids and blood pressure were variable (Figure 4.2.). Among the novel identified and replicated variants (rs759359281 in *SLC30A10*, and rs13107325 in *SLC39A8*), only the latter was available in a non-UK Biobank cohort with available liver blood tests. The cT1-increasing allele in rs13107325 showed similar direction of effect on ALT (n = 46,316, beta = 0.01, p = 0.27) and AST (n = 39,015, beta = 0.014, p = 0.0005) levels in an independent cohort (Table 7.2.4.).[182]

Figure 4.2. Forest plot of the associations of liver cT1 variants with liver and metabolic phenotypes. Effects are in standard deviations (SD) for continuous traits and log(OR) for disease outcomes per copy of the risk allele. ALT = Alanine transferase, AST = Aspartate transferase, GGT = gamma-glutamyl transferase, ALP = alkaline phosphatase, LDL_C = LDL cholesterol,

HDL_C = HDL cholesterol, T2DM = Type 2 Diabetes, CAD = coronary artery disease. A linear mixed model was used for genetic associations. Levels of significance: p < 0.05.



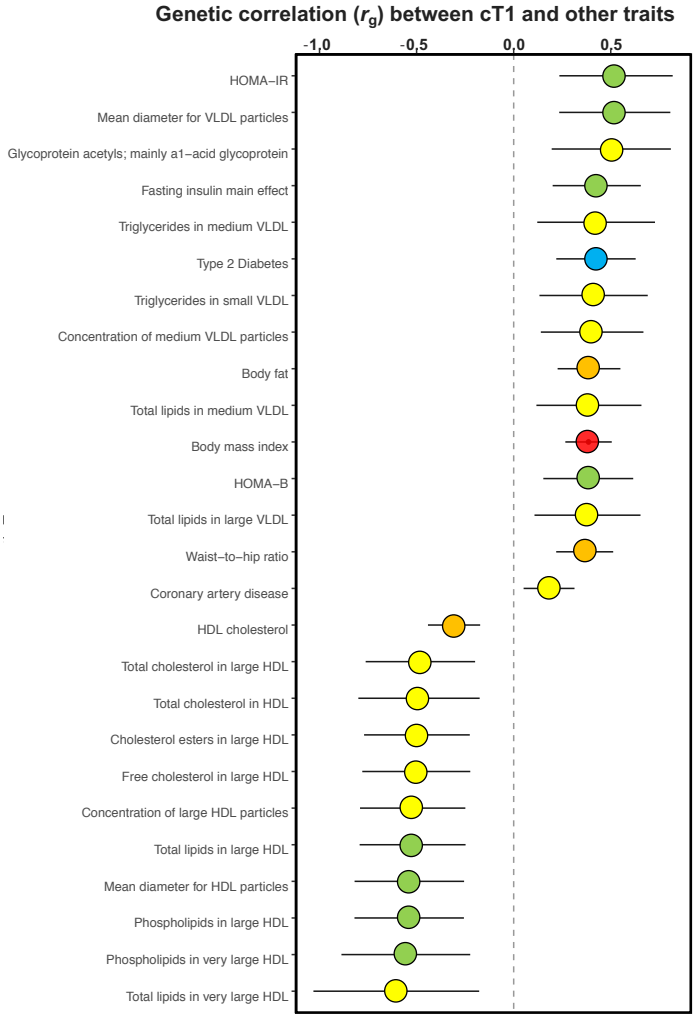
4.7.5 Liver cT1 measures correlate genetically with components of metabolic syndrome.

We calculated genetic correlations using the GWAS summary statistics (120 predominantly metabolic traits/diseases) in LD score regression analysis (Figure 4.3., Table 7.2.5.). Measures

of insulin resistance, triglycerides, VLDL, type 2 diabetes, coronary artery disease, body fat percentage, BMI and waist-to-hip ratio were genetically positively correlated with liver cT1 measures after correcting p-values for multiple testing (false discovery rate (FDR) < 0.05). The most genetically correlated traits were homeostatic model for insulin resistance (HOMA IR, r_G= 0.53, P = 0.0004) and mean diameter of VLDL particles (r_G=0.52, P = 0.0004), whereas the strongest inverse correlation was seen with total cholesterol in very large HDL (r_G = -0.62, P = 0.04).

Figure 4.3. Figure demonstrating the significant genetic correlations (rg) between cT1 and metabolic traits following correction for multiple testing (levels of significance: p false discovery rate < 0.05) among more than 120 traits. The colours correspond to significance of correlation (t-test); red: $p < 1x10^{-8}$; orange: $1x10^{-6} < p < 1x10^{-5}$; blue: $1x10^{-5} < p < 1x10^{-4}$; green: $1x10^{-4} < p < 1x10^{-3}$; yellow: $0.001 < p < 0.01$. Higher cT1 is genetically positively correlated with VLDL, type 2 diabetes, coronary artery disease, and inversely correlated with HDL. HOMA-IR = Homeostatic

model assessment insulin resistance, HOMA-B = Homeostatic model assessment β cell function, VLDL = very large density lipoprotein, HDL = High density lipoprotein.



4.7.6 Association of liver cirrhosis variants with liver cT1

We investigated the effects of all-cause cirrhosis risk variants on cT1 values. Among six variants associated with all-cause cirrhosis in a recent GWAS of 5,770 cases and 572,850 controls [27], four variants (those in or near *MARC1*, *HSD17B13*, *TM6SF2* and *PNPLA3*) demonstrated associations with cT1 (Table 4.7.3) where alleles associated with higher risk of liver cirrhosis were also associated with higher cT1. The *HFE* haemochromatosis risk allele

(in rs1800562) was inversely associated with cT1; however this is to be expected since cT1 measures are corrected for liver iron content. Consistently, this association became remarkably attenuated (from beta = -0.11, p = 8 x 10^{-7} to beta = -0.055, p = 0.02) in our sensitivity analysis correcting for liver iron content. In the GWAS of all-cause cirrhosis, the effect of a1-antitrypsin risk variant (rs28929474 in *SERPINA1*) was very weak (p = 0.01) and present only when a recessive model was carried out (Table 7.2.1.).[27] We did not have any risk allele homozygotes in our liver cT1 cohort and therefore could not perform a recessive model of associations with cT1.

Table 4.7.3. Effects of all-cause cirrhosis risk alleles on liver cT1.* indicates recessive models were run for the previously published all cause cirrhosis GWAS; all other association analyses used additive models. Logistic regression was used for the genetic associations with cirrhosis; a linear mixed model was used for the genetic associations with cT1 (levels of significance: p < $5x10^{-8}$, suggestive p < 0.05).

SNP	CHR	EA	OA	EAF	BETA Cirrhosis	P Cirrhosis	BETA cT1	SE cT1	P cT1	Gene
rs2642438	1	G	A	0.297	0.12	8.7 x 10^{-7}	0.036	0.0127	0.0049	*MARC1*
rs72613567	4	T	TA	0.722	0.16	4.5 x 10^{-8}	0.030	0.0129	0.02	*HSD17B13*
rs58542926	19	T	C	0.927	0.35	9.7 x 10^{-24}	0.124	0.0221	1.4 x 10^{-8}	*TM6SF2*
rs738409	22	G	C	0.211	0.38	2.2 x 10^{-67}	0.095	0.0141	9.6 x 10^{-13}	*PNPLA3*
rs1800562*	6	A	G	0.925	1.16	1.3 x 10^{-14}	-0.111	0.0223	8 x 10^{-7}	*HFE*
rs28929474*	14	T	C	0.0186	0.29	0.01	-0.037	0.0430	0.47	*SERPINA1*

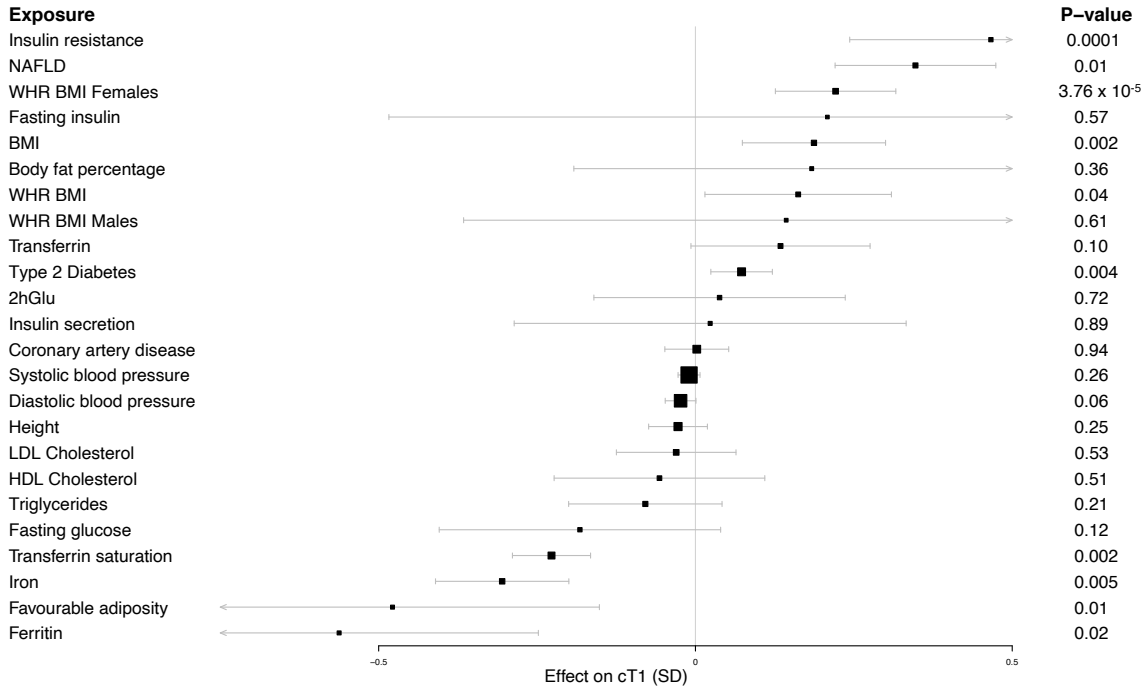
CHR = Chromosome; EA = Effect allele; OA = Other Allele; EAF = Effect allele frequency; SE = Standard error; * = recessive models used for all cause cirrhosis analysis. Beta Cirrhosis is the effect on all-cause cirrhosis in log(Odds Ratio) and Beta cT1 is the effect on cT1 in standard deviation (SD).

4.7.7 Mendelian randomisation analysis provides genetic evidence that non-alcoholic fatty liver, insulin resistance and obesity causally elevate liver cT1.

Demonstrating causality using observational studies can be challenging due to the presence of confounders such as other features of metabolic syndrome and behaviours including smoking and alcohol intake.[184] In UK Biobank, we detected a strong correlation between cT1 and BMI (r^2: 0.36, p = 5 x 10^{-324}) and also between cT1 and MRI determined liver fat PDFF (r^2: 0.62, p = 5 x 10^{-324}), and a weak but significant inverse correlation with liver iron (r^2: -0.069, p = 6.6 x 10^{-18}), which is to be expected since cT1 measures were corrected for liver iron (Figure 7.15.). We used genetic methods (Mendelian randomisation, Figure 4.4.) that are generally free of biases such as confounding and reverse causation to examine the potential causal effect of metabolic traits on liver cT1. We found evidence of a causal association between insulin resistance (IVW p = 0.0001), non-alcoholic fatty liver (IVW *P* = 0.01), type 2 diabetes (IVW p = 0.004), BMI (IVW p = 0.002) and higher cT1. We also found evidence for a protective role of favourable adiposity variants (variants associated with higher adiposity but lower risk of cardiometabolic diseases and lower ectopic fat)[185] and cT1 (IVW p = 0.01) (Table 7.2.6.). Our analyses were robust across a range of sensitivity analyses (Table 7.2.6.).

Figure 4.4. Mendelian randomisation investigating the effect of 24 predominantly metabolic traits on liver cT1. We used two sample Mendelian randomisation analysis to investigate the causal effects of metabolic traits on liver cT1. For full results, including sensitivity analyses, please see Supplementary Table 4. NAFLD = Non-alcoholic fatty liver disease, 2hrGlu = 2 hour glucose

tolerance test, WHR_BMI = Waist hip ratio adjusted for BMI. The inverse variance weighted test (IVW) was used as the main analysis. Levels of significance: p < 0.05.



4.8 Discussion

We identified associations between six independent genetic variants and MRI-based liver cT1, a non-invasive marker of liver inflammation and fibrosis, in 14,440 participants in UK Biobank. These include 5 missense variants (in *SLC39A8*, *PCK2*, *TM6SF2*, *PNPLA3*, and *TMPRSS6*) and one intronic variant (in *SLC30A10*). The cT1-increasing alleles in four genes (*SLC39A8, SLC30A10, PNPLA3,* and *TM6SF2*) were also associated with higher AST (N=360,731) and higher ALT (N=361,940) in UK Biobank and also in an independent GWAS of liver enzymes (except for *SLC30A10* and *TM6SF2* where data was not available).[182] *SLC30A10* and *SLC39A8* encode metal ion transporters and *PNPLA3* and *TM6SF2* are known genes associated with fatty liver and cirrhosis.

cT1 is a continuous trait, and analysed as such in our GWAS in line with other continuous traits such as blood pressure, BMI and height.[186–188] In some earlier publications, cT1 was reported using the Liver Inflammation and Fibrosis (LIF) score (Supplementary Material). The LIF score is a tri-linear mapping of cT1 onto a continuous scale from 0 to 4 based on the association of cT1 with histological fibrosis[10]. LIF categories were defined as having no (LIF <1), mild (LIF 1–1.99), moderate (LIF 2–2.99), or severe (LIF 3–4) liver disease.[9] The LIF cut-off of 1.4 had a sensitivity of 91% and a specificity of 52% for the diagnosis of NASH versus steatosis (AUROC = 0.80), and corresponds to a cT1 value of 780ms; a slightly higher cutoff of 800ms is used in clinical trials[183] and is under evaluation by the FDA and European Medicines Agency as a diagnostic enrichment biomarker for non-alcoholic steatohepatitis;[10,189] The LIF score is no longer used since the medical and MRI physics community is more familiar with T1 for the assessment of inflammation and fibrosis across all specialties including cardiology and neurology.[12,174,190–194] In this GWAS study, the cT1 values reported are standardised across the MRI scanner model and field strength and show very high repeatability and reproducibility.[195]

The missense variant (rs13107325-C > T) in *SLC39A8* is predicted to be deleterious in both Polyphen-2 and SIFT, and is associated with lower expression of *SLC39A8* in human liver.[69] *SLC39A8* encodes ZIP8 which has important roles in inflammation and immunity, and is a negative regulator of the NF-kB pathway.[196] ZIP8 is a divalent cation importer capable of transporting zinc, manganese, iron, cadmium and selinate; the substitution of C for T allele impairs the cellular uptake of metals by this protein.[197] It is not known which metal is involved in liver pathogenicity but there is evidence that hepatic ZIP8 regulates manganese metabolism in the liver, a metal ion that is hepatotoxic at high levels.[198] Zinc and selenium

also have important roles in liver cellular injury, oxidative stress and dysregulated inflammation; dietary supplementation of both has shown benefit in animal models of liver disease.[199,200]

The pathogenic role of *SLC39A8* in liver inflammation and fibrosis is supported by studies in mice which provide mechanistic evidence for the critical role of ZIP8 in liver disease. Liu *et al.* [201] used two mouse models to study the function of *SLC39A8* in the liver. In the first model, they studied the chronic effect of *SLC39A8* knockdown. The *SLC39A8(neo/neo)* homozygous mice died before or immediately after birth. The *SLC39A8(+/neo)* heterozygous mice had moderate ZIP8 deficiency which led to disruption of normal hepatocellular architecture, necrosis, inflammation, fibrosis and development of liver tumours with histopathological features consistent with hepatocellular neoplasms.[201] In the second model, they studied liver specific *SLC39A8* knockdown by adenovirus delivered shRNA and demonstrated that liver damage in the chronic model is not due to some extrahepatic process. Liver specific ZIP8 downregulation for seven days resulted in substantial hepatomegaly, inflammation, proliferation, oxidative stress, liver injury and cell death.[201]

The intronic variant in *SLC30A10,* a gene which codes a predominantly manganese metal ion transporter, was also associated with elevated cT1 measures in our study, as well as elevated transaminases in UK Biobank. Manganese is an essential metal required for the adequate functioning of numerous enzymes, however it is toxic and induces cell death at elevated cellular levels.[202] Loss-of-function mutations in *SLC30A10* have previously been associated with cirrhosis, higher manganese levels in liver biopsy samples and neurotoxicity including parkinsonian like movement disorders.[202,203]

The association between cT1 increasing alleles at the two novel loci (*SLC39A8* and *SLC30A10)* and higher ALT and AST adds supportive evidence for their pathogenic role in the liver. The missense variant in *SLC39A8* has previously been shown to be associated with multiple traits including alcohol intake, BMI, schizophrenia, Crohn's disease, lower brain grey matter volume and microbiome diversity;[160,187,204,205] we show for the first time a further novel association with higher diabetes and triglyceride levels, whilst highlighting variable effects on cholesterol levels. The associations of both variants with cT1 were independent of BMI, alcohol intake, liver fat percentage and liver iron content in our sensitivity GWAS models.

We identified a further two missense variants that were associated with cT1 but not with elevated transaminases; therefore, further research is required to validate these findings and explore their potential role in liver inflammation and fibrosis. The cT1-increasing allele in rs111723834 (missense variant in *PCK2,* also an intronic variant in *NRL*) was associated with lower transaminases, lower risk of type 2 diabetes, and lower triglycerides. *PCK2* encodes a mitochondrial enzyme that catalyses the conversion of oxaloacetate to phosphoenolpyruvate and has a key role in hepatic gluconeogenesis. Mitochondrial phosphoenolpyruvate carboxykinase deficiency (M-PEPCKD) is a rare autosomal recessive disorder resulting from impaired gluconeogenesis, and clinical characteristics include hypotonia, hepatomegaly, failure to thrive, lactic acidosis and hypoglycaemia.[206] The missense variant in *PCK2* is also an intronic variant in *NRL,* and it is unclear which gene is associated with elevated cT1 measures. *NRL* however encodes for neural retinal leucine zipper transcription factor that is specifically expressed in neuronal retina cells, making it an unlikely causal gene candidate for liver cT1. The cT1-increasing allele (rs4820268-A > G) in *TMPRSS6* has previously been reported to be associated with lower plasma iron levels and lower liver iron content.[74,177] It

is also associated with a dysmetabolic profile including higher LDL cholesterol, higher cardiovascular disease risk and hypertension. Its effect on cT1 however remained significant even after correcting for liver iron content in sensitivity analyses, making it unlikely that the association was secondary to bias resulting from iron correction when calculating cT1. Previous Mendelian randomisation studies have shown that higher circulating iron may be cardioprotective,[167] possibly through reduced circulating LDL-cholesterol and lower blood pressure.[207] The same mechanisms may explain why the allele associated with lower circulating iron levels is associated with higher cT1.

Known NAFLD and cirrhosis risk alleles in *PNPLA3* and *TM6SF2* were also associated with both elevated cT1 and MRI derived PDFF in our cohort. These associations provide strong positive controls for our study and validate for the first time the association with MR determined liver PDFF. The risk alleles in these two genes were further associated with higher risk of type 2 diabetes, but with lower serum triglycerides, LDL cholesterol, and lower risk for cardiovascular disease, as previously described.[64,68] In our GWAS on PDFF, alongside *PNPLA3* and *TM6SF2*, we further identified variants in *GCKR* (another known NAFLD variant which we have replicated) and *APOE* (apolipoprotein E, a gene which encodes a major cholesterol carrier).[61,68] The *APOE* risk allele (T) for PDFF is associated with higher risk of diabetes, and lower risk of cardiovascular disease and LDL cholesterol in independent GWASs.[119] This data provide evidence that cT1 and PDFF phenotypes share some but not all aetiopathogenic mechanisms.

We demonstrated that four of five variants associated with all-cause liver cirrhosis (in *PNPLA3*, *TM6SF2, HSD17B13*, and *MARC1)*[27] were also associated with liver cT1 with the first two reaching genome-wide significance. The paradoxical inverse association between the liver

iron-increasing allele in *HFE* may be due to overcorrection since cT1 measures are corrected for liver iron content and were inversely correlated in our cohort. Adjustment for liver iron content in our sensitivity analysis remarkably attenuated the association with cT1. *SERPINA1* variant was associated with all-cause cirrhosis only in a recessive model.[27] We did not have any homozygotes in our liver cT1 cohort to detect a recessive model of association with cT1.

Identifying causal mechanisms to steatohepatitis is crucial since interventions targeting these modifiable exposures may prevent liver disease progression. Our Mendelian randomisation study investigated 24 possible metabolic traits that may cause steatohepatitis. We provide genetic evidence that insulin resistance, non-alcoholic fatty liver and higher BMI causally increase cT1. Recent genetic studies have further identified variants associated with higher BMI but lower risk of type 2 diabetes, hypertension and heart disease.[208] These "favourable adiposity" variants are also associated with higher subcutaneous-to-visceral adipose tissue ratio and may protect from disease through higher adipose storage capacity, by promoting lipid deposition in subcutaneous tissue rather than within the circulation and ectopic places. The inverse link between favourable adiposity and steatohepatitis provides supportive evidence for the protective effects of this phenotype on a variety of cardiometabolic diseases, underlying mechanisms of which can be further explored and point to future preventive and therapeutic strategies.

Our study had a few limitations. We did not have any independent cohort to replicate our findings. To overcome this limitation, we investigated associations between cT1 variants and ALT and AST levels both in UK Biobank and an independent GWAS of liver enzymes.[7] While MRI derived cT1 is clinically available and is used to assess the severity of steatohepatitis, this measure is still novel, and further research is needed to determine the

relative contributions of inflammation and fibrosis to cT1.[11] Whilst it would be useful to have histological reference data for cT1, pathologist-interpreted liver biopsies do not lend themselves to large studies of this nature because of risk to patients and inter-rater variance in assessment of histology. This may be improved with advances in digitally processed histology to address variance and centralised collection of pathology for large consortia like the European LITMUS study. While cT1 has demonstrated excellent repeatability[191,195] and good correlation with fibro-inflammation and clinical outcomes,[8,10] other histological phenomena such as simple steatosis and ballooning have been shown to contribute to an increased T1 signal.[8] Only two of the six cT1 variants were associated with liver steatosis which highlights the complementarity of cT1 and liver fat PDFF as biomarkers of liver status, and their potential to recognise different mechanisms predisposing to liver disease.

4.9 Conclusion

cT1 and PDFF phenotypes share some but not all aetiopathogenic mechanisms. We identified novel associations between an MRI derived measure of fibroinflammatory liver disease and variants in *SLC30A10* and *SLC39A8* that replicated with blood biomarkers of hepatocyte injury. These genes have a critical role of transporting heavy metal cofactors for a multitude of biological processes. Future studies may determine whether targeting *SLC30A10* and *SLC39A8* are possible therapeutic options to prevent liver disease in at risk individuals. Our Mendelian randomisation study provides genetic evidence that addressing weight gain and insulin resistance are useful strategies in the prevention of steatohepatitis.

4.10 Data availability

Full data including individual cT1 and PDFF measures have been returned to UK Biobank and made publicly available via application (amsportal.ukbiobank.ac.uk).

5 Suspected liver fibrosis and incidence of 5 Cardiovascular diseases; a CALIBER study

5.1 Scientific Abstract

5.1.1 Background & Aims

It is uncertain whether non-alcoholic fatty liver disease (NAFLD) independently confers any additional cardiovascular disease (CVD) risk, or if the observed link with CVDs is due to similar risk factors and confounding. There is a growing body of literature suggesting that higher CVD risk is associated with the severity of NAFLD. Elucidating whether liver disease in NAFLD contributes additional CVD risk is important, as it can help improve CVD risk scores; furthermore, it is plausible that treatment of liver disease may ameliorate risk of CVD, over and above treatment of NAFLD associated risk factors. The associations of suspected liver fibrosis with a wide range of incident cardiovascular diseases (CVDs) have not previously been compared.

We aimed to study associations between suspected liver fibrosis and 5 CVDs.

5.1.2 Methods

We used linked primary care (CPRD), hospital admission (HES), and death certificate (ONS) records from the CALIBER programme which links clinical data for people in England. Where possible, we calculated the non-invasive Fibrosis-4 (FIB-4) score, a proxy for liver fibrosis, from results recorded in primary care health records of individuals who were 30 years or older and free from CVDs at baseline. We excluded individuals with conditions that may affect LBTs (excess alcohol, steatogenic medication, hepatopancreatobiliary disease except NASH/NAFLD, haematological conditions, cancer, symptoms of acute illness). We used Cox models to estimate cirrhosis and HCC hazard ratios (HRs) as a way of validating FIB-4 in routinely collected blood tests in primary care. We used Cox models to estimate cause-specific

HRs for 5 CVDs (IS, AF, HF, PAD, MI) and 1 composite endpoint. We completed 3 separate analyses adjusting for a) sex only b) fully adjusted for cardiovascular risk factors and medication (sex, BMI, smoking status, alcohol consumption, diabetes, antihypertensive medication, antilipid medication) and c) fully adjusted including age (which is part of the FIB-4 score). All models were stratified by GP practice.

5.1.3 Results

Among 3,160,383 individuals over the age of 30, a cohort of 49,956 individuals with at least one calculated FIB-4 score and no cardiovascular disease history was created, after exclusion criteria were applied. Individuals with suspected liver fibrosis (FIB-4 ≥ 1.3) showed strong, positive associations with a new diagnosis of cirrhosis (Hazard ratio (HR) 9.2 [7.4 – 11.4], $p < 2 \times 10^{-16}$) and HCC (HR 13.1 [6.3 – 27.4], $p = 8.7 \times 10^{-12}$) when compared to individuals with a FIB-4 < 1.3, thereby validating the score as a predictor of liver fibrosis using real world data.

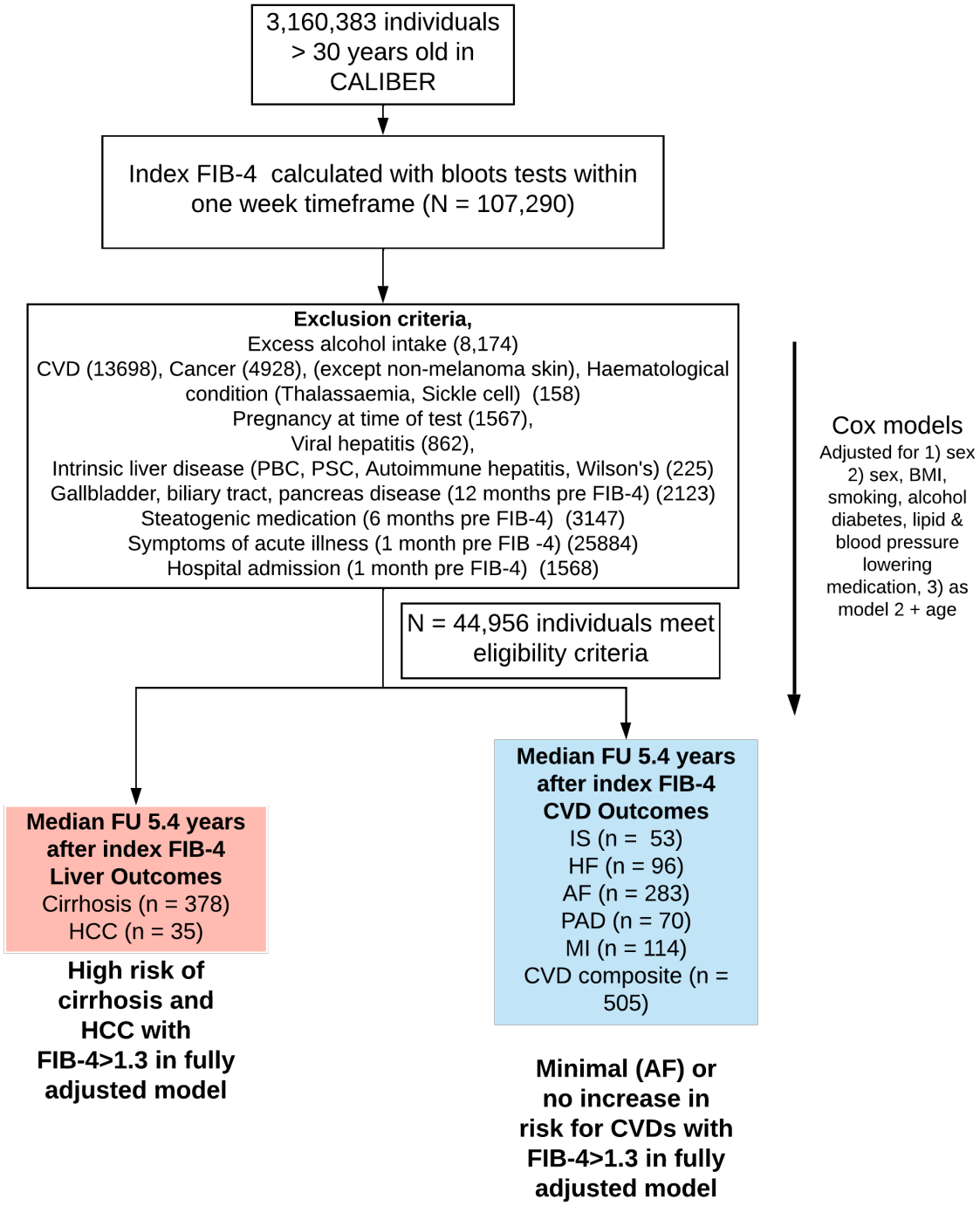
Over a median of 5.4 years follow up, 505 individuals developed at least 1 CVD. Adjusted HRs showed moderate positive associations of "suspected liver fibrosis" (FIB-4 ≥ 1.3) with most CVDs (HRs ranging from 1.72 to 2.05), however these associations were explained by age being part of the index. When adjusted for age, FIB-4 was not associated with higher CVD risk, except minimally with AF (IS HR 1.13 [0.81, 1.58], **AF HR 1.18 [1.01 – 1.37)]**, HF HR 1.05 [0.81, 1.34], PAD HR 0.97 [0.73, 1.28], MI HR 0.94 [0.75, 1.18], CVD composite HR 1.11 [0.99, 1.24]).

5.1.4 Conclusion

Using blood tests routinely collected in primary care, FIB-4 was strongly associated with incident cirrhosis and HCC across all models, validating its use as a marker of suspected liver

fibrosis. FIB-4 was associated with incident CVDs, however these associations were due to age being part of the index, except in the case of AF were a small increase in risk persisted. The assumption that liver fibrosis has an independent and possibly causal association with the occurrence of CVDs is not supported by the findings of this thesis.

5.2 Visual Abstract



5.3 Lay Abstract

The liver is responsible for multiple essential functions in the body. These include allowing blood to clot appropriately, maintaining a healthy immune system, and regulating cholesterol and triglyceride levels. Fatty liver is a common condition where excess fat builds up around the liver; with time this may lead to significant liver damage and scarring. There is growing evidence that the inflammation and scarring caused by liver fat accumulation may predispose individuals to conditions such as heart attacks and strokes. We aimed to investigate if this is the case by analysing the blood tests and electronic health records of thousands of individuals. We find no evidence that suspected liver fibrosis is a strong additional risk factor for most heart and blood vessel conditions, although it may possibly slightly increase the risk of an irregular heart rhythm known as atrial fibrillation; individuals with suspected liver fibrosis should be risk assessed for cardiovascular disease risk in a similar way to the rest of the population.

5.4 Highlights

- FIB-4, a validated non-invasive predictor of liver fibrosis, can be easily calculated via measurements and blood tests taken routinely in primary care across the UK, and linked with primary, secondary care and mortality outcomes on a national scale.
- FIB-4, as calculated in primary care is strongly associated with a higher risk of incident cirrhosis and hepatocellular carcinoma, validating its use as a marker of liver fibrosis
- Although FIB-4 index is associated with incident CVDs, this is mostly explained by age being part of the index. We find no evidence that suspected liver fibrosis, as estimated by the FIB-4 index, is an independent risk factor for CVDs including myocardial infarction, ischaemic stroke, heart failure and peripheral arterial disease.
- There is a suggestion that FIB-4 may be independently linked with atrial fibrillation risk, however the excess risk is modest and possibly due to residual confounding.

5.5 Introduction

NAFLD is a common condition, defined as the presence of hepatic steatosis on imaging (such as ultrasound) or histology, in the absence of excessive alcohol consumption, liver diseases such as viral hepatitis, autoimmune or metabolic liver disease, or medication associated with hepatic steatosis such as amiodarone, steroids or tamoxifen.[209] It is a common condition affecting 23-28% of the adult population.[209] In time, NAFLD progresses from simple steatosis to significant liver fibrosis in approximately 10-30% of individuals. NAFLD is emerging as the most common cause of chronic liver disease in the developed world and is associated with an increased risk of mortality (compared to individuals without NAFLD) with cardiovascular diseases (CVDs) being the most common cause - there is emerging evidence that NAFLD is an independent risk factor for CVDs.[3]

Whether non-alcoholic fatty liver disease (NAFLD) confers any additional cardiovascular disease (CVD) risk, or whether the increase in CVD risk in NAFLD is due to associated risk factors is uncertain. Elucidating whether liver disease in NAFLD contributes to additional CVD risk is important, as it is plausible that treatment of liver disease may ameliorate risk of CVD, over and above treatment associated risk factors (e.g. blood pressure, high cholesterol).

NAFLD has been inadequately studied in a population setting in relation to distinct CVD endpoints (CVDs), with studies primarily focusing on CVD mortality, MI and stroke. Studies focusing on the incidence of other CVDs, such as AF and HF, are limited, despite such conditions being associated with significant morbidity and mortality. Detection, primary and secondary prevention strategies as well as treatment for these conditions can differ from those for MI. Whether simple steatosis (NAFL) is associated with CVD risk is still unclear;[3] previous studies for example have suggested that only patients with the more severe,

inflammatory liver phenotype (NASH) and liver fibrosis have an increased cardiovascular mortality.[210] Regardless of causation, the association of NAFLD and CVDs might have implications for cardiovascular risk prediction; current prognostic scores for myocardial infarction and stroke such as the QRISK now include other non-cardiovascular co-morbidities such as chronic kidney disease and rheumatoid arthritis.[211]

Blood tests are the most common route to diagnosis of NAFLD and liver fibrosis. Such tests, as well as non-invasive scores that can reliably exclude the presence of fibrosis are available (or can be readily calculated) at scale, with no extra cost, in clinically collected EHR data. The Fibrosis -4 (FIB-4 score) clinical prediction score is easy to calculate (on the basis of age, AST/ALT levels and platelet count) and performs well in excluding significant liver fibrosis, with an area under the receiving operating characteristic curve (AUROC) of 0.76-0.80, and associated high negative predictive value for of 93% for liver fibrosis at the 1.3 cut-off.[212] EASL have emphasized the high predictive value and clinical relevance of non-invasive tests such as FIB-4 in excluding significant liver fibrosis. However, the predictive value of the FIB-4 score for the development CVDs, the number one cause of death in individuals with NAFLD has not been adequately investigated, particularly in a population setting.

5.6 Objectives

This study aims to investigate the association of the FIB-4 index, as a proxy for liver fibrosis, with initial presentation of 5 cardiovascular endpoints in a large population-based cohort from linked electronic health record database: the CALIBER program (Clinical Research Using Linked Bespoke Studies and Electronic Health Records).[120] CALIBER is a unique research

platform consisting of “research ready” variables extracted from linked electronic health records (EHR) from primary care, coded hospital records, and cause specific mortality data in England. CALIBER has been extensively validated, replicating known prospective associations of CVDs with alcohol, sex,[213] smoking,[214] blood pressure,[215] ethnicity,[123] socioeconomic deprivation,[216] type 2 diabetes,[217] and neutrophil counts.[218]

5.7 Methods

5.7.1 Study population

The study population was drawn from the CALIBER program, which links three sources of EHR in England: primary care health records (coded diagnosis, coded symptoms, clinical measurements, blood tests, and prescriptions from general practices contributing to the Clinical Practice Research Datalink (CPRD)), coded hospital diagnoses and procedures (hospital episode statistics) and death registrations (ONS). Individuals were eligible for inclusion if they were (or turned) 30 years or older between Jan 1 1998, and March 2016, and if they had been registered for at least one year in a practice that met research data recording standards. The research project was approved by CPRD in 2017 (Study number 17062 RAR). CPRD is jointly sponsored by the Medicines and Healthcare products Regulatory Agency (MHRA) and the National Institute for Healthcare Research (NIHR), as part of the Department of Health and Social Care.

5.7.2 Cohort creation

We extracted a cohort where a FIB-4 score could be calculated. The FIB-4 index was calculated as:

(age[years] x AST[U/L] / ((PLT[10^9/L]) * (√ALT[U/L])

We calculated FIB-4 counts if all blood tests required (platelet count, ALT, AST) were available within 7 days.

5.7.3 Exclusions

We excluded individuals with a history of CVDs or cancer (except non-melanoma skin cancer) prior to the index blood tests. We excluded women who were pregnant at the time of the test. We further excluded individuals with viral or intrinsic liver disease (hepatitis B, hepatitis C, HIV, autoimmune hepatitis, primary sclerosing cholangitis, primary biliary cirrhosis, haemochromatosis, Wilson's disease) or recent pancreatobiliary disease (e.g. gallstones, cholecystitis, pancreatitis) within 12 months of the exposure, as these may affect platelet and LBT levels).

We further excluded individuals with a history of excess alcohol use. We used the most recent record of alcohol consumption in the five years before entry into the study to classify participants' drinking behaviour. We defined five categories of drinking: non-drinkers (Read codes such as "teetotaller" and "non-drinker", former drinkers (those with codes for "stopped drinking alcohol" and/or "ex-drinker", occasional drinkers (those with codes for "drinks rarely" and/or "drinks occasionally", current moderate drinkers (codes such as "alcohol intake within recommended sensible limits" and "light drinker") and heavy drinkers (codes including "alcohol intake above recommended sensible drinking limits" and "hazardous alcohol use");

excluded from the study); We also used data fields with information entered on daily and weekly amount of alcohol consumed to define participants as non-drinkers, moderate drinkers (drank within daily and/or weekly amount of alcohol (less than or equal to 14 units of alcohol per week)), and heavy drinkers (in excess of 14 units per week). We further excluded individuals with use of medication that may cause hepatic steatosis (ciclosporin, tacrolimus, tamoxifen, amiodarone, 6-mercaptopurine, 5-fluorouracil, azathioprine, methotrexate, oral contraceptive pill, glucocorticoids, sodium valproate) in the preceding 6 months before the blood test.

LBTs can also be affected by acute illness, medication, and hematologic conditions. We classified the patient state at the time of the blood test as acute or stable. An "acute clinical state" was defined as any of the following conditions: in hospital on the date of blood test; vaccination in the previous 7 days; anaemia diagnosis within the previous 30 days; symptoms or diagnosis of infection within the previous 30 days; prior diagnosis of myelodysplastic syndrome; haemoglobinopathy, cancer chemotherapy, or injection of granulocyte colony-stimulating factor within the previous 6 months; or the use of drugs affecting the immune system, such as methotrexate or steroids, within the previous 3 months. Patients were considered stable if they did not fulfil the criteria for an acute clinical state. Patients on dialysis, those with human immunodeficiency virus infection, or a history of splenectomy were also excluded from this study. These criteria were based on those proposed by the eMERGE (Electronic Medical Records and Genomics). We excluded "acutely unwell" individuals, who were hospitalised within a month of the calculated FIB-4 count or had symptoms in keeping with an infective illness such as fever, diarrhoea, vomiting, cough. The codes used to identify

the phenotypes above were designed by a team including myself, other medical and epidemiology trained staff, and can be accessed through the HDR UK Phenotype library.[219]

5.7.4 Covariates

For continuous variables (e.g. BMI), we used, as a baseline value, the most recent measurement recorded in CPRD in the 5 years before study entry. Data recorded before study entry were used to classify participants' smoking (never, ex, current) and alcohol status (teetotal, ex, light, moderate) at baseline. We used as covariates the use of antihypertensive and cholesterol medications in the preceding six months prior to study entry. Multiple imputation by chained equations was used for missing data. The codes used to identify the phenotypes above can be accessed through the HDR UK Phenotype library, with links provided in the supplementary material (Table 7.3.2.).[219]

5.7.5 Outcomes

To validate the calculated FIB-4 score from routinely collected primary health blood tests, we initially compared the risk of developing cirrhosis and HCC in individuals with high (≥ 1.3) and low (<1.3) FIB-4. The primary endpoint of the study was the first record of one of the following cardiovascular presentations in any of the data sources (CPRD/HES/ONS), similar to previous CALIBER studies:[123,215,217,218] IS, AF, HF, PAD, MI, and a cardiovascular disease (CVD) composite endpoint. The codes used to identify the phenotypes above can be accessed through the HDR UK Phenotype library.[219]

5.7.6 Statistical analysis

Study entry occurred on the date of the first calculated FIB-4. Follow up was censored at the occurrence of a primary endpoint, death, de-registration from the practice, or the last data collection for the practice, whichever occurred first. We used multivariable Cox regression to calculate cause-specific hazards for association between FIB-4 index (as a continuous variable, quintiles, but also using a binary cut-off at 1.3) and initial presentations of cirrhosis, HCC, and CVD endpoints. We extracted demographic variables, and cardiovascular risk factors around the time of diagnosis from CPRD. We further extracted phenotypes which would allow us to reliably apply our exclusion criteria (Figure 5.1.). Testing for proportional hazards was not performed as current opinion is that this should not affect interpretation; the majority of hazards in clinical medicine are not proportional over time.[220]

In the primary analysis, we compared individuals with first recorded FIB-4 index, adjusting for sex, BMI, smoking status (never, former, current), alcohol consumption (tee-total, ex, occasional, moderate), diabetes, lipid lowering therapy and blood pressure lowering therapy in the six months before study entry. In other models, we also corrected for b) age + fully adjusted model used in primary analysis (since age is part of FIB-4 we left it out of the fully adjusted primary analysis model) and c) sex only. Multiple imputation by chained equations was used for missing data was carried out using package MICE.[221]

Analysis was carried out using R (version 4.1.2) and STATA (version 17). For data wrangling, Hadley Wickhams tidyverse collection of packages (containing dplyr, ggplot2, tidyr, readr, purrr, tibble, stringr, forcats) were used.[222] For cox regression analysis, the package survival was used.[223]

5.8 Results

5.8.1 The characteristics of FIB-4 cohort

Among 3,160,383 individuals in the cohort over the age of 30 with at least one blood test, we were able to calculate a FIB-4 score in 107,290 (3.4%) individuals. Following application of our exclusion criteria, 49,956 individuals with an index FIB-4 were eligible for inclusion in the study. Individuals had a median follow up of 5.4 years. Individuals with a FIB-4 score > 1.3 were more likely to be older, male, have diabetes, and be on blood pressure and lipid lowering therapy. They were also more likely to have a coded diagnosis of NAFLD/NASH or cirrhosis (Table 5.8.1.).

Figure 5.1. Schematic of patients in study of FIB-4 and initial presentation of cardiovascular diseases.

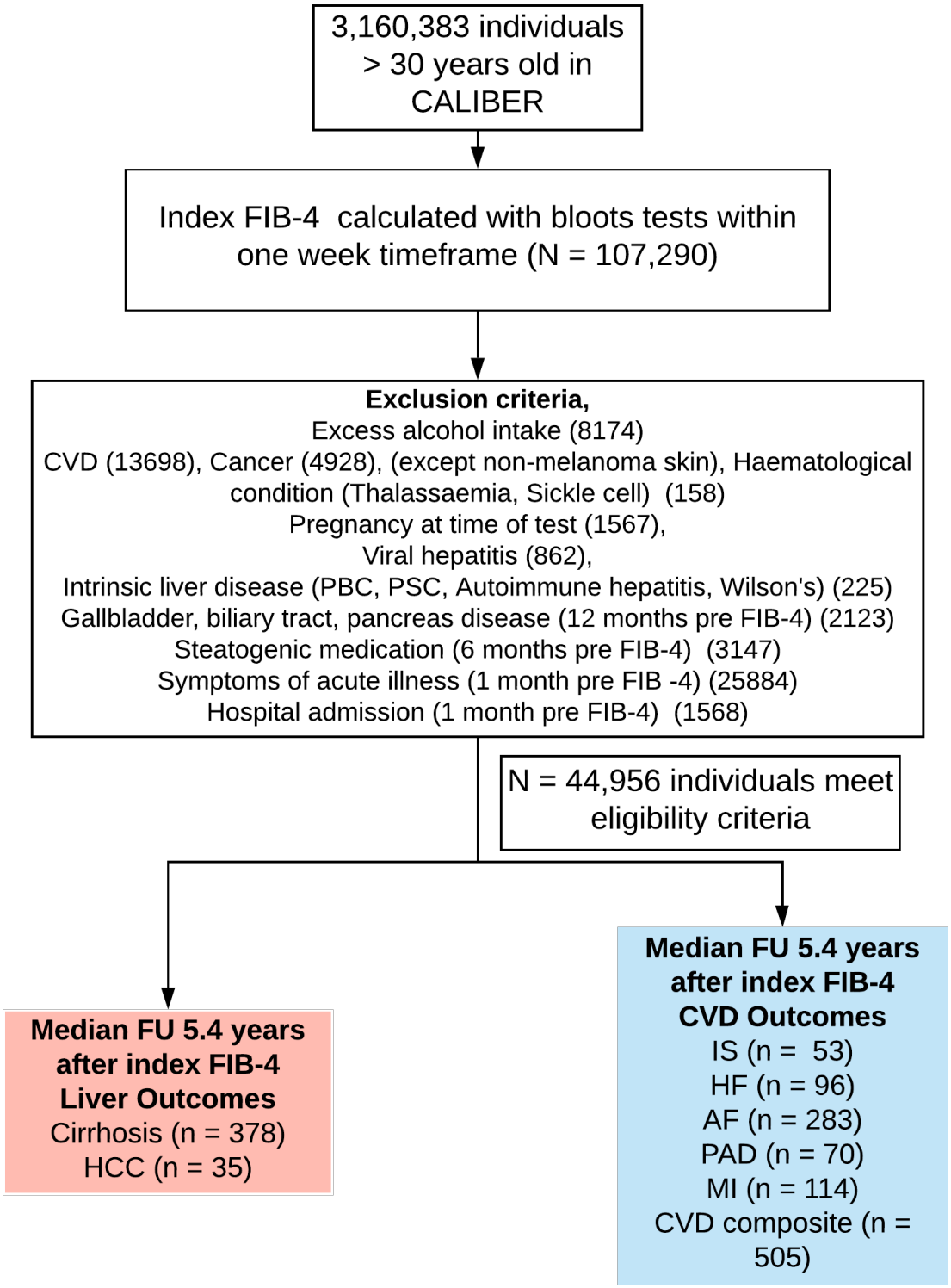


Table 5.8.1. Descriptive characteristics of the participants.

	Fib low (<1.3) N=42344	**Fib high (≥1.3) N=7612**
Age, median (IQR)	44.2 (37.5 – 51.9)	55.1 (48.4 – 60.3)
BMI, median (IQR)	26.3 (23.3 – 30.1)	25.8 (23.0 – 29.3)
Missing: N (%)	Missing: 6964 (16%)	Missing: 1231 (19%)
Sex, N (%)		
Men	18682 (56%)	4253 (56%)
Women	23662 (44%)	3359 (44%)
Smoking status, N (%)		
Never	24254 (57%)	4396 (58%)
Former	5861 (14%)	1172 (15%)
Current	11363 (27%)	1905 (25%)
Missing	886 (2%)	139 (2%)
Alcohol		
No	6305 (15%)	1006 (13%)
Ex	424 (1%)	84 (1%)
Occasional	5473 (13%)	895 (12%)
Moderate	1333 (3%)	295 (4%)
Missing	28809 (68%)	5332 (70%)
Diabetes, N (%)	967 (2%)	210 (3%)
Blood pressure lowering medication, N (%)	1709 (4%)	548 (7%)
Lipid lowering medication, N (%)	1059 (3%)	390 (5%)
NAFLD / NASH, N (%)	614 (1.5%)	161 (2.1%)
Cirrhosis, N (%)	14 (0.03%)	57 (0.75%)

5.8.2 FIB-4 is associated with higher risk of cirrhosis and HCC

During the study, 378 individuals developed cirrhosis, and 35 developed HCC. Individuals with "suspected liver fibrosis" (FIB-4 ≥ 1.3) showed positive hazards with a new diagnosis of cirrhosis (9.21 [7.46 – 11.36], $p < 2 \times 10^{-16}$) and hepatocellular carcinoma (13.09 [6.26 – 27.40], $p = 8.7 \times 10^{-12}$) when compared to individuals with a FIB-4 < 1.3, thereby validating the score as a predictor of liver disease using real world data. The positive associations persisted when adjusting the model for age in cirrhosis (13.64 [10.79 – 17.26], $p < 2 \times 10^{-16}$) and HCC (11.64 [5.15 – 26.31], $p = 3.5 \times 10^{-9}$).

5.8.3 FIB-4 is associated with higher risk of CVDs however this risk is explained by age being part of the score

Adjusted HRs comparing the "no suspected liver fibrosis" category (FIB-4 < 1.3) with "suspected liver fibrosis" (FIB-4 ≥ 1.3) showed positive associations with CVDs (IS 1.92 [1.4 – 2.65], AF 2.05 [1.78 – 2.37], HF 1.92 [1.51 – 2.44], PAD 1.72 [1.3 – 2.06], CVD composite 1.9 [1.71 – 2.11]). When adjusting for age however, a FIB-4 > 1.3 showed no associations with most CVD endpoints (IS 1.13 [0.81 – 1.58], HF 1.05 [0.81 – 1.34], PAD 0.97 [0.73 – 1.28], CVD composite 1.11 [0.99 – 1.24]). It showed weak associations with AF (1.18 [1.01 – 1.37]), A similar pattern of effects was seen when the FIB-4 score was examined as a continuous variable and in quintiles (**Error! Reference source not found.**., Figure 7.16.).

Figure 5.2. Sex adjusted, fully adjusted and fully adjusted including age models comparing hazard ratios of high (≥1.3) versus low (<1.3) Fib-4 score.

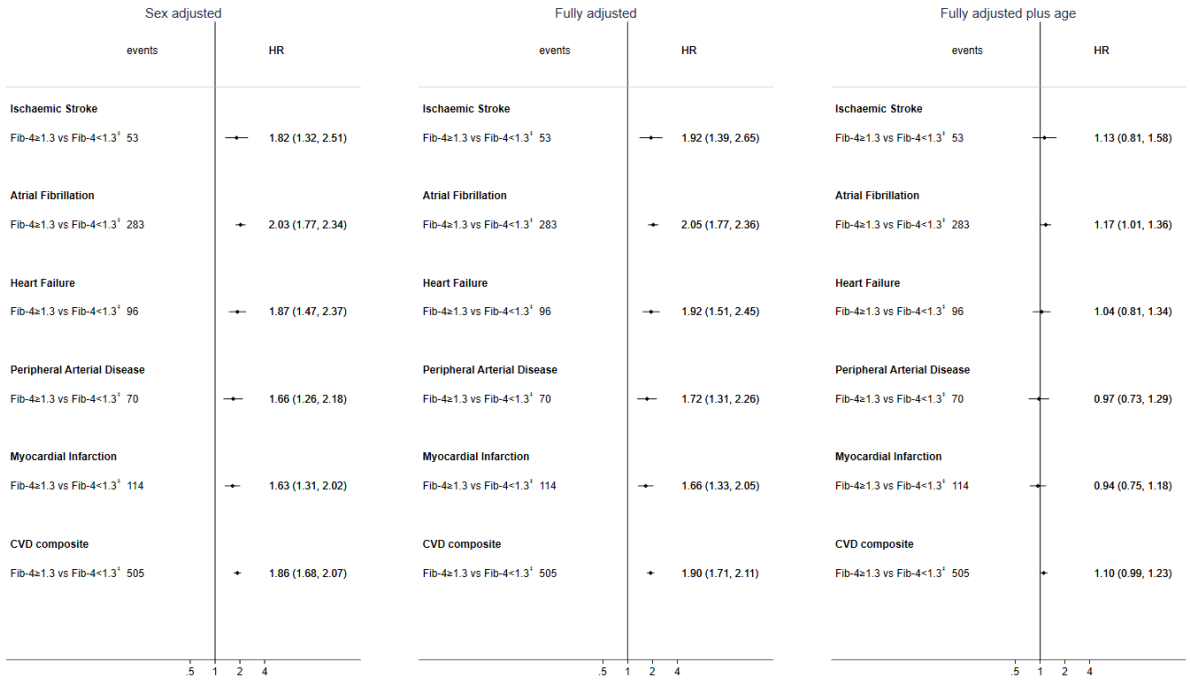
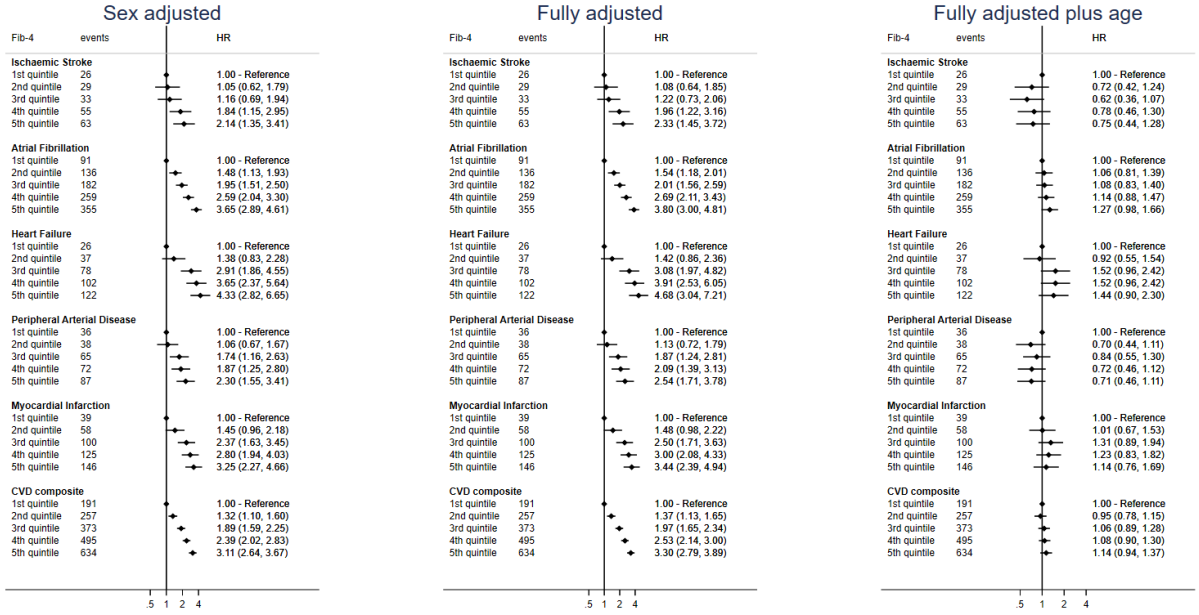


Figure 5.3. Hazard ratios for cardiovascular diseases per FIB-4 quintile (sex adjusted, fully adjusted, fully adjusted + age models). Quintile groups (FIB-4 < 0.66, 0.66-0.86, 0.86-1.12, > 1.12)



5.9 Discussion

Our results suggest that a FIB-4 index ≥ 1.3, when adjusting for age, is not associated with 4 incident CVDs when compared to a FIB-4 index < 1.3 and is only weakly associated with a higher risk of AF. Similar results are seen when the FIB-4 score was analyzed in quintiles as well as a continuous variable (**Error! Reference source not found.**., Figure 7.16.).

Recent, large, population based studies have also failed to show an increase in incident CVDs in patients with NAFLD / liver fibrosis, when diagnosis was based on EHR codes[99] but also biopsy samples.[100] We have added further evidence to this literature using a continuous variable, FIB-4, a score which is reliably associated with liver fibrosis.[224] This study also supports our Mendelian randomisation studies in Chapter 4 which failed to show a causal association between genetically determined cT1 and coronary artery disease (Figure 4.4.).

There is accumulating evidence that NAFLD is associated with cardiac arrhythmias. The mechanisms underpinning the link between NAFLD and arrhythmias are complex and not fully understood, and no studies have established a cause-effect relationship.[89] Hypotheses previously suggested in the literature include that the ongoing pro-inflammatory factors driving liver fibrosis may lead to the development and persistence of AF, and that NAFLD may be a marker of ectopic and lipotoxic fat accumulation in other tissues, including the myocardium and pericardium.[89,225] Some studies have also suggested that NAFLD is associated with cardiac calcification, in both aortic and mitral valves, conditions that are established risk factors for AF.[226–228] A recent meta-analysis of 7,012,960 individuals showed that NAFLD was independently associated with higher risks of AF (HR 1.12 [1.11-1.13]), with moderate heterogeneity.[229] Our findings of a positive association supports current literature, however given the small increase in hazard, further research is needed to ensure this result is not secondary to residual confounding.

Our study has significant strengths. Our large sample size allowed us for the first time to our knowledge to investigate the association of suspected fibrosis and multiple CVD outcomes that are associated with significant morbidity and mortality. The use of multiple, linked EHR datasets, allowed for more complete data and higher resolution, both in terms of covariates used and outcomes detected. We were further able to apply stringent exclusion criteria, and correct for multiple possible confounders, including demographic data, acute and chronic diseases, smoking status, alcohol status and hepatotoxic medications. To eliminate confounding from liver disease potentially driven by alcohol, we excluded adults with coded liver conditions such as alcoholic liver disease and those with a diagnosis of alcohol misuse,

harmful drinking habits (e.g. code for binge drinking) or a recorded average unit consumption above the government guidelines (Table 7.3.2.)

The absence of strong associations between liver fibrosis and cardiovascular disease risk might be explained if the abnormality causing increased liver fat is due predominantly to a reduction of very low-density lipoprotein (VLDL) secretion by the liver. This would lead to consequent accumulation of liver fat but a reduction in circulating triglycerides and LDL cholesterol. Genetic evidence in humans supports this hypothesis; a recent exome-wide association study of plasma lipids in > 300000 individuals demonstrated that the two most robustly associated NAFLD variants (I148M in *PNPLA3*, and E167K in *TM6SF2*), associated with higher liver fat, higher risk for type 2 diabetes, but with lower serum triglycerides, LDL cholesterol, and lower risk of CHD. Another explanation is the fact that ALT levels seem to decrease with increasing age and frailty; despite adjusting our models for age, residual confounding cannot be excluded.[230] It should not be assumed that people with a diagnosis of NAFLD or liver fibrosis are automatically at increased risk of CVDs; it is important to perform a cardiovascular risk assessment in people with a diagnosis of liver fibrosis, similar to the general population.

Another explanation is that associations between liver fibrosis and CVDs are mostly explained by confounding. These conditions both occur in the milieu of the metabolic syndrome, sharing multiple metabolic and behavioural risk factors such as insulin resistance, dyslipidaemia, central obesity, and high blood pressure.

Although our study has significant strengths – its large size, population base, extensive adjustment for potential confounders and stringent exclusion criteria – it also has important limitations. The measurement of LBTs was taken in usual clinical care, and by different

laboratories across the country without study wide protocols. This limitation was partially corrected by stratifying our analysis by GP practice. Heterogeneity of measured methods and heterogeneity among the study population itself may have led to biased estimates of association, but these would tend to be underestimates. Because our study was based on EHRs, some values of baseline variables were missing for some patients. We did not adjust for other confounders such as ethnicity and deprivation.

A major limitation of the current study was the use of FIB-4 to classify suspected fibrosis. These blood tests were taken in the "real world", risking significant ascertainment bias. However, should we have elected to use EHR codes to investigate NAFLD, NASH and liver fibrosis, our sample would have been significantly smaller, with a much-reduced follow up time, since most individuals with advanced fibrosis are captured in the more advanced stages of their disease.[5,231] Indeed, a recent multi-national study using disease codes to examine the effect of NAFLD on CVD risk was criticised methodologically for this reason, and the potential for misclassification bias (this study also failed to show an independent link between NAFLD, MI and stroke).[99] Furthermore, we would not have been able to study a continuous variable such as FIB-4, which enables greater power but also allows us to study the shape of any associations. In the current study, we used strict exclusion criteria to exclude excessive alcohol, viral hepatitis, hepatotoxic medication, hepatopancreatobiliary disease, acute illness and the presence of cancer, all of which can alter LBTs.

5.10 Conclusion

A validated, non-invasive marker of liver fibrosis (FIB-4) showed minimal associations with the incidence of specific CVDs when adjusting for age. There is no evidence that suspected liver fibrosis is associated with higher risk of CVDs, except in the case of AF were a small increase in risk persisted. Individuals with chronic liver disease should still be risk assessed for CVD risk since cardiovascular related mortality is still the most common cause of death in this cohort.

6 Conclusion

The new epidemic in chronic liver disease and cirrhosis is related to NAFLD. NAFLD is a continuum of liver abnormalities, from NAFL to the more advanced state, non-alcoholic steatohepatitis (NASH), which in turn can lead to cirrhosis and liver cancer. Metabolic disease also contributes to hepatic iron overload, DIOS, a condition that may expedite chronic liver disease progression. Cirrhosis appears to receive less public health attention that other chronic diseases, such as congestive heart failure, chronic obstructive pulmonary disease and chronic kidney disease; this is in part attributable to its stigmatisation and the perception that the disease is driven through behavioural factors such as alcohol consumption; around 2 million deaths worldwide are attributable to liver disease, per year.[5]

This thesis aimed to use large scale biomedical databases and population based linked EHR to further our understanding on some fundamental questions within chronic liver disease. These include 1) investigating the genetic susceptibility of phenotypes linked with chronic liver disease, such as MRI quantified liver iron and liver cT1, a proxy of steatohepatitis 2) investigating causality between metabolic traits and liver disease using novel genetic methodologies and 3) investigating whether suspected liver fibrosis is an independent risk factor for CVDs, the primary cause of death in this cohort.

We report a large GWAS of MRI derived liver iron content and identify 3 susceptibility loci previously implicated in circulating iron traits in an unselected population within UK Biobank. We provide genetic validation for multi-parametric MRI as a novel, non-invasive and radiation free imaging modality for liver iron content. Our genetic study suggests that higher liver iron content may be caused in part by higher central adiposity, suggesting that weight loss may help as a strategy to reduce liver iron in selected patients. The deposition of excess iron in the liver seems to share common mechanisms with circulating iron accumulation, which eventually

results in widespread damage to parenchymal tissue, leading to several pathologies through a common mechanism; such conditions may be proactively screened for during clinic consultations, and a multidisciplinary, multispecialty management approach may be better suited strategy for individuals with liver iron overload.

We further report a large GWAS for MRI derived liver cT1, a proxy for steatohepatitis / fibroinflammatory disease. We find that cT1 and PDFF phenotypes share some but not all aetiopathogenic mechanisms. We identified novel associations between MRI cT1 measures and variants in *SLC30A10* and *SLC39A8* that replicated with blood biomarkers of hepatocyte injury. These genes have a critical role of transporting heavy metal cofactors for a multitude of biological processes. Future studies may determine whether targeting *SLC30A10* and *SLC39A8* are possible therapeutic options to prevent liver disease in at risk individuals. A recent GWAS has further provided supportive evidence of a pathogenic role of metal cofactors in liver disease.[232] Our MR study provides genetic evidence that addressing weight gain and insulin resistance are useful strategies in the prevention of steatohepatitis. We further identified that variants associated with favourable adiposity are protective against steatohepatitis, similar to other metabolic phenotypes such as insulin resistance, high blood pressure and CVDs.[208]

Recent, large population based studies have failed to show a convincing signal of increased risk in incident CVDs in patients with NAFLD / liver fibrosis, when diagnosis was based on EHR codes[99] but also histopathology.[100] We have added further evidence to this literature using a continuous variable, FIB-4, a validated score which is reliably associated with liver fibrosis, and reliably associated with high risk of cirrhosis and HCC in our cohort.[224] Our population based study also supports our MR studies which failed to show a causal association between genetically determined steatohepatitis and cardiovascular disease risk. We report an

independent (albeit modest) association with incident AF, in line with a recent large meta-analysis.[229] Further studies are required however to ensure this finding is not due to residual confounding, however screening individuals in the clinic with chronic liver disease for AF, with simple tests (e.g. ECG) may be appropriate if this signal continues to remain across studies. Despite the absence of causality or prognostic value, all patients with chronic liver disease should undergo a thorough cardiovascular risk assessment since cardiovascular disease is still the primary cause of death in this cohort.

It is important to highlight once again the limitations of this thesis, starting with the liver iron study. The UK Biobank is not a completely unbiased sample of the population. The UK Biobank MRI cohort is slightly more healthy, wealthy, and well educated compared to the whole cohort of 40-69 year olds in the UK.[168] The population studied in this work has a slightly lower average BMI and waist circumference than the UK Biobank population as a whole, while the average Townsend Deprivation Index is slightly lower in this study cohort than the remainder of the UK Biobank. Larger GWAS (e.g. on completion of the full 100,000 UK Biobank imaging cohort, with genetic data moving to whole genome resolution compared to microarrays) may elucidate further susceptibility loci.

Our study on cT1 measures also has limitations. We did not have any independent cohort to replicate our findings. To overcome this limitation, we investigated associations between cT1 variants and ALT and AST levels both in UK Biobank and an independent GWAS of liver enzymes.[7] While MRI derived cT1 is clinically available and is used to assess the severity of steatohepatitis, this measure is still novel, and further research is needed to determine the relative contributions of inflammation and fibrosis to cT1.[11] Whilst it would be useful to have histological reference data for cT1, pathologist-interpreted liver biopsies do not lend

themselves to large studies of this nature because of risk to patients and inter-rater variance in assessment of histology, a contributor to the "placebo response".[49] This may be improved with advances in digitally processed histology to address variance and centralised collection of pathology for large consortia like the European LITMUS study.[233] While cT1 has demonstrated excellent repeatability[191,195] and good correlation with fibro-inflammation and clinical outcomes,[8,10] other histological phenomena such as simple steatosis and ballooning have been shown to contribute to an increased T1 signal.[8] Only two of the six cT1 variants were associated with liver steatosis which highlights the complementarity of cT1 and liver fat PDFF as biomarkers of liver status, and their potential to recognise different mechanisms predisposing to liver disease.

The major limitation of our CALIBER study was the use of FIB-4 to classify suspected fibrosis. These blood tests were taken in the "real world", risking significant ascertainment bias. However, should we have elected to use EHR codes to investigate NAFLD, NASH and liver fibrosis, our sample would have been significantly smaller, with a reduced follow up time since most individuals with fibrosis are captured in the more advanced stages of their disease. Furthermore, we would not have been able to study a continuous variable, which enables greater power but also allows us to study the shape of any associations. In the current study, we used strict exclusion criteria to exclude excessive alcohol, viral hepatitis, hepatotoxic medication, hepatopancreatobiliary disease, acute illness, and the presence of cancer, all of which can alter LBTs. We further did not correct for other important confounding such as ethnicity and deprivation index.

The COVID-19 pandemic is an example of how real-time analysis of health data can save millions of lives. A further challenge in our CALIBER study was the time that was required to

have access to the requested datasets; the journey from study conception and project application, phenotype derivation, ethical approval, raw dataset extraction and translation to "research ready" variables took approximately 18 months. Although some of these steps can now be expedited due to automated linkage and phenotype pipelines within the academic environment of Health Data Research UK,[234] to which this thesis has contributed to, there are several steps that potentially act as a "red tape" for EHR research. This challenge has been highlighted in an independent review[235] led by Professor Ben Goldacre which aimed to investigate how health data for research and analysis of NHS patient records by academics, NHS analysts, and innovators can be used safely and efficiently. Recommendations in the review include building a small number of "Trusted Research Environments" (also known as data safe havens), improving opportunities for data analysts within the NHS by modernizing their job and career development, encouraging open working for all NHS data analysis, using technology to allow staff to spend more time with patients, and giving people better access to their own data through shared care records and the NHS App. There is also a recommendation about having public conservations on issues such as "commercial use of NHS data for innovation" as well as "building trust by taking concrete action on privacy and transparency". The Goldacre review was the basis of the UK government's ambitious policy paper "Data saves lives: reshaping health and social care with data".[236] There is little doubt that optimising the workspace environment, ensuring timely access to "research ready" variables, enabling and empowering multidisciplinary research teams, together with ongoing public engagement and building trust through transparency is at the heart of efficient health data research.

Through this thesis, we were able to gain insight and identify novel variants and aetiopathogenic mechanisms for novel phenotypes that can be performed at scale. We further

used MR to investigate the causal effect of several metabolic traits on the phenotypes, results which can go on to influence public health policy. We finally add to existing literature highlighting the lack of population based evidence that liver fibrosis is an independent strong risk factor for CVDs.[99,100,103]

Understanding the underlying genetic susceptibility of liver imaging phenotypes, as well as its relationship with the clinical "phenome" may allow new insights on the main pathophysiological mechanisms contributing to chronic liver disease, help identify potential new drug targets and improve risk prediction across all diseases in this cohort of patients. Our thesis highlights the utility of large Biobanks and EHRs to further our understanding in this important area. Such studies are already increasing at dramatic pace in size, follow up time and resolution; future studies based on the concepts of this thesis are almost guaranteed to increase our understanding of chronic liver diseases: UK Biobank aims to have performed a whole-body MRI in 100,000 individuals over the next few years, and whole genome sequencing in all 500,000 participants by early 2023. The European NAFLD Registry is an international, prospectively recruited cohort study that aims to establish a large, highly phenotyped patient cohort and linked bioresource, able to support research and biomarker capacity at scale and pace. Biobanks have generally lacked appropriate representations of diverse ancestries, with most bioresources composed of individuals of European ancestries, a further limitation that is also apparent in this thesis. The Global Biobank Meta-analysis Initiative (GBMI) – a collaborative network of 19 biobanks from 4 continents representing more than 2.1 million consenting individuals with genetic data linked to EHR is another example of the progress made in this field.[237] GBMI brings together results from GWAS analyses across 6 main ancestry groups: approximately 33,000 of African ancestry either from Africa or from

admixed-ancestry diaspora, 18,000 admixed American, 31,000 Central and South Asian, 341,000 East Asian, 1.4 million European, and 1,600 Middle Eastern individuals. Genetic diversity in datasets has already shown its potential in increasing understanding in lipid research, and it makes intuitive sense that similar advances will be made in chronic liver disease research.[238]

This thesis has allowed me to gain essential attributes to further develop my academic career. I have been able to acquire important skills in data science, including the safe handling of large clinical and genetic data, electronic health record phenotyping, data wrangling, as well as bioinformatics, statistics, and data visualisation. The EHR phenotypes that I have developed are all accessible and can be used by the academic community through the HDR UK Phenotype Library, whereas the GWAS summary statistics have been returned to UK Biobank and are available via application.

I was further able to develop very important soft skills in leadership, teaching, collaboration, and management. I put together and coordinated research teams across a wide range of individual skills and expertise, working in academia, clinical healthcare and industry, with the main aim of producing leading research and impact. I regularly taught medical students, data scientists and epidemiologists from small group tutorials to lectures in fields such as clinical medicine, data science, phenotyping, evidence-based medicine, and prognostic score development.

The importance of chronic liver diseases and cirrhosis have been largely underestimated. For example, liver disease does not appear in the WHO list of non-communicable diseases that includes amongst others cardiovascular diseases, cerebrovascular diseases, diabetes, and

chronic respiratory diseases.[239] This lack of appropriate consideration is likely to be a contributory factor to the low awareness of liver diseases. Efforts should be made at the national and international level to place liver diseases at the level required to counteract stigmatisation of the disease and initiate campaigns to promote liver health.[5] A multi-disciplinary approach comprising of the appropriate leadership, research, awareness, treatment and care, policy strategies, patient and community perspectives and defining and implementing models of care is necessary to advance the NAFLD public health agenda. A consensus statement has recently been published with a series of recommendations on how this would be best achieved, and similar initiatives are required across other chronic liver disease such as ALD.[231] This thesis utilises a variety of alternative research methods highlighted in this consensus including electronic health record research and use of non-invasive tests (imaging, FIB-4) to enhance our understanding of aetiopathogenic mechanisms and risk in individuals with suspected liver diseases. Furthermore, efforts to standardize administrative codes used to record exposures and outcomes for NAFLD have been made, with phenotypes across the liver disease spectrum created and now publicly available. We hope our thesis can serve as an example on how large clinical and genetic datasets can be used to further our understanding of cause, consequence, and outcome risks of individuals with chronic liver disease.

7 Supplementary Material

7.1 Genetic studies of MRI liver iron content identify susceptibility loci and yield insights into its link with other diseases.

Collaborators/ Investigators: We would like to thank the following collaborators who work with the IMI DIRECT consortium. **Please include as collaborators on Pubmed**: Christopher Jennison, Beate Ehrhardt, Patrick Baum, Coria Schoelsch, Jan Freijer, Rolf Grempler, Ulrike Graefe-Mody, Anita Henige, Christiane Dings, Thorsten Lehr, Nina Sihinecich, Francois Pattou, Violetta Raverdi, Robert Caiazzo, Fanelly Torres, Helene Verkindt, Andrea Mari, Toni Giorgino, Roberto Bizzotto, Philippe Froguel, Amelie Bonneford, Mickael Canouil, Veronique Dhennin, Caroline Brorsson, Sonek Brunak, Federico De Masi, Valborg Gudmundsdottir, Helle Pedersen, Karina Banasik, Cecilia Thomas, Peter Sackett, Hans-Henrik Staerfeldt, Agnete Lundgaard, Birgitte Nilsson, Agnes Nielsen, Gianluca Mazzoni, Tugce Karaderi, Simon Rasmussen, Joachim Johansen, Rosa Allesoe, Andreas Fritsche, Barbara Thorand, Jurek Adamski, Harald Grallert, Mark Haid, Sapna Sharma, Martinal Troll, Jonathan Adam, Jorge Ferrer, Heather Eriksen, Gary Frost, Ragna Haussler, Mun-gwan Hong, Jochen Schwenk, Mathias Uhlen, Claudia Nicolay, Imre Pavo, Birgit Steckel Hamann, Melissa Thomas, Kofi Adragni, Han Wu, Leen ‘t Hart, Slieker Roderick, Nienke van Leeuwen, Koen Dekkers, Francesca Frau, Johann Gassenhuber, Bernd Jablonka, Petra Musholt, Hartmut Ruetten, Joachim Tillner, Tania Baltauss, Oana Bernard Poenaru, Nathalie de Preville, Marianne Rodriquez, Manimozhiyan Arumugam, Kristine Allin, Line, Engelbrechtsen, Torben Hansen, Tue Hansen, Annemette Forman, Anna Jonsson, Oluf Pedersen, Avirup Dutta, Josef Vogt, Henrik Vestergaard, Markku Laakso, Tarja Kokkola, Teemu Kuulasmaa, Paul Franks, Nick Giordano, Hugo Fitipaldi, Pascal Mutie, Maria

Klintenberg, Margit Bergstrom, Leif Groop, Martin Ridderstrale, Naeimeh Atabaki Pasdar, Harshal Deshmukh, Alison Heggie, Dianne Wake, Donna McEvoy, Ian McVittie, Mark Walker, Andrew Hattersley, Anita Hill, Angus Jones, Timothy McDonald, Mandy Perry, Rachel Nice, Michelle Hudson, Claire Thorne, Emmanouil Dermitzakis, Ana Vinuela, Louise Cabrelli, Heather Loftus, Adem Dawed, Louise Donnelly, Ian Forgie, Ewan Pearson, Colin Palmer, Andrew Brown, Robert Koivula, Agata Wesolowska-Andersen, Moustafa Abdalla, Nicky McRobert, Juan Fernandez, Yunlong Jiao, Neil Robertson, Stephen Gough, Jane Kaye, Miranda Mourby, Anubha Mahajan, Mark McCarthy, Nisha Shah, Harriet Teare, Reinhard Holl, Anitra Koopman, Femke Rutters, Joline Beulens, Lenka Groeneveld, Anitra Koopman, Brandon Whitcher.

Figure 7.1. Power estimates for the GWAS of liver iron content. We used Quanto to calculate our GWAS power in 8,289 individuals of European ancestry from UK Biobank. The line shows 80% power for the minimum standardized effect sizes (in SD units) on liver iron content that could be identified for a given effect-allele frequency at α level 5×10−8. The variants reached GWAS significance level in our discovery set (UK Biobank) are shown in black circles.

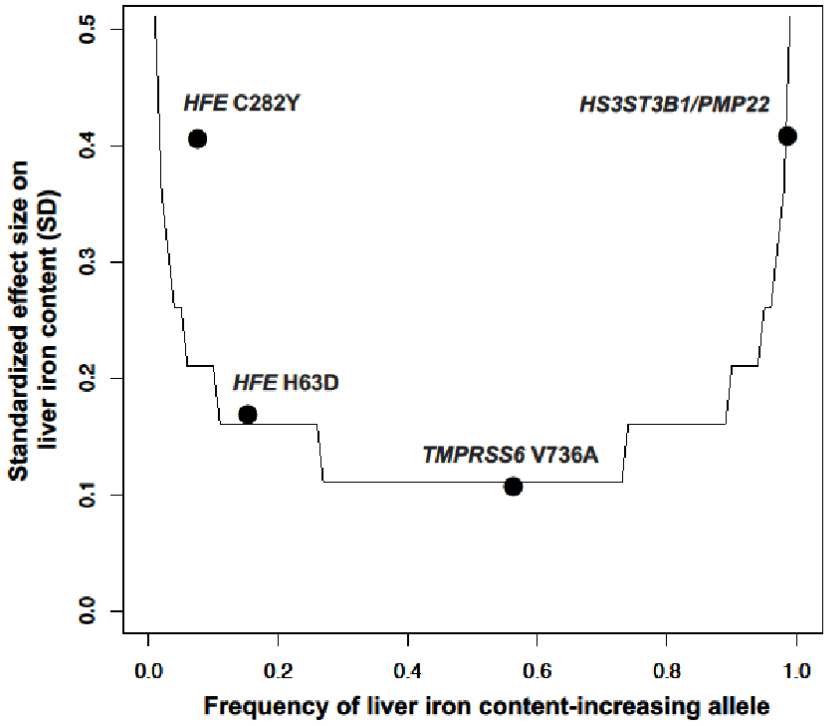


Figure 7.2. Histograms of liver iron content distributions in UK Biobank participants, stratified by sex.

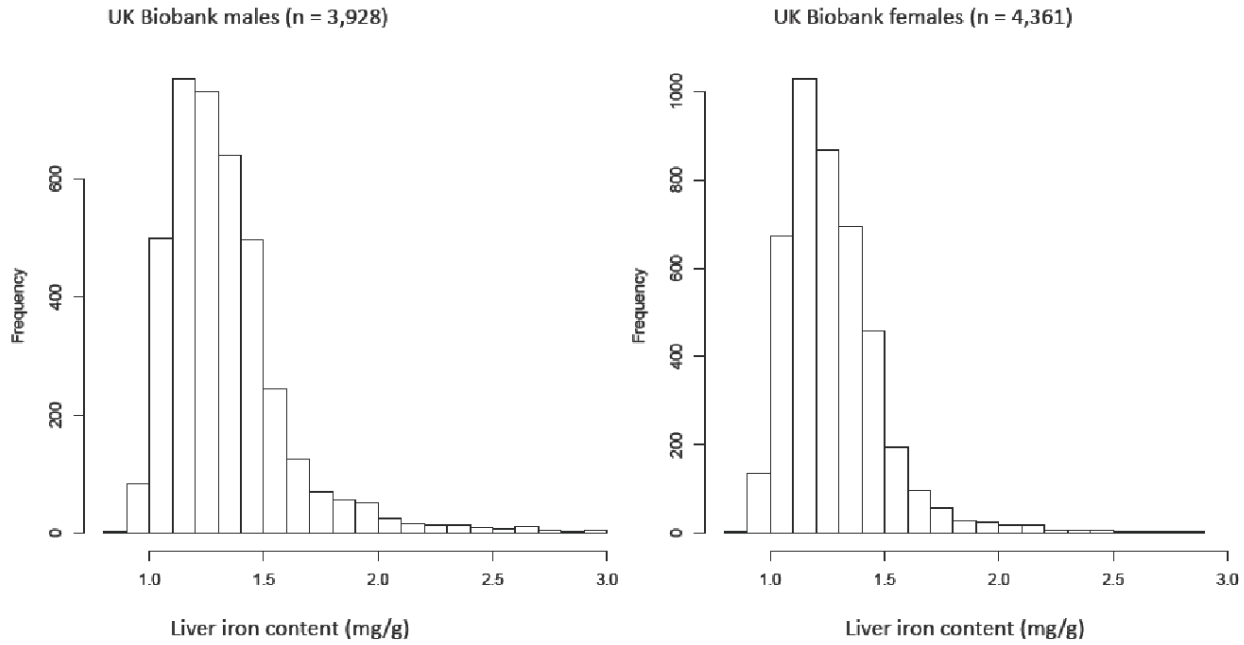


Figure 7.3. Supplementary Figure 3. Quantile quantile (QQ) plot illustrating results of genome wide association study (GWAS) for liver iron in UKB participants (8,289 individuals). Illustrates deviation of observed values (black dots) from expected values (null hypothesis, red line).

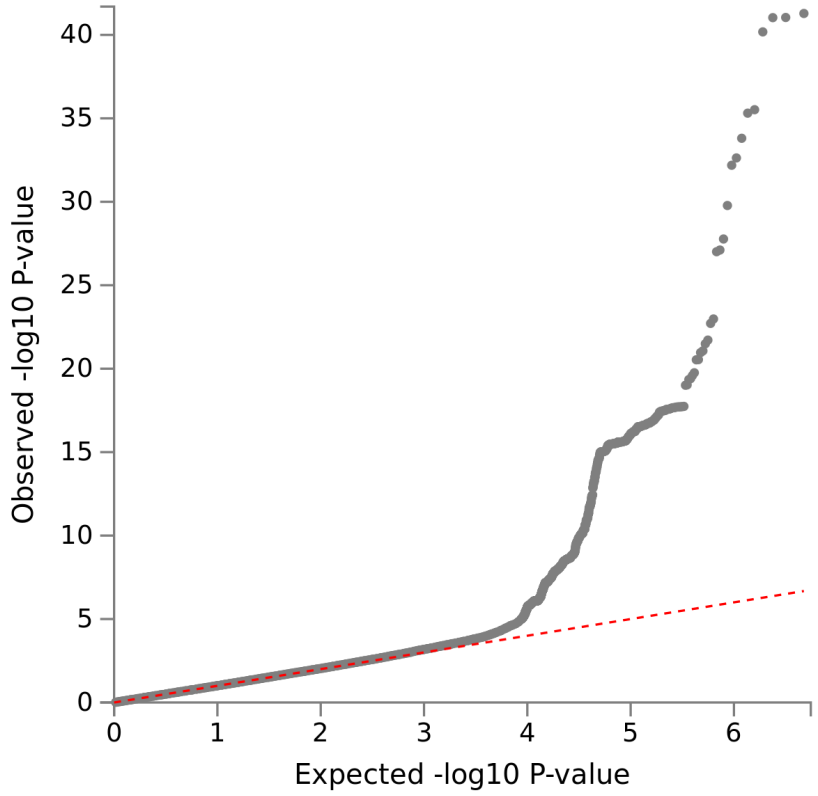


Figure 7.4. Correlation plots showing strong agreement between effect estimates from GWAS carried out separately in PLINK and GEMMA. Variants with MAF < 1 % are not included.

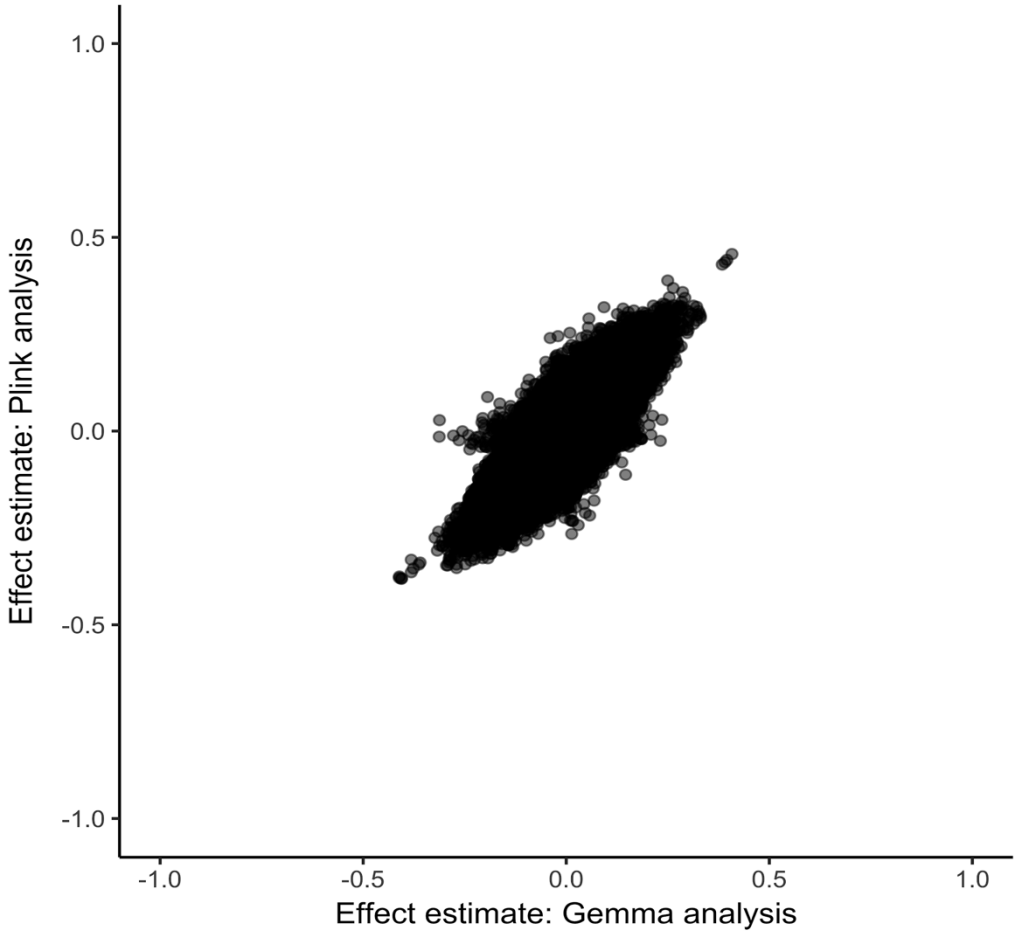


Figure 7.5. Correlation plots showing strong agreement between effect estimates from GWAS carried out separately in PLINK and GEMMA. Variants with MAF < 1 % are not included.

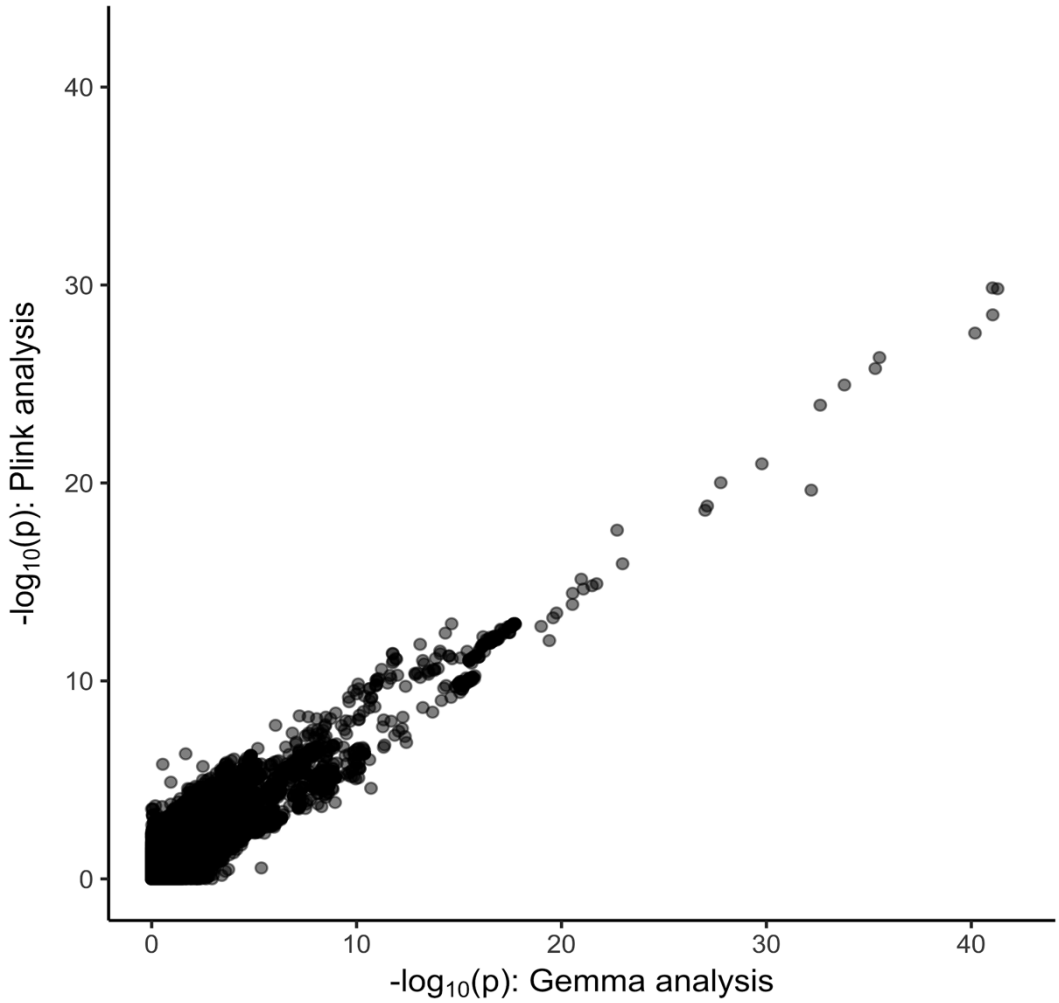


Figure 7.6. Gender specific GWAS Manhattan plots do not reveal obvious gender differences.

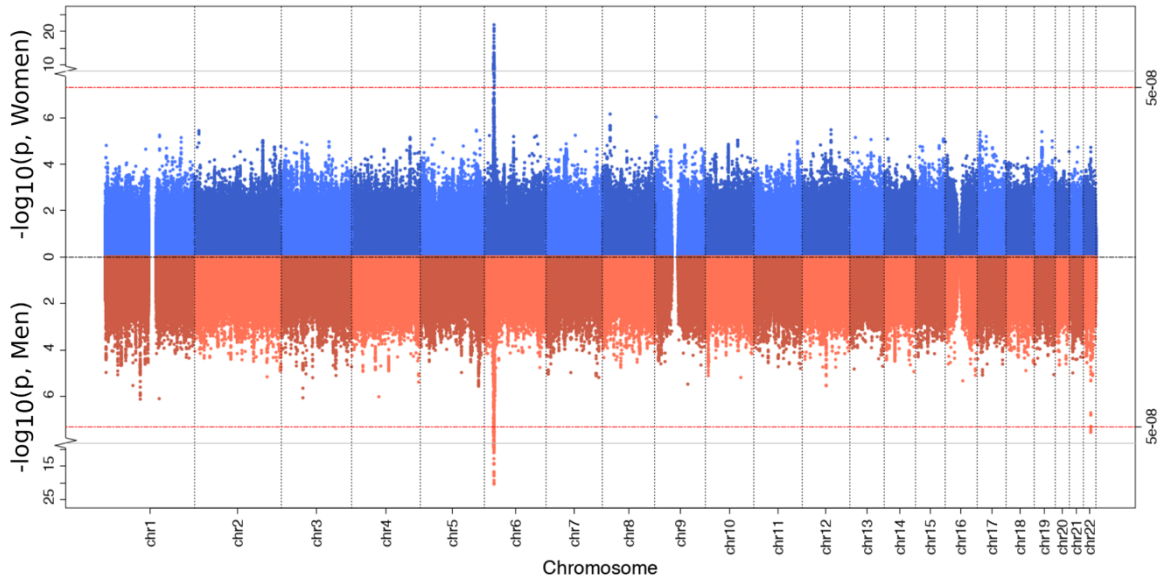


Figure 7.7. Proportion of overlapping genes in gene sets, high p-value enrichment seen with autism spectrum disorder and schizophrenia.

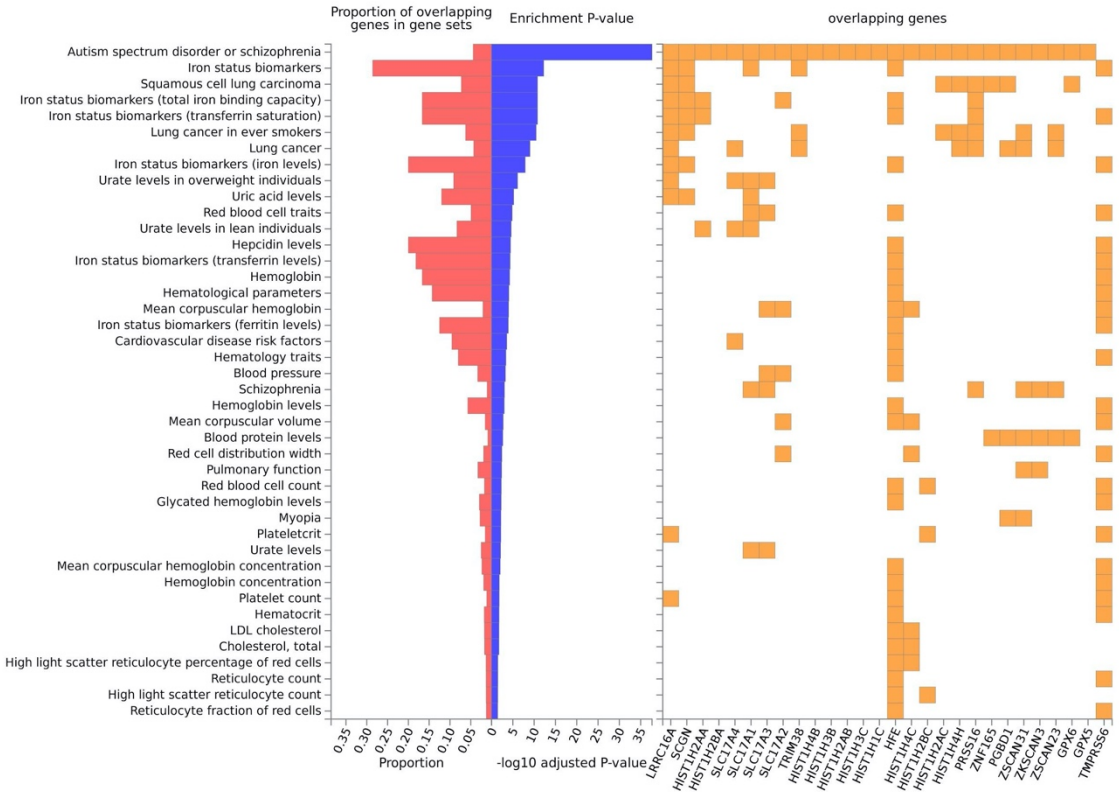


Figure 7.8. Locuszoom plot for lead independent locus in *HFE*.

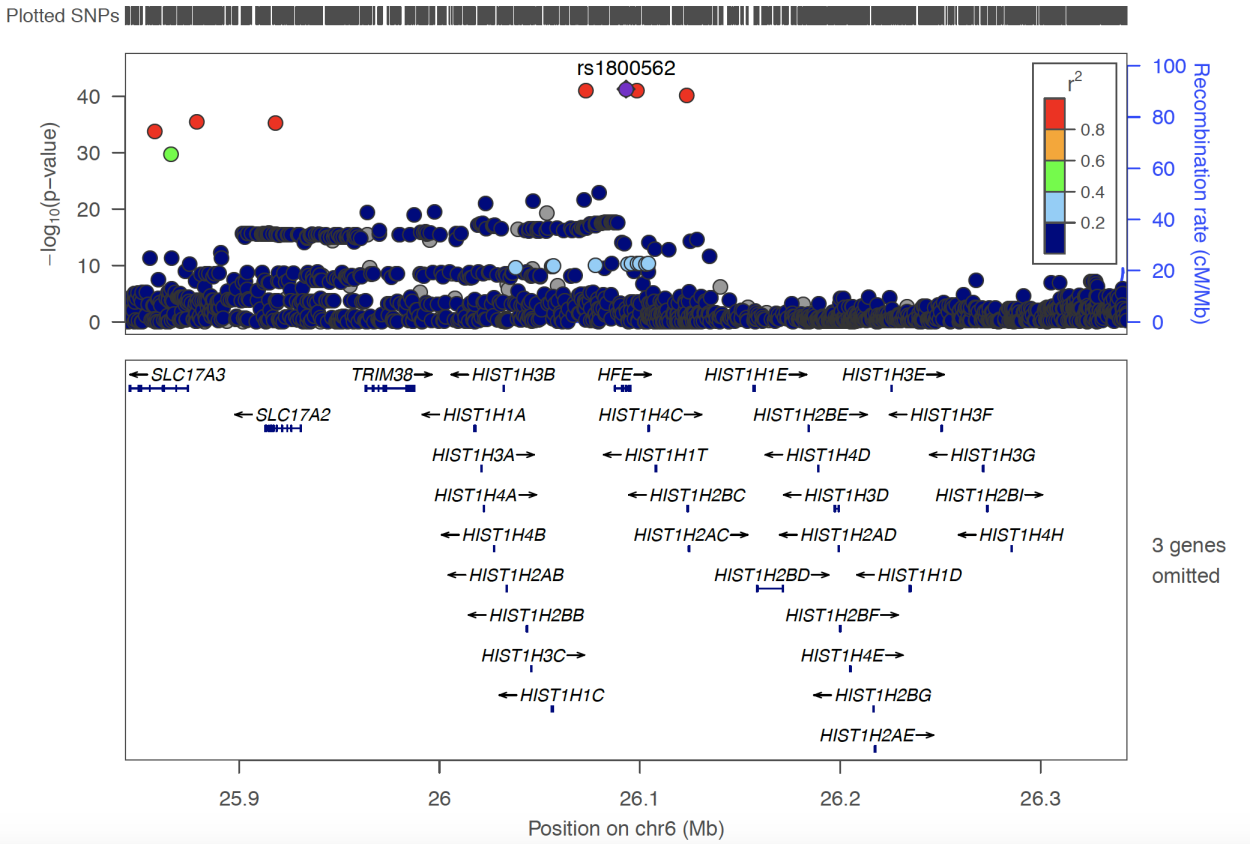


Figure 7.9. Locuszoom plot for independent locus in TMPRSS6.

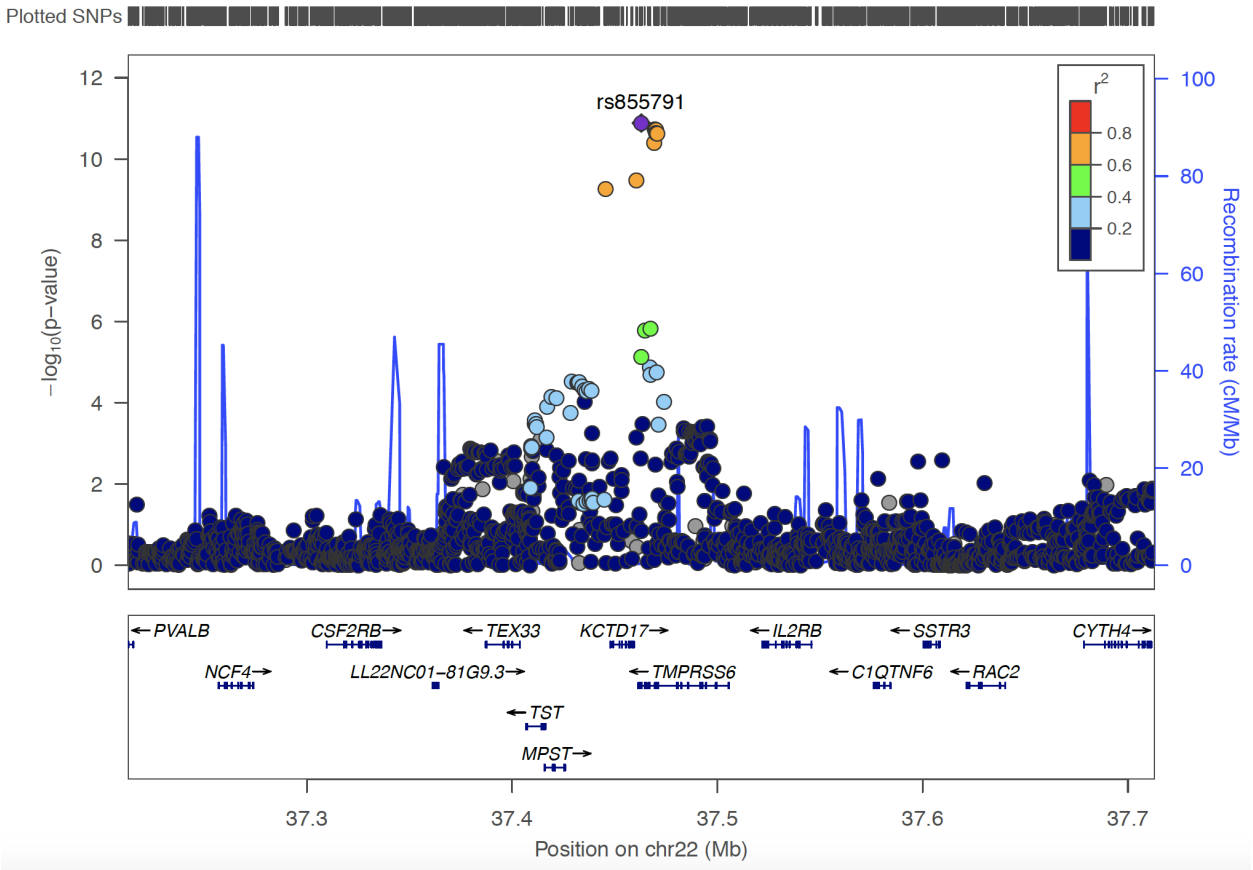


Figure 7.10. Mendelian randomisation scatter plot of waist-to-hip ratio adjusted for BMI (WHR BMIadj) vs. liver iron content. Plot shows WHR BMIadj variants and their effects (standard deviation (SD)) on liver iron content. Lines identify the slopes of the four methods tested. Error bars represent standard errors of effect sizes.

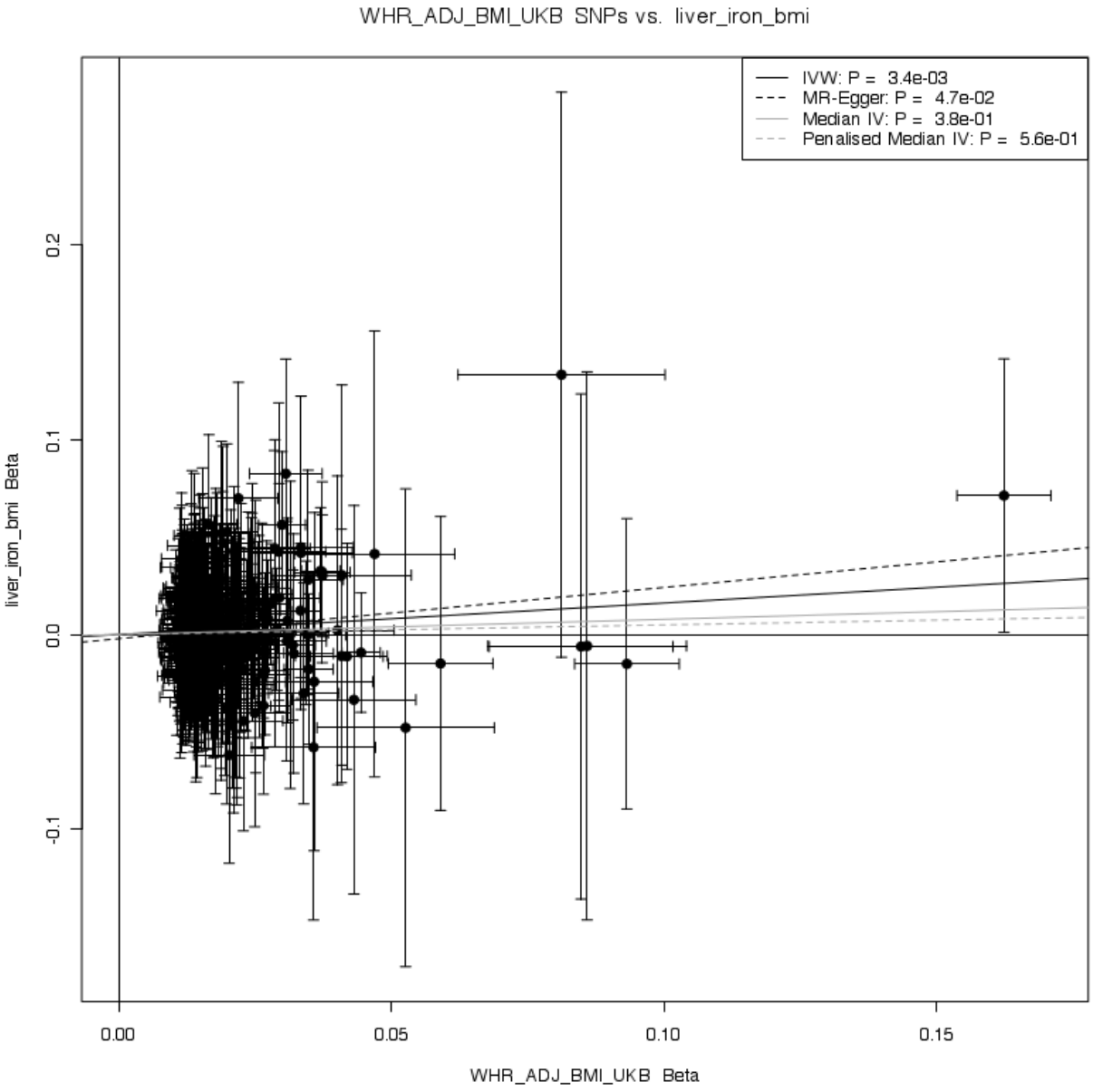


Table 7.1.1. UK Biobank fields used in main analysis

UK Biobank Field number	Description
22800-22823	Imputed genotypes (Version 3)
22001	Genetic Sex
22009	Genetic principal components
22000	Genotype measurement batch
21001	BMI (instance 2, at imaging visit)
21003	Age (instance 2, at imaging visit)
31	Sex
22006	Genetic ethnic grouping
22027	Outliers for heterozygosity or missing rate
22021	Genetic kinship to other participants
21000	Ethnic background (self-reported)
22400	Liver Iron
22402	Liver fat percentage

Table 7.1.2. Phenome-wide association study (PheWAS) between rs1800562 and 770 traits, disease codes self-reported conditions in UK Biobank and publicly available GWAS summary statistics (we present here, for clarity, traits where p < 0.05, Beta = log(odds ratio) per effect allele, FDR = False discovery rate, ICD = Summary statistics from GWAS on UK Biobank ICD codes, SR = Summary statistics derived from GWAS on self reported diseases in UK Biobank, Other_GWAS = Summary statistics outwith UK Biobank, from publicly available summary statistics).

Trait	Beta	P_value	Category	FDR
Disorders of mineral metabolism	0.013277	< E-200	ICD	1.18E-300
Transferrin Saturation	0.5772	1.52E-178	Other_GWAS	1.18E-175
Transferrin	-0.5496	1.26E-153	Other_GWAS	4.89E-151
Iron	0.3724	3.96E-77	Other_GWAS	1.02E-74
Mean cell haemoglobin	0.4252	6.50E-56	Other_GWAS	1.26E-53
Mean cell volume	0.9425	1.25E-42	Other_GWAS	1.94E-40
Ferritin	0.211	1.43E-29	Other_GWAS	1.85E-27
Standing Height	0.16057	1.89E-24	Traits	2.10E-22
HbA1C	-0.0636	2.59E-20	Other_GWAS	2.51E-18
Comparative height size at age 10	0.019167	2.18E-19	Traits	1.88E-17
Metabolic disorders	0.0093352	2.48E-16	ICD	1.92E-14
Impedance of left arm	0.91402	5.58E-14	Traits	3.94E-12
Haemoglobin concentration	0.1098	5.57E-13	Other_GWAS	3.60E-11
Hypertension	0.010657	9.65E-13	SR	5.76E-11

Impedance of right arm	0.82811	1.48E-12	Traits	8.20E-11
Mean cell haemoglobin concentration	0.0197	2.03E-10	Other_GWAS	1.05E-08
Fibrosis/cirrhosis of liver	0.0010061	7.88E-09	ICD	3.82E-07
Impedance of whole body	1.1305	3.17E-08	Traits	1.45E-06
Comparative body size at age 10	-0.012726	4.21E-08	Traits	1.81E-06
Essential (primary) hypertension	0.0068553	4.99E-07	ICD	2.04E-05
Hypertensive diseases	0.0067368	8.04E-07	ICD	3.08E-05
Coxarthrosis	0.0031267	8.34E-07	ICD	3.08E-05
Other diseases of blood	7.45E-04	8.79E-07	ICD	3.10E-05
Arthrosis	0.005578	1.23E-06	ICD	4.15E-05
Cellulitis	0.0024228	1.78E-06	ICD	5.76E-05
Infections of the skin and subcutaneous tissue	0.0029053	2.18E-06	ICD	6.77E-05
Diseases of veins	0.0052224	2.88E-06	ICD	8.60E-05
Sitting height	0.049223	3.34E-06	Traits	9.60E-05
Other anaemias	-0.0029734	3.63E-06	ICD	1.01E-04
Diseases of liver	0.0019732	3.79E-06	ICD	1.01E-04

Aplastic and other anaemias	-0.0030663	4.12E-06	ICD	1.07E-04
Total cholesterol	-0.0519	4.66E-06	Other_GWAS	1.17E-04
Other diseases of liver	0.0016064	5.56E-06	ICD	1.35E-04
LDL cholesterol	-0.0526	6.91E-06	Other_GWAS	1.62E-04
Demyelinating diseases of the central nervous system	0.001006	1.70E-05	ICD	3.88E-04
Varicose veins of lower extremities	0.0026244	1.79E-05	ICD	3.97E-04
Other arthrosis	0.0028573	1.44E-04	ICD	0.00301969
Number of self-reported non-cancer illnesses	0.024934	1.42E-04	SR	0.00301969
Multiple sclerosis	8.19E-04	1.65E-04	ICD	0.00336008
Alcoholic liver disease	6.40E-04	2.75E-04	ICD	0.00546821
Gout	0.0012223	3.24E-04	ICD	0.00629452
Non-insulin-dependent diabetes mellitus	0.0027233	3.43E-04	ICD	0.00634325
Oesophageal varices	4.94E-04	3.43E-04	ICD	0.00634325
Polyarthrosis	0.0013382	4.21E-04	ICD	0.00760101

Chronic/degenerative neurological problem	0.0010099	9.15E-04	SR	0.01614539
BMI	-0.049401	0.0010255	Traits	0.01768418
Impedance of right leg	-0.056706	0.0012024	Traits	0.01985239
Impedance of left leg	-0.052419	0.0016392	Traits	0.0265004
Osteoarthritis	0.0032516	0.0017763	SR	0.02813079
Osteomyelitis	4.31E-04	0.0019486	ICD	0.03011489
Number of treatments/medications taken	0.028575	0.0019792	SR	0.03011489
haematology	0.0014072	0.0021074	SR	0.03144889
Gall bladder disease	-0.0015547	0.0023865	SR	0.03494196
Cutaneous abscess, furuncle and carbuncle	8.82E-04	0.0025553	ICD	0.03672061
osteomyelitis	4.56E-04	0.0030174	SR	0.04257277
Maternal care for other conditions predominantly related to pregnancy	0.0013672	0.0032596	ICD	0.04516874
Hand grip strength (right)	-0.07552	0.0037987	Traits	0.05171564
Epilepsy	0.0010596	0.0038821	ICD	0.05193982

Malignant neoplasm of kidney, except renal pelvis	5.02E-04	0.0041326	ICD	0.05271919
Joint disorder	0.0035837	0.0041987	SR	0.05271919
Other dermatitis	6.76E-04	0.0042121	ICD	0.05271919
Scoliosis	5.63E-04	0.004075	ICD	0.05271919
Iron deficiency anaemia	-0.0014948	0.0045075	ICD	0.05552095
Diabetes mellitus	0.0022228	0.004908	ICD	0.0595095
Seborrhoeic keratosis	9.18E-04	0.0062934	ICD	0.07513351
Disorders of vitreous body and globe	-6.82E-04	0.0064438	ICD	0.07576347
Other diseases of gallbladder	-6.93E-04	0.0068246	ICD	0.07904313
Hernia	0.0032157	0.0070088	ICD	0.07998278
Pneumothorax	-4.70E-04	0.0074926	ICD	0.0842646
Other diseases of digestive system	0.0015279	0.0081383	ICD	0.09021887
Nutritional anaemias	-0.0014561	0.0087115	ICD	0.09389061
Spontaneous rupture of synovium and tendon	3.67E-04	0.0085996	ICD	0.09389061
obstetric problem	-9.79E-04	0.0092214	SR	0.09802475

Unspecified maternal hypertension	-7.03E-04	0.0096896	ICD	0.10160986
Coagulation defects, purpura and other haemorrhagic conditions	7.96E-04	0.010151	ICD	0.10364705
Ease of skin tanning	-0.0082678	0.010109	SR	0.10364705
deep venous thrombosis (dvt)	0.0013793	0.010681	SR	0.10764229
Pleural plaque	-4.25E-04	0.01084	ICD	0.1078441
Inguinal hernia	0.0018168	0.011232	ICD	0.11032952
Ulcerative colitis	8.56E-04	0.011667	ICD	0.11259664
Other disorders of bone density and structure	5.03E-04	0.011753	ICD	0.11259664
Other soft tissue disorders, not elsewhere classified	0.0016468	0.01249	ICD	0.11819805
Acquired deformities of fingers and toes	0.0014249	0.014362	ICD	0.13400781
Spondylosis	0.0012694	0.014506	ICD	0.13400781
Phlebitis and thrombophlebitis	9.02E-04	0.01508	ICD	0.13767153
eye trauma	-3.53E-04	0.015467	SR	0.1395627
Oesophagitis	0.001381	0.015892	ICD	0.14174933

Atrophic disorders of skin	-7.70E-04	0.016671	ICD	0.14700791
Hand grip strength (left)	-0.061979	0.017494	Traits	0.15083716
Hayfever/allergic rhinitis	0.0020856	0.017318	SR	0.15083716
Other diseases of urinary system	0.0022962	0.0204	ICD	0.17396044
Other diseases of anus and rectum	0.0020638	0.021017	ICD	0.17727383
Diabetes	0.0017854	0.022166	SR	0.18198008
Glaucoma	8.85E-04	0.022506	SR	0.18198008
Venous thromboembolic disease	0.0013882	0.022513	SR	0.18198008
Prolonged pregnancy	8.55E-04	0.025276	ICD	0.202208
Disorders of vitreous body	-4.93E-04	0.026942	ICD	0.21333665
Chronic nephritic syndrome	3.92E-04	0.027676	ICD	0.21693511
Gout	0.15	0.03010851	Other_GWAS	0.23057851
Hip circumference	0.018	0.03	Other_GWAS	0.23057851
Other disorders of urinary system	0.0017628	0.030308	ICD	0.23057851
Other arthritis	0.0013456	0.034365	ICD	0.25890524
Purpura and other haemorrhagic conditions	5.17E-04	0.036492	ICD	0.26465226
fracture head & neck	4.28E-04	0.036387	SR	0.26465226

Other retinal disorders	-7.46E-04	0.035812	ICD	0.26465226
Spondylopathies	0.0013426	0.036003	ICD	0.26465226
Other disorders of ear, not elsewhere classified	3.18E-04	0.037073	ICD	0.26637637
Infectious arthropathies	2.76E-04	0.038766	ICD	0.27598547
Neoplasms of uncertain or unknown behaviour	5.29E-04	0.039637	ICD	0.27962102
Other disorders of external ear	3.89E-04	0.040379	ICD	0.28119607
Fissure and fistula of anal and rectal regions	-6.95E-04	0.040585	ICD	0.28119607
high cholesterol	-0.0023749	0.044088	SR	0.29241272
kidney stone/ureter stone/bladder stone	6.89E-04	0.043154	SR	0.29241272
Cervical disk disorders	4.74E-04	0.043681	ICD	0.29241272
Trunk fat mass	0.032755	0.043875	Traits	0.29241272
Urinary albumin-to-creatinine ratio	-0.16	0.043	Other_GWAS	0.29241272
Postprocedural endocrine and metabolic disorders, not elsewhere classified	4.38E-04	0.046817	ICD	0.30529405

Disorders of skin appendages	0.0012031	0.046798	ICD	0.30529405
Corrected insulin response	0.1	0.04764	Other_GWAS	0.308072
Haemorrhoids	0.0017526	0.049741	ICD	0.31900013

Table 7.1.3. Phenome-wide association study (PheWAS) between rs1799985 and 770 traits, disease codes self-reported conditions in UK Biobank and publicly available GWAS summary statistics (we present here, for clarity, traits where p < 0.05, Beta = log(odds ratio) per effect allele, FDR = False discovery rate, ICD = Summary statistics from GWAS on UK Biobank ICD codes, SR = Summary statistics derived from GWAS on self reported diseases in UK Biobank, Other_GWAS = Summary statistics outwith UK Biobank, from publicly available summary statistics).

Trait	Beta	P_value	Category	FDR
Transferrin Saturation	0.228	2.98E-60	Other_GWAS	2.32E-57
Mean cell haemoglobin	0.2172	4.01E-47	Other_GWAS	1.56E-44
Iron	0.1896	1.65E-42	Other_GWAS	4.27E-40
Mean cell volume	0.4633	2.35E-33	Other_GWAS	4.56E-31
Haemoglobin concentration	0.0938	3.60E-26	Other_GWAS	5.59E-24
Hypertension	0.011698	1.48E-25	SR	1.92E-23
Transferrin	-0.1188	5.59E-17	Other_GWAS	6.20E-15
Mean cell haemoglobin concentration	0.0129	1.50E-12	Other_GWAS	1.46E-10
Essential (primary) hypertension	0.0066033	1.05E-10	ICD	9.06E-09
Hypertensive diseases	0.0065562	1.49E-10	ICD	1.16E-08
Ferritin	0.0589	7.38E-06	Other_GWAS	5.21E-04
HbA1C	-0.0181	1.43E-04	Other_GWAS	0.00925925
Malabsorption/coeliac disease	-7.01E-04	2.47E-04	SR	0.01463202

Meantime to correctly identify matches	-1.0768	2.64E-04	SR	0.01463202
Ankylosing spondylitis	5.58E-04	3.05E-04	SR	0.01578398
Variation in diet	0.0058715	4.21E-04	SR	0.02045647
Malignant neoplasm of bladder	4.27E-04	7.06E-04	ICD	0.03228298
Intestinal malabsorption	-6.85E-04	9.68E-04	ICD	0.04180087
gynaecological disorder (not cancer)	0.0035203	0.0011799	SR	0.0482517
uterine problem	0.0028743	0.0014695	SR	0.05709008
Nutritional anaemias	-0.0012992	0.0017931	ICD	0.0663447
Iron deficiency anaemia	-0.0012222	0.0019474	ICD	0.06877863
Other anaemias	-0.0014475	0.0026304	ICD	0.08886177
Dorsalgia	-0.0015782	0.0028488	ICD	0.0922299
soft tissue inflammation	-3.48E-04	0.003833	SR	0.11746746
urinary frequency / incontinence	4.98E-04	0.0039307	SR	0.11746746
Hand grip strength (right)	0.056029	0.0041738	Traits	0.12011269
Unspecified maternal hypertension	5.74E-04	0.0048111	ICD	0.13350802
AUCins/AUCglu	-0.089	0.007847	Other_GWAS	0.19157421
Aplastic and other anaemias	-0.0013261	0.0078898	ICD	0.19157421

Schizophrenia, schizotypal and delusional disorders	-3.78E-04	0.0077531	ICD	0.19157421
Number of self-reported non-cancer illnesses	0.013108	0.0076262	SR	0.19157421
Viral agents as the cause of diseases classified to other chapters	-2.69E-04	0.011049	ICD	0.25036667
Comparative height size at age 10	0.0040296	0.011564	Traits	0.25036667
Ischemic stroke	0.0589	0.0116	Other_GWAS	0.25036667
other abdominal problem	-8.33E-04	0.011555	SR	0.25036667
Diseases of veins, lymphatic vessels and lymph nodes, not elsewhere classified	0.0021004	0.012054	ICD	0.253134
Other dorsopathies	-0.0015424	0.012395	ICD	0.25344513
Postprocedural endocrine and metabolic disorders, not elsewhere classified	4.10E-04	0.013031	ICD	0.25404015
Other specific joint derangements	-5.74E-04	0.013078	ICD	0.25404015
Inflammatory bowel disease	-0.058599	0.0136	Other_GWAS	0.25773659
Varicose veins of lower extremities	0.0011122	0.01531	ICD	0.25860587
LDL cholesterol	-0.0203	0.01425	Other_GWAS	0.25860587

mumps / epidemic parotitis	3.29E-04	0.015122	SR	0.25860587
Other diseases of urinary system	-0.0018048	0.015036	ICD	0.25860587
Standing height	0.028643	0.015167	Traits	0.25860587
Malignant neoplasm of colon	-5.29E-04	0.01615	ICD	0.26699043
Maternal care for other known or suspected foetal problems	8.52E-04	0.016562	ICD	0.26809738
Insulin at 30 minutes	-0.077	0.017936	Other_GWAS	0.27872544
Urate	0.03714019	0.01769045	Other_GWAS	0.27872544
Chronic ischaemic heart disease	-0.0015585	0.018906	ICD	0.28803847
cerebrovascular disease	8.87E-04	0.023869	SR	0.33905455
Crohn's disease	-0.0729995	0.02376	Other_GWAS	0.33905455
Phobic anxiety disorders	-2.43E-04	0.023984	ICD	0.33905455
Obesity class 1	0.041	0.024	Other_GWAS	0.33905455
menorrhagia (unknown cause)	6.20E-04	0.024994	SR	0.34679175
pneumothorax	3.44E-04	0.025717	SR	0.35056332
AUCins	-0.073	0.027434	Other_GWAS	0.350686
Malignant neoplasms of digestive organs	-7.37E-04	0.028434	ICD	0.350686

colitis/not crohns or ulcerative colitis	-3.29E-04	0.02739	SR	0.350686
Height	0.04034753	0.02761489	Other_GWAS	0.350686
Other diseases of stomach and duodenum	-8.40E-04	0.028112	ICD	0.350686
Neuromuscular dysfunction of bladder, not elsewhere classified	-4.00E-04	0.028245	ICD	0.350686
Number of treatments/medications taken	0.015113	0.029067	SR	0.35289155
Diseases of arteries, arterioles and capillaries	-8.32E-04	0.031678	ICD	0.37293645
Unspecified lump in breast	-3.84E-04	0.031667	ICD	0.37293645
Malignant neoplasm of breast	0.001655	0.034513	ICD	0.3876564
Malignant neoplasm of breast	0.001655	0.034513	ICD	0.3876564
Osteonecrosis	2.26E-04	0.034924	ICD	0.3876564
Acute tonsillitis	2.26E-04	0.03665	ICD	0.39103323
Other disorders of bladder	-8.72E-04	0.036738	ICD	0.39103323
Total cholesterol	-0.0163	0.03608	Other_GWAS	0.39103323
Trunk fat mass	0.025299	0.03785	Traits	0.397425

Trunk fat percentage	0.035428	0.040957	Traits	0.42431452
Gout	0.094	0.04160193	Other_GWAS	0.42532503
Urticaria and erythema	2.98E-04	0.042818	ICD	0.42653315
miscarriage	4.91E-04	0.042474	SR	0.42653315
Malignant neoplasms of urinary tract	3.73E-04	0.045212	ICD	0.43912155
Fissure and fistula of anal and rectal regions	-5.10E-04	0.045165	ICD	0.43912155
Impedance of arm (right)	0.17281	0.048858	Traits	0.46867489
Recurrent and persistent haematuria	2.37E-04	0.049943	ICD	0.47324038

Table 7.1.4. Phenome-wide association study (PheWAS) between rs855791 and 770 traits, disease codes self-reported conditions in UK Biobank and publicly available GWAS summary statistics (we present here, for clarity, traits where p < 0.05, Beta = log(odds ratio) per effect allele, FDR = False discovery rate, ICD = Summary statistics from GWAS on UK Biobank ICD codes, SR = Summary statistics derived from GWAS on self reported diseases in UK Biobank, Other_GWAS = Summary statistics outwith UK Biobank, from publicly available summary statistics).

Trait	Beta	P_value	Category	FDR
Transferrin Saturation	0.1921	3.50E-80	Other_GWAS	2.72E-77
Iron	0.1868	4.31E-77	Other_GWAS	1.67E-74
Mean cell haemoglobin	0.1928	1.01E-69	Other_GWAS	2.62E-67
Mean cell volume	0.4257	2.40E-54	Other_GWAS	4.66E-52
Haemoglobin concentration	0.0791	4.65E-40	Other_GWAS	7.23E-38
Mean cell haemoglobin concentration	0.0124	3.11E-17	Other_GWAS	4.03E-15
HbA1C	-0.0271	2.74E-14	Other_GWAS	3.04E-12
Aplastic and other anaemias	-0.0023713	4.96E-11	ICD	4.82E-09
Other anaemias	-0.0021435	7.13E-10	ICD	6.16E-08
Ferritin	0.0513	5.81E-08	Other_GWAS	4.51E-06
Disorders of mineral metabolism	5.92E-04	8.42E-05	ICD	0.00594758
Transferrin	-0.0395	1.29E-04	Other_GWAS	0.00835275
Chronic ischaemic heart disease	-0.0017122	3.48E-04	ICD	0.02077519
Angina pectoris	-0.0014759	4.03E-04	ICD	0.02237483

abdominal hernia	-5.49E-04	5.57E-04	SR	0.02884587
Nutritional anaemias	-0.0010314	6.07E-04	ICD	0.02904152
Lipidaemias	-0.0019186	6.35E-04	ICD	0.02904152
Iron deficiency anaemia	-8.71E-04	0.0022528	ICD	0.09724587
Ischaemic heart diseases	-0.0015564	0.0027755	ICD	0.11350334
Gingivitis and periodontal diseases	-2.70E-04	0.0043173	ICD	0.16772711
Other disorders of brain	-3.50E-04	0.0072543	ICD	0.2684091
Malignant neoplasms of digestive organs	6.40E-04	0.0086194	ICD	0.30442154
chronic/degenerative neurological problem	-4.20E-04	0.010599	SR	0.34751325
Premature rupture of membranes	-5.01E-04	0.010734	ICD	0.34751325
peripheral vascular disease	2.91E-04	0.011528	SR	0.35829024
Lung cancer	-0.045758	0.012554	Other_GWAS	0.37517146
Nasal polyp	-4.55E-04	0.013605	ICD	0.39152167
dermatology	-0.0010167	0.015101	SR	0.41905275
Haemorrhoids	-0.0011652	0.016093	ICD	0.4249672
other abdominal problem	-5.73E-04	0.016408	SR	0.4249672

Unspecified renal colic	-3.38E-04	0.017789	ICD	0.44587268
Coffee intake	0.0096721	0.020428	SR	0.4535016
Comparative height size at age 10	0.0026506	0.01988	Traits	0.4535016
Disorders of muscles	-2.57E-04	0.020078	ICD	0.4535016
Diseases of Bartholin's gland	-3.83E-04	0.018862	ICD	0.4535016
Lung adenocarcinoma	-0.0648425	0.021971	Other_GWAS	0.47420742
Extrapyramidal and movement disorders	-3.27E-04	0.024441	ICD	0.513261
Malignant neoplasm of colon	3.55E-04	0.025919	ICD	0.52202446
Other disorders of muscle	-2.19E-04	0.026202	ICD	0.52202446
eczema/dermatitis	-7.10E-04	0.031241	SR	0.53614689
Parkinson's disease	-2.42E-04	0.028552	ICD	0.53614689
Diseases of tongue	2.49E-04	0.030177	ICD	0.53614689
Redundant prepuce, phimosis and paraphimosis	-6.99E-04	0.031741	ICD	0.53614689
Obesity class 2	0.043	0.028	Other_GWAS	0.53614689
oesophageal disorder	-0.0011093	0.030462	SR	0.53614689
Triglycerides	0.0106	0.02945	Other_GWAS	0.53614689

Bacterial diseases	4.62E-04	0.042392	ICD	0.55084205
Hypothyroidism	-7.82E-04	0.034171	ICD	0.55084205
Other disorders of fluid, electrolyte and acid-base balance	4.80E-04	0.043162	ICD	0.55084205
Demyelinating diseases	-2.62E-04	0.037883	ICD	0.55084205
Other disorders of the nervous system	-4.40E-04	0.035772	ICD	0.55084205
heart/cardiac problem	-0.0010641	0.042489	SR	0.55084205
Other forms of heart disease	-0.0010709	0.03837	ICD	0.55084205
Internalizing problems	-0.0668	0.04297	Other_GWAS	0.55084205
irritable bowel syndrome	6.53E-04	0.04067	SR	0.55084205
Other disorders of teeth and supporting structures	-3.57E-04	0.034685	ICD	0.55084205
Femoral hernia	-1.67E-04	0.042801	ICD	0.55084205
Renal failure	6.15E-04	0.03951	ICD	0.55084205
Unspecified lump in breast	2.62E-04	0.043245	ICD	0.55084205
Number of self-reported cancers	0.0013253	0.04072	SR	0.55084205

Maternal care related to the fetus and amniotic cavity and possible delivery problems	-9.57E-04	0.034907	ICD	0.55084205
other fractures	-1.72E-04	0.044586	SR	0.55876326
Essential (primary) hypertension	-0.0014729	0.045465	ICD	0.560735
Hypertensive diseases	-0.0014694	0.046218	ICD	0.56111541
Other peripheral vascular diseases	-3.94E-04	0.047072	ICD	0.56269145
Cervical disk disorders	-2.51E-04	0.049426	ICD	0.57319406
Complications predominantly related to the puerperium	-2.97E-04	0.04875	ICD	0.57319406

Table 7.1.5. Baseline characteristics of UK Biobank participants who took part in the imaging study compared to rest of the UK Biobank cohort.

Characteristics	UK Biobank liver iron cohort		UK Biobank rest of cohort	
	Men	Women	Men	Women
No	3,928	4,361	184637	217434
Age (IQR) (years)	57 (51,62)	56 (49,61)	59 (51, 64)	58 (51, 63)
Waist Circumference (IQR) (cm)	94 (88, 101)	80 (74, 89)	96 (89, 104)	83.0 (75.2, 92.0)
Townsend deprivation index (IQR)	-2.72 (-3.95, -0.76)	-2.63 (-3.87,-0.80)	-2.33 (-3.74, 0.16)	-2.34 (-3.73, 0.03)
Self-reported diabetes (%)	134 (3.7%)	88 (2.2%)	12102 (6.6%)	7482 (3.5%)
Liver iron (IQR) (mg/g)	1.28 (1.155,1.437)	1.23 (1.13,1.38)	NA	NA
BMI (IQR) (kg/m2)	26.49 (24.3, 29)	25.08 (22.59,28.35)	27.3 (25.0, 30.1)	26.1 (23.5, 29.7)
No consuming alcohol daily (%)	1,088 (27.7%)	825 (18.9%)	48497 (26.3%)	36336 (16.7%)

Table 7.1.6. Sensitivity analyses for BMI unadjusted, alcohol adjusted and sex specific analyses in UK Biobank and separate results for participants with or without diabetes in IMI DIRECT (Effect Allele = ALLELE1, Other Allele = ALLELE0, Beta = log(odds ratio) per effect allele, SE = standard error)

UK Biobank Women								
SNP	**chr**	**pos**	**ALLELE1**	**ALLELE0**	**A1_FREQ**	**BETA**	**SE**	**P_value**
rs1800562	6	26093141	A	G	0.076	4.04E-01	4.09E-02	2.42E-22
rs1799945	6	26091179	G	C	1.53E-01	1.71E-01	2.98E-02	1.15E-08
rs855791	22	37462936	G	A	5.63E-01	9.11E-02	2.16E-02	2.57E-05
rs149275125	17	14831420	C	T	9.86E-01	3.85E-01	9.37E-02	4.16E-05
UK Biobank BMI unadjusted								
SNP	**chr**	**pos**	**ALLELE1**	**ALLELE0**	**A1_FREQ**	**BETA**	**SE**	**P_value**
rs1800562	6	26093141	A	G	0.076	4.02E-01	2.93E-02	2.27E-41
rs1799945	6	26091179	G	C	1.53E-01	1.75E-01	2.16E-02	8.49E-16
rs855791	22	37462936	G	A	5.63E-01	1.07E-01	1.57E-02	1.15E-11
rs149275125	17	14831420	C	T	9.86E-01	4.03E-01	6.87E-02	4.71E-09
UK Biobank alcohol adjusted								
SNP	**chr**	**pos**	**ALLELE1**	**ALLELE0**	**A1_FREQ**	**BETA**	**SE**	**P_value**
rs1800562	6	26093141	A	G	0.04997	4.46E-01	3.66E-02	8.80E-33
rs1799945	6	26091179	G	C	0.1196	1.87E-01	2.77E-02	1.90E-11
rs855791	22	37462936	G	A	0.59145	1.08E-01	1.99E-02	7.06E-08

rs149275125	17	14831420	C	T	9.86E-01	5.09E-01	8.62E-02	4.24E-09
DIRECT Non-Diabetics only (nr = 1010)								
SNP	**chr**	**pos**	**ALLELE1**	**ALLELE0**	**A1_FREQ**	**BETA**	**SE**	**P_value**
rs1800562	6	26093141	A	G	0.04997	0.34217	0.11693	0.0035
rs1799945	6	26091179	G	C	0.1196	0.17716	0.07098	0.0127
rs855791	22	37462936	G	A	0.59145	0.18474	0.04387	2.77E-05
rs149275125	17	14831420	C	T	0.99069	-0.311771	0.3051	0.307
DIRECT Diabetics only (nr = 503)								
SNP	**chr**	**pos**	**ALLELE1**	**ALLELE0**	**A1_FREQ**	**BETA**	**SE**	**P_value**
rs1800562	6	26093141	A	G	0.04997	0.32545	0.11389	0.00444
rs1799945	6	26091179	G	C	0.1196	0.23808	0.08316	0.00437
rs855791	22	37462936	G	A	0.59145	0.01412	0.0641	0.826
rs149275125	17	14831420	C	T	0.99069	-0.162332	0.26725	0.544

Table 7.1.7. Genetic correlations of liver iron content against 448 LD Hub traits phenotypes ordered by P-value (we present here, for clarity, traits where p < 0.2, SE = standard error).

Trait	SE	P_value
Transferrin	0.3785	0.0392
Ferritin	0.6433	0.0544
Anorexia Nervosa	0.1545	0.1061
Non-cancer illness code_ self-reported: malabsorption/coeliac disease	0.1948	0.1255
Diagnoses - main ICD10: M10 Gout	0.6034	0.1326
Urate	0.3247	0.1334
Non-cancer illness code_ self-reported: kidney stone/ureter stone/bladder stone	0.3441	0.1675
Non-cancer illness code_ self-reported: gout	0.3166	0.1735
Type 2 Diabetes	0.3258	0.1776
Tinnitus: Yes_ but not now_ but have in the past	0.4983	0.193
Age started oral contraceptive pill	0.2231	0.1982

Table 7.1.8. GTEx tissue enrichment results from FUMA (MAGMA) (Beta = log(odds ratio), SE = standard error, FDR = False discovery rate).

TISSUE	BETA	SE	P_value	FDR
Blood_Vessel	0.024	0.0123	0.025484	0.32273
Adipose_Tissue	0.0252	0.0137	0.032273	0.32273
Lung	0.0229	0.0115	0.022835	0.32273
Muscle	0.0111	0.00711	0.058704	0.34373
Blood	0.0105	0.00669	0.059211	0.34373
Nerve	0.0171	0.0115	0.068746	0.34373
Heart	0.0135	0.0097	0.082246	0.35248286
Breast	0.0175	0.0166	0.1459	0.547125
Fallopian_Tube	0.0145	0.0152	0.17044	0.56813333
Spleen	0.00662	0.00803	0.20486	0.61458
Testis	0.00382	0.00551	0.24447	0.66673636
Liver	0.00217	0.00679	0.37473	0.88763077
Thyroid	0.0034	0.0116	0.38464	0.88763077
Uterus	-0.0033	0.013	0.59997	0.97684
Brain	-0.00261	0.00707	0.64422	0.97684
Cervix_Uteri	-0.00923	0.0167	0.71013	0.97684
Skin	-0.00817	0.0112	0.76778	0.97684
Adrenal_Gland	-0.00832	0.0104	0.78835	0.97684
Ovary	-0.00929	0.0107	0.8071	0.97684
Bladder	-0.0132	0.015	0.81052	0.97684
Esophagus	-0.0166	0.0172	0.8333	0.97684
Vagina	-0.015	0.0135	0.86674	0.97684
Small_Intestine	-0.0109	0.0097	0.86847	0.97684
Salivary_Gland	-0.0144	0.0116	0.89176	0.97684

Colon	-0.0221	0.0167	0.9073	0.97684
Stomach	-0.0194	0.0144	0.91025	0.97684
Prostate	-0.0218	0.0141	0.93928	0.97684
Kidney	-0.0172	0.0106	0.94844	0.97684
Pancreas	-0.0178	0.00954	0.96913	0.97684
Pituitary	-0.0165	0.00829	0.97684	0.97684

Table 7.1.9. Gene-set enrichment results from FUMA (MAGMA) on 10,651 gene-sets (for clarity, we are presenting enriched pathways with p values < 0.001, Beta = log(odds ratio), SE = standard error, FDR = False discovery rate).

Pathway	Number of Genes	Beta	SE	P_value	FDR
GO_bp:go_mitochondrial_calcium_ion_transport	11	0.988	0.259	6.92E-05	0.56
Curated_gene_sets:chassot_skin_wound	10	1.05	0.284	1.10E-04	0.56
Curated_gene_sets:wu_hbx_targets_2_up	21	0.681	0.192	2.02E-04	0.56
Curated_gene_sets:yang_muc2_targets_duodenum_6mo_dn	20	0.581	0.165	2.10E-04	0.56
GO_bp:go_cellular_hormone_metabolic_process	99	0.27	0.0792	3.21E-04	0.61
Curated_gene_sets:kim_germinal_center_t_helper_up	61	0.37	0.11	3.83E-04	0.61
GO_bp:go_establishment_of_spindle_orientation	26	0.498	0.149	4.27E-04	0.61
GO_bp:go_uv_protection	12	0.723	0.218	4.60E-04	0.61
GO_bp:go_urate_metabolic_process	12	0.736	0.227	5.91E-04	0.64
GO_bp:go_negative_regulation_of_interleukin_2_production	15	0.709	0.219	6.05E-04	0.64
GO_bp:go_calcium_ion_import	62	0.334	0.105	7.61E-04	0.65
GO_bp:go_metal_ion_transport	545	0.111	0.0357	8.87E-04	0.65
GO_mf:go_dna_dependent_atpase_activity	72	0.268	0.0859	8.97E-04	0.65

Table 7.1.10. Mendelian randomisation studies of 29 predominantly metabolic traits (added as positive controls are 4 serum iron markers) and liver iron content. Presented here are results from main analysis (inverse variance weighting) and sensitivity analyses (Egger, penalised weighted median). BMIadj = BMI adjusted, SHBG = sex hormone binding globulin. Columns with the associated p-values for each MR method are in bold.

Exposure	betaIVW2	sebetaIVW2	tIVW	pIVW	IVR_int	p.hetero	betaEgger	sebetaEgger	tegger	pEgger	egger_int	int_p	betaWM	sebetaWM	tWM	pWM	betaPWM
Transferrin saturation	0.645	0.068	9.47	**0.001**	0	0.011949102	0.77	0.09	8.77	**0.003**	-0.034	0.171	0.68	0.048	14.162	**1.58E-45**	0.691
WHR BMIadj	0.162	0.055	2.95	**0.003**	0	0.261095254	0.26	0.13	1.99	**0.047**	-0.002	0.401	0.08	0.089	0.875	**0.381719644**	0.049
Iron	0.833	0.181	4.59	**0.010**	0	9.02E-10	1.33	0.24	5.52	**0.012**	-0.091	0.096	0.84	0.095	8.778	**1.66E-18**	0.794
Ferritin	1.512	0.384	3.93	**0.011**	0	1.54E-14	2.11	0.76	2.77	**0.051**	-0.055	0.412	1.96	0.206	9.554	**1.25E-21**	1.69
Fasting glucose	0.216	0.099	2.18	**0.037**	0	0.437732612	0.33	0.17	1.89	**0.068**	-0.004	0.429	0.35	0.133	2.653	**0.007976399**	0.35
NAFLD	0.205	0.059	3.48	**0.040**	0	0.980189373	0.21	0.11	1.99	**0.185**	-0.002	0.790	0.21	0.063	3.324	**0.000887441**	0.21
Body fat %	-0.360	0.154	2.34	**0.044**	0	0.58846871	0.43	0.63	0.68	**0.518**	-0.029	0.171	-0.42	0.203	2.081	**0.0374111**	-0.42
Alanine transaminase	0.008	0.003	3.13	**0.050**	0	0.500519197	0.01	0.01	1.39	**0.298**	0.000	0.989	0.01	0.003	2.759	**0.005800579**	0.01
Transferrin	-0.360	0.163	2.21	**0.058**	0	0	-0.27	0.23	1.16	**0.284**	-0.024	0.577	-0.17	0.082	2.010	**0.044428187**	-0.28
Fasting insulin BMIadj	0.740	0.361	2.05	**0.065**	0	0.213272348	2.89	1.92	1.51	**0.163**	-0.035	0.281	0.91	0.431	2.102	**0.035541546**	0.92
Insulin secretion	0.253	0.122	2.07	**0.065**	0	0.769708328	0.45	0.21	2.14	**0.061**	-0.010	0.167	0.38	0.148	2.538	**0.011160081**	0.38
BMI	-0.129	0.072	1.80	**0.075**	0	0.05693282	-0.11	0.16	0.73	**0.466**	0.000	0.914	0.01	0.118	0.064	**0.94900973**	0.01
Coronary artery disease	-0.052	0.029	1.80	**0.077**	0	0.432771004	0.04	0.06	0.64	**0.524**	-0.008	0.079	-0.04	0.044	0.899	**0.368783993**	-0.04
2hGlu	-0.109	0.084	1.31	**0.232**	0	0.276652622	-0.43	0.33	1.31	**0.239**	0.028	0.352	-0.09	0.100	0.896	**0.37007294**	-0.09
Favourable adiposity	-0.277	0.272	1.02	**0.326**	0	0.086257008	0.04	0.81	0.05	**0.960**	-0.007	0.684	-0.41	0.293	1.402	**0.160923723**	-0.42
LDL-C	-0.069	0.073	0.94	**0.350**	0	0	-0.13	0.11	1.12	**0.267**	0.004	0.503	0.03	0.060	0.421	**0.673994471**	0.03

Leptin BMIadj	-0.096	0.102	0.94	**0.380**	0	0.591483977	0.33	0.58	0.57	**0.591**	-0.026	0.446	-0.11	0.134	0.839	**0.401547798**	-0.11
Type 2 diabetes	-0.021	0.024	0.86	**0.393**	0	0.18031828	-0.05	0.05	1.13	**0.264**	0.003	0.428	-0.07	0.036	1.857	**0.06338137**	-0.07
SHBG	-0.134	0.151	0.89	**0.397**	0	0.09672251	-0.56	0.29	1.91	**0.093**	0.024	0.138	-0.38	0.154	2.456	**0.014048854**	-0.38
Alcohol	0.440	0.500	0.88	**0.419**	0	0.070913546	2.40	2.56	0.94	**0.400**	-0.044	0.477	0.78	0.491	1.587	**0.112481391**	0.78
HDL-C	0.033	0.044	0.75	**0.458**	0	0.098009444	0.07	0.07	1.09	**0.281**	-0.002	0.431	0.07	0.065	1.032	**0.302275499**	0.07
Diastolic blood pressure	0.006	0.008	0.70	**0.484**	0	0.001082401	0.02	0.02	1.02	**0.311**	-0.005	0.409	-0.01	0.010	1.059	**0.289390137**	-0.01
Systolic blood pressure	0.004	0.006	0.69	**0.491**	0	8.70E-06	0.00	0.02	0.14	**0.890**	0.003	0.707	-0.01	0.006	1.164	**0.244246416**	-0.01
GGT	0.001	0.001	0.52	**0.608**	0	0.055656125	0.01	0.00	1.61	**0.121**	-0.020	0.143	0.00	0.001	1.100	**0.271253814**	0.00
CRP	-0.037	0.074	0.49	**0.631**	0	0.004257521	0.01	0.14	0.08	**0.939**	-0.006	0.695	-0.04	0.060	0.620	**0.535011022**	-0.04
Insulin resistance	0.067	0.147	0.45	**0.652**	0	0.132019	0.66	0.43	1.53	**0.131**	-0.012	0.149	0.17	0.199	0.842	**0.399819736**	0.32
Adiponectin	-0.012	0.030	0.41	**0.709**	0	0.729913306	0.03	0.32	0.10	**0.927**	-0.016	0.875	-0.01	0.032	0.200	**0.841143306**	-0.01
ALP	0.000	0.002	0.20	**0.841**	0	0.020129606	0.00	0.00	0.70	**0.498**	0.008	0.492	0.00	0.002	0.368	**0.713140831**	0.00

Triglycerides	0.001	0.065	0.01	**0.994**	0	0.00104925 4	-0.10	0.09	1.02	**0.314**	0.006	0.1 70	-0.09	0.089	1.069	**0.285023 462**	-0.09

Table 7.1.11. Associations previously reported in GWAS Catalog.

SNP	Nearest Gene	Associations of SNPs in Genomic Loci - GWAS Catalogue
rs1800562	*HFE*	a1c measurement, alcohol drinking, diastolic blood pressure, erythrocyte count, ferritin measurement, hematocrit, hemoglobin measurement, hepcidin:ferritin ratio, iron biomarker measurement, low density lipoprotein cholesterol measurement, mean corpuscular hemoglobin, mean corpuscular volume, serum hepcidin measurement, serum iron measurement, total cholesterol measurement, total iron binding capacity, transferrin measurement, transferrin saturation measurement
rs1799945	*HFE*	diastolic blood pressure, ferritin measurement, hematocrit, hemoglobin measurement, hypertension, mean arterial pressure, mean corpuscular hemoglobin, mean corpuscular hemoglobin concentration, platelet count, red blood cell distribution width, reticulocyte count, serum iron measurement, smoking status measurement, systolic blood pressure, transferrin saturation measurement
rs855791	*TMPRSS 6*	a1c measurement, bilirubin measurement, blood protein measurement, clinical laboratory measurement, ferritin measurement, hematocrit, hemoglobin measurement, hepcidin:ferritin ratio, hepcidin:transferrin saturation ratio, iron biomarker measurement, low density lipoprotein cholesterol measurement, mean corpuscular hemoglobin, mean corpuscular hemoglobin concentration, mean corpuscular volume, red blood cell distribution width, serum hepcidin measurement, serum iron measurement, soluble transferrin receptor measurement, transferrin measurement, transferrin saturation measurement

7.2 Genome wide and Mendelian randomisation analysis of magnetic resonance imaging of the liver yield insights into the pathogenesis of steatohepatitis.

7.2.1 Supplementary Methods

7.2.1.1 The relationship between cT1 and Liver inflammation and Fibrosis (LIF) score.

A promising, non-invasive measure of steatohepatitis and fibrosis severity is magnetic resonance imaging (MRI) based corrected T1 (cT1).[8–10] T1 relaxation time reflects extracellular fluid which is characteristic of fibrosis and inflammation. The presence of iron, which can be determined from T2* maps, has an opposing effect. Combining T2* and T1 values can correct for this opposing effect, from which cT1 (in milliseconds) is derived. Higher cT1 values are associated with both histological liver inflammation and fibrosis, although their relative contributions to the score are still unknown.[10,11] cT1 has already been used as a non-invasive outcome in randomised controlled trials for non-alcoholic steatohepatitis (NASH)[12] and is associated with liver disease outcomes.[9]

cT1 is a continuous trait, and analysed as such in our GWAS in line with other continuous traits such as blood pressure, BMI and height.[186–188] In some earlier publications, cT1 was reported using the Liver Inflammation and Fibrosis (LIF) score. The LIF score is a tri-linear mapping of cT1 onto a continuous scale from 0 to 4 based on the association of cT1 with histological fibrosis.[10] LIF categories were defined as having no (LIF <1), mild (LIF 1–1.99), moderate (LIF 2–2.99), or severe (LIF 3–4) liver disease.[9] The LIF cut-off of 1.4 had a sensitivity of 91% and a specificity of 52% for the diagnosis of non-alcoholic steatohepatitis (NASH) versus steatosis (AUROC = 0.80), and corresponds to a cT1 value of 780ms; a slightly

higher cutoff of 800ms is used in clinical trials[183] and is under evaluation by the FDA and European Medicines Agency as a diagnostic enrichment biomarker for NASH;[10,189] The LIF score is no longer used since the medical and MRI physics community is more familiar with T1 for the assessment of inflammation and fibrosis across all specialties including cardiology and neurology.[12,174,190–194] In this GWAS, the cT1 values reported are standardised across the MRI scanner model and field strength and show very high repeatability and reproducibility.[195]

7.2.1.2 Principle of Mendelian Randomisation

Observational epidemiologic studies measure the association between an exposure and an outcome, however they are vulnerable to confounding, reverse causation and other forms of bias, and do not provide evidence that an observed association is causal. MR uses genetic variants associated with an exposure of interest (e.g. BMI, cholesterol, blood pressure) to assess its causal effect on an outcome of interest (e.g. steatohepatitis). In the classic MR paradigm, genetic associations are free from confounding since they are assigned randomly at conception from parents to offspring (according to Mendel's second law) and reverse causation is precluded since the sequence of the germline is not modifiable by disease. MR can be thought of as analogous to a randomised controlled trial (RCT) that uses naturally randomised genetic variation rather than randomised allocation to a drug or treatment, as the 'intervention'.[76]

Genetic polymorphisms that are associated with an exposure of interest are used as an instrument to randomly allocate study participants to higher or lower levels of the exposure under study. Because allocation to genetic variation in levels of the exposure is random, this study design should be less susceptible to confounding. In addition, the allocation of the polymorphism occurs at conception so this study design should not be vulnerable to reverse causation. The results of a Mendelian randomisation study can be interpreted as follows: If a polymorphism (or a collection of polymorphisms, a genetic instrument) is associated with an exposure and the outcome of interest, then the observed association between the exposure and outcome is likely to be causal. If not, then the observed association between the exposure and outcome is likely to be an artefact of confounding, reverse causation or other study bias.

We used Mendelian randomisation to evaluate the causal association between multiple metabolic traits and diseases (e.g. insulin resistance, obesity, coronary artery disease)

previously observationally associated with our outcome of interest (steatohepatitis). We were able to detect exposures that are likely to be causal (e.g. central obesity, insulin resistance, type 2 diabetes), protective (favourable adiposity), but also detect exposures where previous associations were likely to be down to confounding or other biases (e.g. blood pressure).

Mendelian Randomisation also has limitations and results need to be interpreted alongside other evidence in the field in the spirit of triangulation of evidence.[77] Despite presumed random allocation of genetic polymorphisms according to Mendel's law of independent assortment, this study design is still vulnerable to confounding e.g. by population structure or pleiotropy. Confounding by population structure can be addressed by performing studies within ethnically homogeneous study populations (as in this study). Confounding by pleiotropy can be addressed by selecting polymorphisms that are only associated with the exposure of interest, but not with other exposures that are known to be causally associated with the outcome under study, as well as using statistical methods (e.g. MR Egger) as this manuscript has done to correct for pleiotropy.

To create genetic instruments for possible causes of steatohepatitis, we constructed genetic scores for 24 predominantly metabolic traits. We combined multiple independently inherited polymorphisms to create genetic instruments. These genetic scores are instruments that reflect the combined effect of the polymorphisms included on the exposure of interest. As a result, each score has a much larger effect than any individual polymorphism included in the score. Genetic instruments were constructed by using the effect sizes of independent, genome-wide significant genetic variants ($R^2 < 0.1$) associated with a particular exposure from previous GWASs. Using genetic instruments from a separate population compared to where the outcome

of interest was measured results in less bias and more power, a method known as two-sample Mendelian randomisation.[78] We investigated the potential causal associations between 24 predominantly metabolic traits on cT1 using two-sample Mendelian randomisation analysis.

List of genetic instruments used, number of SNPs comprising each genetic instrument, and PMIDs of the manuscripts from which the genetic instruments were derived.

Exposure	N_snp	PMID
Favourable adiposity	14	PMID: 30352878
Insulin resistance	53	PMID: 27841877
Insulin secretion	11	PMID: 22885924
Type 2 diabetes	89	PMID: 29632382
2 hour Glucose	8	PMID: 22885924
Fasting glucose	33	PMID: 22885924
Fasting Insulin	14	PMID: 22885924
NAFLD	4	PMID: 21423719
Body fat percentage	10	PMID: 26833246

BMI	73	PMID: 25673413
Coronary artery disease	64	PMID: 26343387
Systolic blood pressure	89	PMID: 27618452
Diastolic blood pressure	109	PMID: 27618452
LDL Cholesterol	80	PMID: 24097068
Triglycerides	62	PMID: 24097068
HDL Cholesterol	98	PMID: 24097068
Waist Hip ratio BMI adjusted	53	PMID: 30239722
Waist Hip ratio BMI adjusted (Female specific variants)	47	PMID: 30239722
Waist Hip ratio BMI adjusted (Male specific variants)	6	PMID: 30239722
Ferritin	6	PMID: 25352340
Transferrin	9	PMID: 25352340

Transferrin saturation	5	PMID: 25352340
Iron	5	PMID: 25352340
Height	809	PMID: 30124842

Figure 7.11. Histograms of cT1 values in UK Biobank stratified by sex. 2.6% of women (169 / 6,455) and 5.3% of men (299 / 5,595) had values above 800ms, a threshold that has been set in current clinical trials as a cut-off for steatohepatitis (800ms shown by red dotted line).

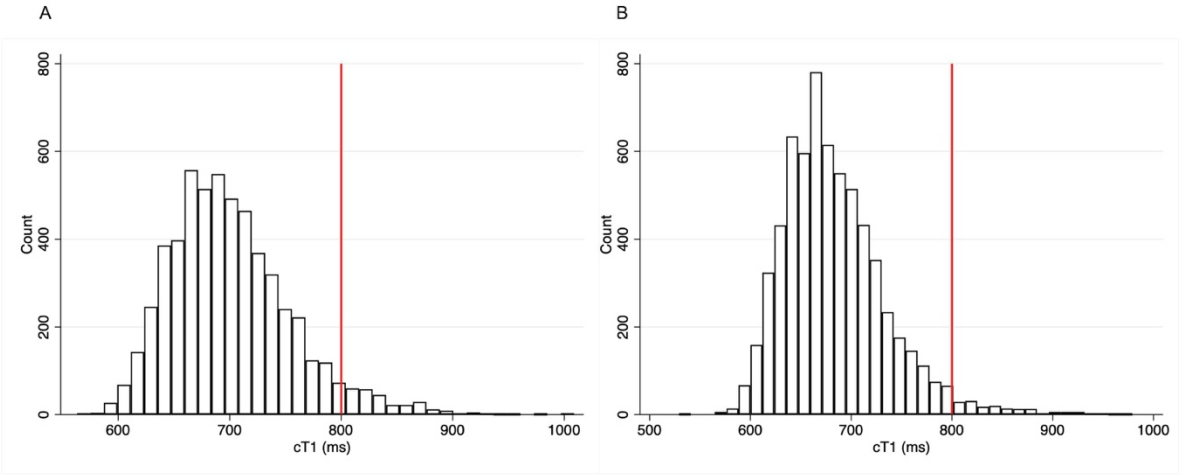


Figure 7.12. Quantile quantile plot for cT1 GWAS. The observed versus expected -log10(p-values) in our GWAS supports normality and shows no evidence of inflation.

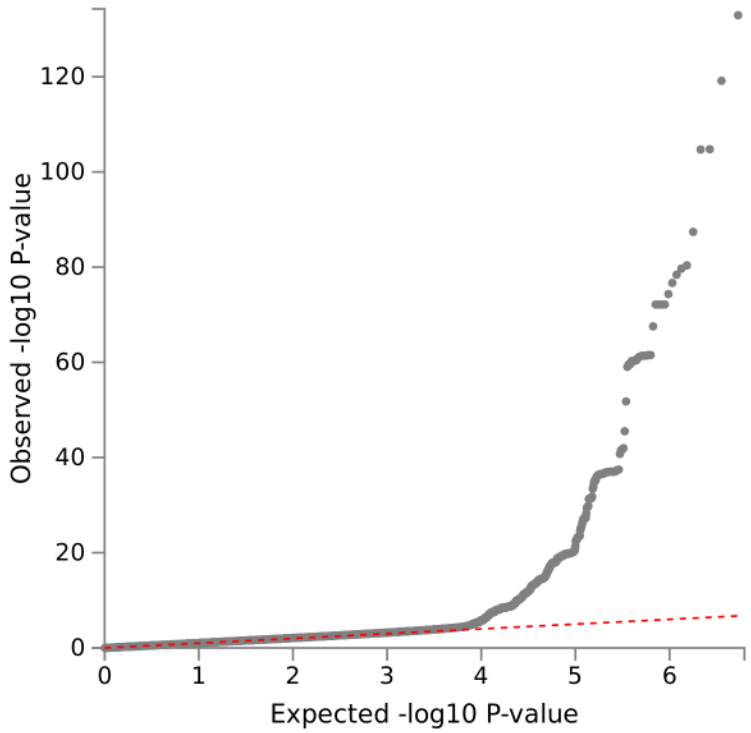


Figure 7.13. cT1 values per *SLC39A8* genotype group. Red lines indicate the median values, blue lines indicate the lower and upper quartiles.

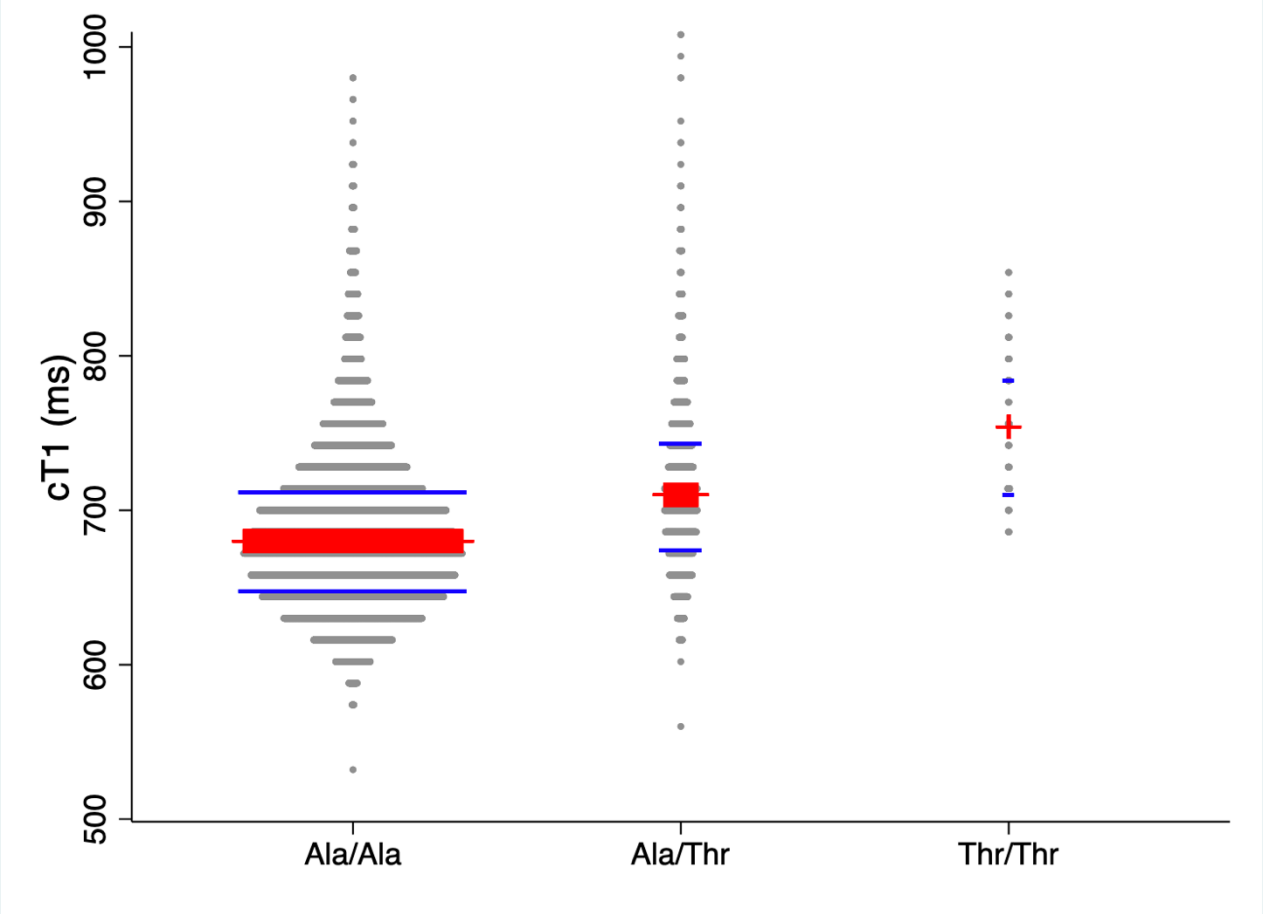


Figure 7.14. Manhattan plot illustrating GWAS of liver PDFF measurements in 14,440 UK Biobank individuals (~12 million imputed variants). The x-axis is the chromosomal position and y-axis is the significance of association for each variant in log10(p-values). Grey line indicates genome-wide significance level. For the GWAS, a linear mixed model was used. Level of significance: $p < 5 \times 10^{-8}$.

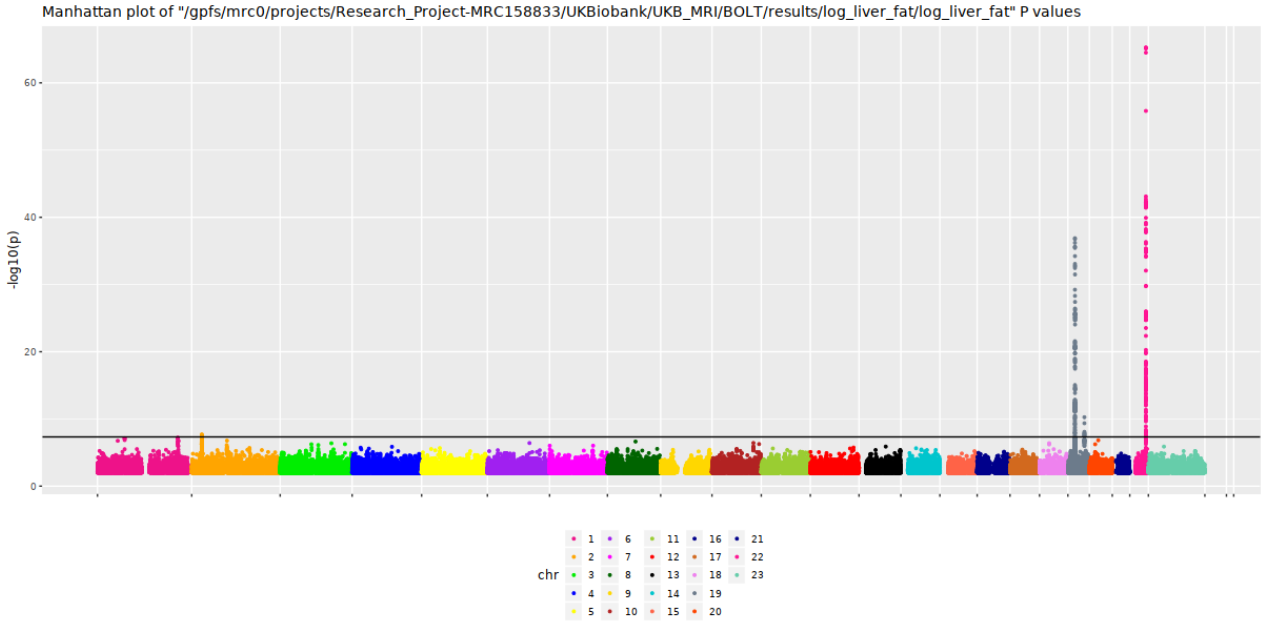


Figure 7.15. Scatterplot of correlation between cT1 and a) BMI b) liver fat% (PDFF), and c) liver iron. P values calculated using t-test. Levels of significance: p < 0.05.

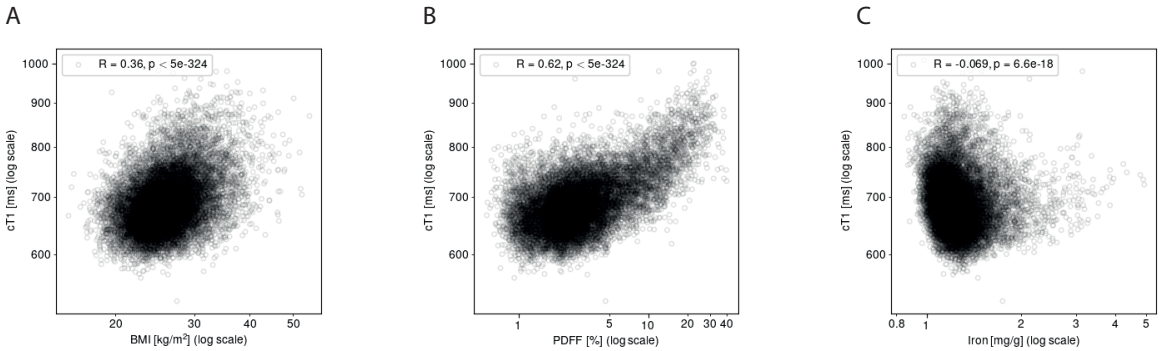


Table 7.2.1. GWAS Sensitivity analyses, with models correcting for a) BMI b) BMI & alcohol intake (in units) c) liver fat PDFF d) liver iron e) males only f) females only. A linear mixed model was used for genetic associations. Levels of significance: $p < 5 \times 10^{-8}$.

								cT1 (BMI) (N=14,405)		
SNP	**CHR**	**BP**	**EA**	**OA**	**EA5F**	**Gene**	**variant type**	**BETA**	**SE**	**P**
rs759359281	1	220100497	C	CA	0.06	*SLC30A10*	Intron	0.131773	0.0242194	1.40E-08
rs13107325	4	103188709	T	C	0.07	*SLC39A8*	Missense Variant	0.530298	0.0210205	2.40E-143
rs111723834	14	24572932	A	G	0.02	*PCK2, NRL*	Missense Variant, Intron Variant	0.29092	0.0435268	2.20E-12
rs58542926	19	19379549	T	C	0.07	*TM6SF2*	Missense Variant	0.124756	0.0208328	1.10E-09
rs4820268	22	37469591	G	A	0.46	*TMPRSS6*	Missense Variant	0.108009	0.0132949	2.30E-17
rs738409	22	44324727	G	C	0.21	*PNPLA3*	Missense Variant	0.0656351	0.0109403	2.20E-10
								cT1 (BMI & alcohol) (N=11,893)		
								BETA	**SE**	**P**
rs759359281	1	220100497	C	CA	0.06	*SLC30A10*	Intron	0.13711	0.0266234	1.30E-07

rs13107325	4	103188709	T	C	0.07	*SLC39A8*	Missense Variant	0.536308	0.0229941	4.20E-122
rs111723834	14	24572932	A	G	0.02	*PCK2, NRL*	Missense Variant, Intron Variant	0.279295	0.0478867	4.40E-10
rs58542926	19	19379549	T	C	0.07	*TM6SF2*	Missense Variant	0.132978	0.0228134	5.00E-09
rs4820268	22	37469591	G	A	0.46	*TMPRSS6*	Missense Variant	0.121534	0.0146066	3.50E-18
rs738409	22	44324727	G	C	0.21	*PNPLA3*	Missense Variant	0.0643384	0.0119858	3.20E-08

								cT1 (liver fat) (N=14,440)		
								BETA	**SE**	**P**
rs759359281	1	220100497	C	CA	0.06	*SLC30A10*	Intron	0.126816	0.021926	1.70E-09
rs13107325	4	103188709	T	C	0.07	*SLC39A8*	Missense Variant	0.540024	0.0190318	5.60E-179
rs111723834	14	24572932	A	G	0.02	*PCK2, NRL*	Missense Variant, Intron Variant	0.308174	0.0394095	2.40E-16
rs58542926	19	19379549	T	C	0.07	*TM6SF2*	Missense Variant	-0.024769	0.0189835	1.90E-01
rs4820268	22	37469591	G	A	0.46	*TMPRSS6*	Missense Variant	0.0607271	0.00991015	5.10E-11

rs738409	22	44324727	G	C	0.21	*PNPLA3*	Missense Variant	-0.0308937	0.0121655	1.90E-02
								cT1 (liver iron) (N=14,440)		
								BETA	**SE**	**P**
rs759359281	1	220100497	C	CA	0.06	*SLC30A10*	Intron	0.136471	0.0254803	2.80E-08
rs13107325	4	103188709	T	C	0.07	*SLC39A8*	Missense Variant	0.541497	0.0221264	1.10E-133
rs111723834	14	24572932	A	G	0.02	*PCK2, NRL*	Missense Variant, Intron Variant	0.289944	0.0457921	2.9E-11
rs58542926	19	19379549	T	C	0.07	*TM6SF2*	Missense Variant	0.127372	0.0219207	4.00E-09
rs4820268	22	37469591	G	A	0.46	*TMPRSS6*	Missense Variant	0.0542081	0.0115437	7.40E-07
rs738409	22	44324727	G	C	0.21	*PNPLA3*	Missense Variant	0.0559764	0.01399	1.10E-12
								cT1 (males) (N=6,640)		
								BETA	**SE**	**P**
rs759359281	1	220100497	C	CA	0.06	*SLC30A10*	Intron	0.156337	0.0391243	5.20E-05

rs13107325	4	103188709	T	C	0.07	*SLC39A8*	Missense Variant	0.56999	0.0331774	1.10E-66
rs111723834	14	24572932	A	G	0.02	*PCK2, NRL*	Missense Variant, Intron Variant	0.344737	0.0711227	4.50E-07
rs58542926	19	19379549	T	C	0.07	*TM6SF2*	Missense Variant	0.161972	0.0334273	4.80E-07
rs4820268	22	37469591	G	A	0.46	*TMPRSS6*	Missense Variant	0.0674141	0.017446	9.30E-05
rs738409	22	44324727	G	C	0.21	*PNPLA3*	Missense Variant	0.121098	0.0211575	7.60E-09

cT1 (females) (N=7,800)

								BETA	**SE**	**P**
rs759359281	1	220100497	C	CA	0.06	*SLC30A10*	Intron	0.124803	0.0351335	2.60E-04
rs13107325	4	103188709	T	C	0.07	*SLC39A8*	Missense Variant	0.551638	0.0311469	2.90E-70
rs111723834	14	24572932	A	G	0.02	*PCK2, NRL*	Missense Variant, Intron Variant	0.268713	0.0625125	8.90E-06
rs58542926	19	19379549	T	C	0.07	*TM6SF2*	Missense Variant	0.0943526	0.0303544	2.10E-03
rs4820268	22	37469591	G	A	0.46	*TMPRSS6*	Missense Variant	0.0701305	0.0160201	7.20E-06

rs738409	22	44324727	G	C	0.21	*PNPLA3*	Missense Variant	0.0770984	0.0195081	2.60E-05

Table 7.2.2. The association between four independent genetic variants and liver PDFF in 14,440 UK Biobank participants. A linear mixed model was used for genetic associations. Levels of significance: $p < 5 \times 10^{-8}$.

SNP	CHR	BP	EA	OA	EAF	BETA	SE	P	Gene	Consequence
rs1260326	2	27730940	T	C	0.39117	0.046	0.009	3.90E-08	*GCKR*	Missense Variant
rs58542926	19	19379549	T	C	0.926558	0.206	0.016	6.30E-37	*TM6SF2*	Missense Variant
rs429358	19	45411941	T	C	0.848199	0.073	0.011	5.60E-11	*APOE*	Missense Variant
rs738409	22	44324727	G	C	0.788816	0.173	0.010	5.40E-66	*PNPLA3*	Missense Variant

Table 7.2.3. Associations of cT1 variants with liver blood tests, liver fat PDFF, liver iron content, cardiometabolic traits and diseases. A linear mixed model was used for genetic associations. Levels of significance: p < 0.05.

SNP	Chr	Outcome	BETA	SE	P	N
rs759359281	1	liver fat %	0.0127587	0.0253171	6.00E-01	14440
rs759359281	1	liver iron content	-7.53E-03	3.42E-02	8.26E-01	8282
rs759359281	1	ALT	0.01971676	0.00492174	0.00006175	361940
rs759359281	1	AST	0.02097398	0.00501843	0.00002924	360731
rs759359281	1	GGT	0.01092781	0.00485173	0.02430062	361888
rs759359281	1	ALP	0.01569743	0.00512336	0.0021849	362087
rs759359281	1	BMI	0.00760543	0.00506501	0.13321097	378214
rs759359281	1	HDL	-0.0109753	0.00494659	0.0265036	331494
rs759359281	1	LDL	0.01033718	0.00514988	0.04472196	361392
rs759359281	1	Triglycerides	0.01511628	0.00507773	0.00291123	361785
rs759359281	1	SHBG	-0.00536504	0.00499736	0.28301357	328442

rs759359281	1	Systolic blood pressure	-0.00570476	0.00463114	0.21801416	378821
rs759359281	1	Diastolic blood pressure	-0.00812936	0.00493904	0.09977707	378197
rs759359281	1	Type 2 diabetes	0.98768539	0.0207639	0.55066752	379703
rs759359281	1	Heart disease	1.0155469	0.01793243	0.38962282	379706
rs759359281	1	Hypertension	0.98441287	0.0128509	0.2215286	379708
rs13107325	4	liver fat %	0.00785646	0.0220298	6.40E-01	14440
rs13107325	4	liver iron content	-8.15E-02	3.00E-02	6.57E-03	8282
rs13107325	4	ALT	0.02205415	0.0042059	1.58E-07	361940
rs13107325	4	AST	0.06215688	0.00428884	1.39E-47	360731
rs13107325	4	GGT	-0.0247514	0.0041454	2.36E-09	361888
rs13107325	4	ALP	0.00261341	0.00437779	0.55052791	362087
rs13107325	4	BMI	0.05491581	0.00432855	7.11E-37	378214
rs13107325	4	HDL	-0.08265373	0.00422418	3.32E-85	331494

rs13107325	4	LDL	-0.03573906	0.00440007	4.59E-16	361392
rs13107325	4	Triglycerides	0.03163324	0.00433898	3.10E-13	361785
rs13107325	4	SHBG	-0.00723772	0.00426942	0.09002952	328442
rs13107325	4	Systolic blood pressure	-0.03389149	0.00395976	1.14E-17	378821
rs13107325	4	Diastolic blood pressure	-0.04711799	0.00422344	6.74E-29	378197
rs13107325	4	Type 2 diabetes	1.0718579	0.01719493	0.00005444	379703
rs13107325	4	Heart disease	0.98825314	0.0153871	0.44252172	379706
rs13107325	4	Hypertension	0.96248733	0.01100757	0.00051379	379708
rs118045231	10	liver fat %	0.0818138	0.0352523	9.30E-03	14440
rs118045231	10	liver iron content	-5.24E-02	4.83E-02	2.78E-01	8282
rs118045231	10	ALT	0.00166892	0.00648689	0.79696626	361940
rs118045231	10	AST	0.00438794	0.00661627	0.50720039	360731
rs118045231	10	GGT	-0.00173244	0.00639243	0.78638002	361888

rs118045231	10	ALP	0.00236123	0.00675068	0.72650622	362087
rs118045231	10	BMI	0.0032314	0.00666841	0.62797191	378214
rs118045231	10	HDL	-0.01015797	0.00652252	0.11938393	331494
rs118045231	10	LDL	-0.00854632	0.00678571	0.20786529	361392
rs118045231	10	Triglycerides	0.01254281	0.00669167	0.06087712	361785
rs118045231	10	SHBG	-0.00827293	0.00659444	0.20965037	328442
rs118045231	10	Systolic blood pressure	0.00405816	0.00609824	0.50575431	378821
rs118045231	10	Diastolic blood pressure	0.00047661	0.00650397	0.94158377	378197
rs118045231	10	Type 2 diabetes	1.0283803	0.02690784	0.29832486	379703
rs118045231	10	Heart disease	0.97960143	0.02383456	0.38720926	379706
rs118045231	10	Hypertension	1.0222345	0.01679198	0.1903289	379708
rs111723834	14	liver fat %	-0.0240158	0.0455128	6.70E-01	14440
rs111723834	14	liver iron content	-2.40E-02	6.18E-02	6.98E-01	8282

rs111723834	14	ALT	-0.01873864	0.00869896	0.03123117	361940
rs111723834	14	AST	-0.01608606	0.00887277	0.06983735	360731
rs111723834	14	GGT	-0.02653084	0.00857517	0.0019755	361888
rs111723834	14	ALP	-0.02270907	0.00905475	0.01214307	362087
rs111723834	14	BMI	-0.01252042	0.0089537	0.16200836	378214
rs111723834	14	HDL	0.03186024	0.0087415	0.00026773	331494
rs111723834	14	LDL	-0.00258033	0.00910383	0.77684502	361392
rs111723834	14	Triglycerides	-0.03984006	0.00897427	9.03E-06	361785
rs111723834	14	SHBG	0.03205497	0.00882511	0.00028101	328442
rs111723834	14	Systolic blood pressure	-0.01311047	0.00818987	0.10941868	378821
rs111723834	14	Diastolic blood pressure	-0.02248176	0.008737	0.01007761	378197
rs111723834	14	Type 2 diabetes	0.91055144	0.0378863	0.01338647	379703
rs111723834	14	Heart disease	0.97187684	0.03207806	0.37385583	379706

rs111723834	14	Hypertension	0.92386097	0.0230377	0.00058696	379708
rs58542926	19	liver fat %	0.261101	0.0217732	8.10E-32	14440
rs58542926	19	liver iron content	5.22E-02	2.93E-02	7.48E-02	8282
rs58542926	19	ALT	0.08092395	0.00418902	4.17E-83	361940
rs58542926	19	AST	0.05932545	0.00427257	8.00E-44	360731
rs58542926	19	GGT	0.01636427	0.00413194	0.00007483	361888
rs58542926	19	ALP	-0.09005039	0.00436066	1.09E-94	362087
rs58542926	19	BMI	-0.00326425	0.00431546	0.4494056	378214
rs58542926	19	HDL	-0.00408918	0.00420784	0.33115097	331494
rs58542926	19	LDL	-0.13974327	0.00437875	3.51E-223	361392
rs58542926	19	Triglycerides	-0.10360269	0.00432086	6.02E-127	361785
rs58542926	19	SHBG	-0.00340488	0.00425019	0.42306731	328442
rs58542926	19	Systolic blood pressure	0.01470579	0.00394581	0.00019385	378821

rs58542926	19	Diastolic blood pressure	0.02296663	0.00420855	4.84E-08	378197
rs58542926	19	Type 2 diabetes	1.0846765	0.01710389	2.01E-06	379703
rs58542926	19	Heart disease	0.95968678	0.01551553	0.00799994	379706
rs58542926	19	Hypertension	1.024177	0.01084989	0.02767896	379708
rs4820268	22	liver fat %	0.0123012	0.0114339	3.20E-01	14440
rs4820268	22	liver iron content	-1.04E-01	1.55E-02	2.63E-11	8282
rs4820268	22	ALT	-0.0039429	0.00221457	0.07500625	361940
rs4820268	22	AST	-0.00979042	0.00225823	0.00001455	360731
rs4820268	22	GGT	-0.0024286	0.002183	0.26592274	361888
rs4820268	22	ALP	-0.00683427	0.00230523	0.00303023	362087
rs4820268	22	BMI	-0.00240988	0.00227925	0.29036959	378214
rs4820268	22	HDL	-0.00047719	0.00222443	0.83014014	331494
rs4820268	22	LDL	0.00764322	0.00231692	0.00097086	361392

rs4820268	22	Triglycerides	-0.00949943	0.00228489	0.00003218	361785
rs4820268	22	SHBG	-0.00500005	0.00224719	0.0260806	328442
rs4820268	22	Systolic blood pressure	0.00294821	0.00208454	0.15726715	378821
rs4820268	22	Diastolic blood pressure	-0.00187424	0.00222338	0.39924638	378197
rs4820268	22	Type 2 diabetes	1.000291	0.00930691	0.97502587	379703
rs4820268	22	Heart disease	1.019562	0.00810697	0.01686224	379706
rs4820268	22	Hypertension	1.012714	0.0057664	0.02845281	379708
rs738409	22	liver fat %	0.233979	0.0138955	1.20E-63	14440
rs738409	22	liver iron content	5.36E-02	1.89E-02	4.66E-03	8282
rs738409	22	ALT	0.09831972	0.00267411	2.18E-295	361940
rs738409	22	AST	0.09945885	0.00272673	9.25E-291	360731
rs738409	22	GGT	-0.00130433	0.00264091	0.62138227	361888
rs738409	22	ALP	-0.01233248	0.00278871	9.77E-06	362087

rs738409	22	BMI	-0.00899027	0.00275861	0.00111822	378214
rs738409	22	HDL	-0.01596964	0.00269179	2.98E-09	331494
rs738409	22	LDL	-0.01270562	0.00280268	5.81E-06	361392
rs738409	22	Triglycerides	-0.00292861	0.00276422	0.28938526	361785
rs738409	22	SHBG	0.04467104	0.00271863	1.20E-60	328442
rs738409	22	Systolic blood pressure	-0.00002693	0.00252259	0.99148313	378821
rs738409	22	Diastolic blood pressure	0.00254793	0.00269094	0.34371227	378197
rs738409	22	Type 2 diabetes	1.0346154	0.01116244	0.00229915	379703
rs738409	22	Heart disease	0.97555595	0.00985126	0.01200007	379706
rs738409	22	Hypertension	0.99394356	0.00697888	0.38404798	379708

Table 7.2.4. The association between cT1 variants and ALT / AST measures in Chambers *et al.*[182]

SNP	CHR	EA	OA	EAF	Gene	Beta ALT	SE ALT	P ALT	Beta AST	SE AST	P AST
rs759359281	1	C	CA	0.06	*SLC30A10*	NA	NA	NA	NA	NA	NA
rs13107325	4	T	C	0.07	*SLC39A8*	0.01	0.005	0.2745	0.0139	0.004	0.005
rs111723834	14	A	G	0.02	*PCK2, NRL*	NA	NA	NA	NA	NA	NA
rs10401969 (rs58542926 proxy)	19	C	T	0.88	*TM6SF2*	0.0111	0.0038	0.0003	0.0164	0.004	0.00004
rs4820268	22	G	A	0.46	*TMPRSS6*	0.0001	0.0016	0.515	0.0012	0.002	0.189
rs738409	22	G	C	0.21	*PNPLA3*	0.0254	0.0021	1.21E-45	0.0254	0.0023	1.23E-34

Table 7.2.5. Genetic correlation analyses between cT1 measures and 120 predominantly metabolic traits. For clarity, we present traits where p < 0.05 (t-test, levels of significance: p < 0.01).

Phenotype	PMID	rg	SE	P
HOMA-IR	20081858	0.5272	0.1489	0.0004
Mean diameter for VLDL particles	27005778	0.5207	0.146	0.0004
Glycoprotein acetyls; mainly a1-acid glycoprotein	27005778	0.5027	0.1566	0.0013
Triglycerides in small HDL	27005778	0.4957	0.1897	0.009
Leucine	27005778	0.4918	0.2024	0.0151

Tyrosine	27005778	0.4548	0.1755	0.0096
Isoleucine	27005778	0.4422	0.1724	0.0103
Fasting insulin main effect	22581228	0.4277	0.1158	0.0002
Phenylalanine	27005778	0.4266	0.2047	0.0371
Triglycerides in medium VLDL	27005778	0.4247	0.1547	0.0061
Type 2 Diabetes	22885922	0.4239	0.1043	4.79E-05
Triglycerides in small VLDL	27005778	0.4116	0.1423	0.0038
Concentration of medium VLDL particles	27005778	0.4041	0.1348	0.0027
Valine	27005778	0.4034	0.1829	0.0274
Waist circumference	25673412	0.3917	0.0699	2.09E-08
Body fat	26833246	0.388	0.0825	2.59E-06
Total lipids in medium VLDL	27005778	0.3876	0.138	0.005
Body mass index	20935630	0.3856	0.061	2.57E-10
HOMA-B	20081858	0.3842	0.1181	0.0011
Triglycerides in large VLDL	27005778	0.3841	0.1419	0.0068
Total lipids in large VLDL	27005778	0.3802	0.1393	0.0063

Concentration of very large VLDL particles	27005778	0.3752	0.1435	0.0089
Phospholipids in chylomicrons and largest VLDL particles	27005778	0.3727	0.1499	0.0129
Waist-to-hip ratio	25673412	0.3658	0.0749	1.05E-06
Total cholesterol in large VLDL	27005778	0.3603	0.1341	0.0072
Total lipids in very large VLDL	27005778	0.3552	0.1381	0.0101
Concentration of large VLDL particles	27005778	0.3501	0.1472	0.0174
Triglycerides in chylomicrons and largest VLDL particles	27005778	0.3464	0.1544	0.0249
Concentration of small VLDL particles	27005778	0.3443	0.1317	0.009
Serum total triglycerides	27005778	0.3422	0.1385	0.0135
Total lipids in chylomicrons and largest VLDL particles	27005778	0.3325	0.15	0.0266
Obesity class 2	23563607	0.3294	0.0746	1.01E-05
Phospholipids in large VLDL	27005778	0.3276	0.1399	0.0192
Free cholesterol in large VLDL	27005778	0.3249	0.1315	0.0135
Triglycerides in very large VLDL	27005778	0.3247	0.1299	0.0124
Phospholipids in medium VLDL	27005778	0.3244	0.1429	0.0232

Phospholipids in very large VLDL	27005778	0.3237	0.1428	0.0234
Total lipids in small VLDL	27005778	0.3235	0.1331	0.0151
Cholesterol esters in large VLDL	27005778	0.3149	0.1277	0.0137
Concentration of chylomicrons and largest VLDL particles	27005778	0.3128	0.1403	0.0257
Average number of methylene groups per a double bond	27005778	0.303	0.1268	0.0169
Phospholipids in small VLDL	27005778	0.2978	0.1374	0.0302
Free cholesterol in medium VLDL	27005778	0.2928	0.1409	0.0377
Total cholesterol in medium VLDL	27005778	0.2909	0.1358	0.0322
Hip circumference	25673412	0.2898	0.0616	2.56E-06
Obesity class 3	23563607	0.2765	0.1105	0.0123
HbA1C	20858683	0.275	0.1233	0.0258
Cholesterol esters in medium VLDL	27005778	0.2624	0.1325	0.0476
Crohns disease	26192919	0.1897	0.0854	0.0263
Coronary artery disease	26343387	0.1822	0.0669	0.0065
Triglycerides	20686565	0.1508	0.0762	0.0476

Ratio of bisallylic groups to double bonds	27005778	-0.2844	0.1073	0.008
Ratio of bisallylic groups to total fatty acids	27005778	-0.2893	0.117	0.0134
HDL cholesterol	20686565	-0.307	0.0686	7.58E-06
Average number of double bonds in a fatty acid chain	27005778	-0.3073	0.1307	0.0187
Citrate	27005778	-0.3293	0.1562	0.0351
Apolipoprotein A-I	27005778	-0.4093	0.1882	0.0296
Acetate	27005778	-0.4596	0.1846	0.0128
Total cholesterol in HDL	27005778	-0.4876	0.1594	0.0022
Total lipids in large HDL	27005778	-0.5206	0.1392	0.0002
Concentration of very large HDL particles	27005778	-0.5545	0.215	0.0099
Phospholipids in very large HDL	27005778	-0.5554	0.1692	0.001
Total lipids in very large HDL	27005778	-0.6045	0.2178	0.0055
Free cholesterol in very large HDL	27005778	-0.6146	0.2366	0.0094
Total cholesterol in very large HDL	27005778	-0.6232	0.3029	0.0396

Table 7.2.6. Mendelian randomisation sensitivity analyses. Egger test, weighted median (WM) and penalised weighted median (PWM) show similar directional effects with the IVW method. Levels of significance: p < 0.05.

Exposure	betaEgger	pEgger	betaWM	pWM	betaPWM	pPWM	N_snp	PMID
Favourable adiposity	-0.675	0.201	-0.328	0.149	-0.284	0.221	14	PMID: 30352878
Insulin resistance	0.327	0.337	0.512	0.001	0.488	0.001	53	PMID: 27841877
Insulin secretion	0.117	0.697	0.067	0.729	0.123	0.501	11	PMID: 22885924
Type 2 diabetes	0.072	0.1312	0.055	0.068	0.053	0.089	89	PMID: 29632382
2 hour Glucose	-0.055	0.903	0.015	0.851	-0.007	0.931	8	PMID: 22885924
Fasting glucose	0.0382	0.855	-0.176	0.221	-0.172	0.258	33	PMID: 22885924
Fasting Insulin	-4.555	0.014	0.792	0.014	0.838	0.008	14	PMID: 22885924
NAFLD	0.391	0.107	0.365	4.06E-11	-0.049	0.806	4	PMID: 21423719

Body fat percentage	1.173	0.160	0.430	0.008	0.435	0.007	10	PMID: 26833246
BMI	0.355	0.014	0.251	0.007	0.242	0.005	73	PMID: 25673413
Coronary artery disease	-0.025	0.635	-0.036	0.266	-0.045	0.202	64	PMID: 26343387
Systolic blood pressure	-0.046	0.090	-0.002	0.554	-0.011	0.055	89	PMID: 27618452
Diastolic blood pressure	-0.099	0.006	-0.012	0.087	-0.013	0.117	109	PMID: 27618452
LDL Cholesterol	-0.031	0.675	-0.035	0.504	-0.035	0.503	80	PMID: 24097068
Triglycerides	-0.127	0.170	-0.079	0.219	-0.067	0.424	62	PMID: 24097068
HDL Cholesterol	0.033	0.796	0.022	0.646	0.022	0.646	98	PMID: 24097068
Waist Hip ratio BMI adjusted	0.100	0.701	0.072	0.471	0.015	0.881	53	PMID: 30239722

Waist Hip ratio BMI adjusted (Female specific variants)	0.087	0.537	0.247	0.001	0.181	0.012	47	PMID: 30239722
Waist Hip ratio BMI adjusted (Male specific variants)	3.905	0.269	0.187	0.435	0.017	0.947	6	PMID: 30239722
Ferritin	-0.372	0.329	-0.567	8.35E-08	-0.100	0.521	6	PMID: 25352340
Transferrin	0.039	0.679	0.092	0.032	0.025	0.443	9	PMID: 25352340
Transferrin saturation	-0.228	0.029	-0.238	1.48E-11	-0.100	0.378	5	PMID: 25352340
Iron	-0.464	0.009	-0.332	1.72E-12	0.0117	0.942	5	PMID: 25352340
Height	-0.087	0.150	0.001	0.966	0.0124	0.6476	809	PMID: 30124842

7.3 Suspected liver fibrosis and incidence of 5 Cardiovascular diseases; a CALIBER study

7.3.1 Supplementary Figures

Figure 7.16. Hazard ratios for cardiovascular diseases with FIB-4 as a continuous variable (sex adjusted, fully adjusted, fully adjusted + age models).

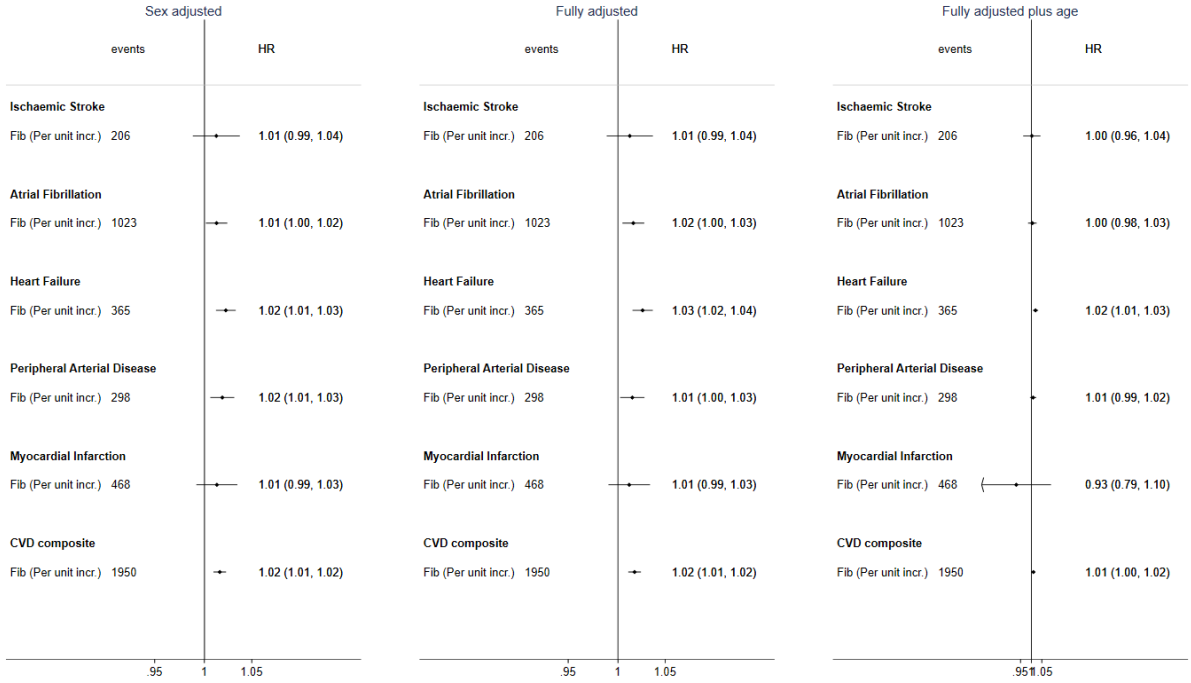
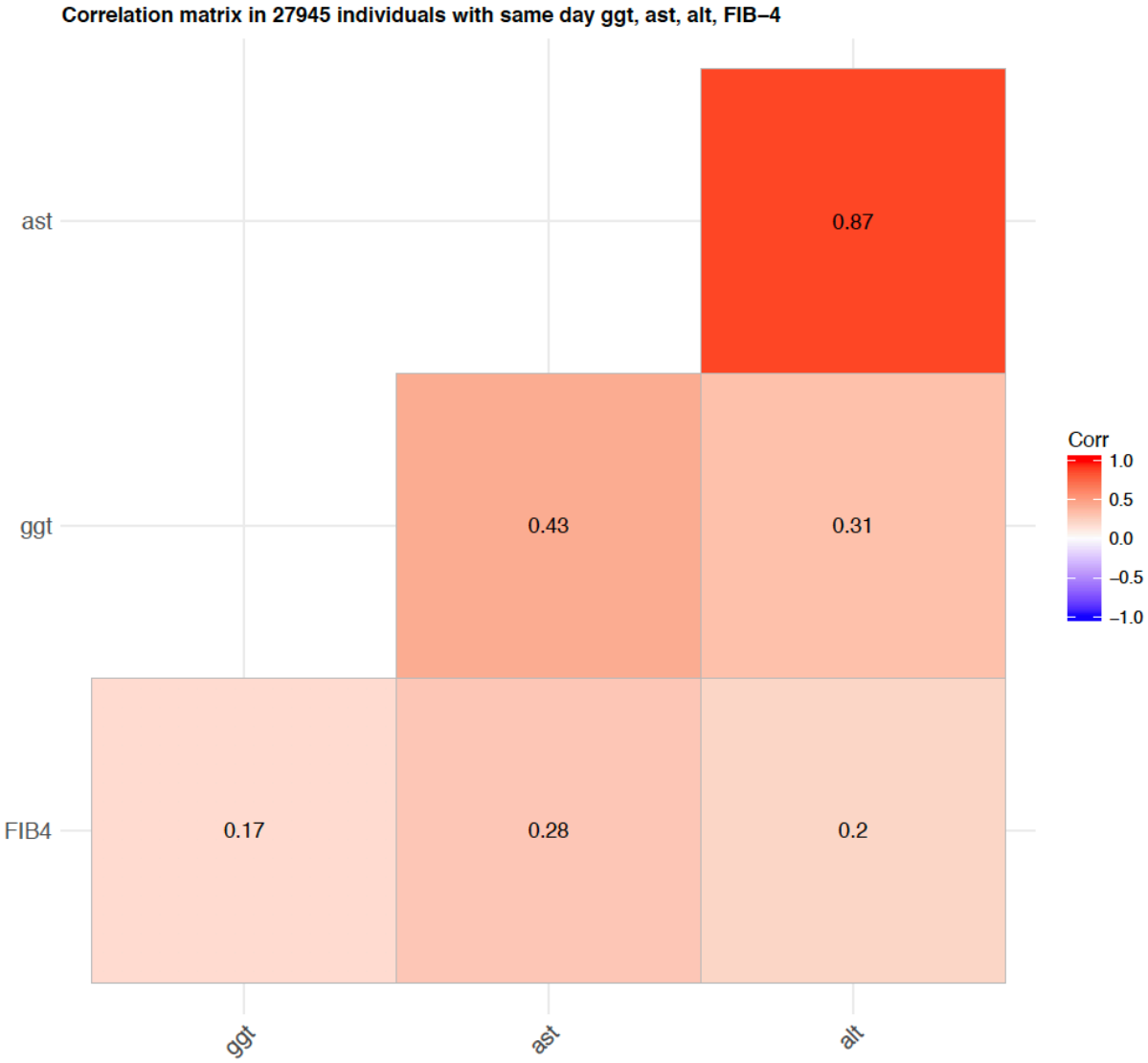


Figure 7.17. Correlation matrix of same day LBTs and FIB-4 index in 27,945 individuals.



7.3.2 Supplementary Tables

Table 7.3.1. Cox model hazard ratios of fully adjusted analysis (including correcting for age) for all cardiovascular outcomes investigated. We present the analysis when FIB-index was used with the binary cutoff of 1.3.

IS		1						
parm	estimate	HR	stderr	dof	t	p	min95	max95
0b.fib_category	0	1	0	0				
1.fib_category	0.1200565	1.127561	0.171405	22719234	0.700427	0.483661	-0.21589	0.456004
age	0.0804601	1.083786	0.009288	2393151	8.663073	4.60E-18	0.062256	0.098664
sex	0.8440683	2.32581	0.15601	34769.6	5.410347	6.33E-08	0.538284	1.149853
BMI_imputed	0.0276155	1.028	0.012415	4032.026	2.22434	2.62E-02	0.003275	0.051956
1b.alcohol	0		0	0				
2.alcohol	0.0586893	1.060446	0.597257	23.57506	0.098265	0.922553	-1.17517	1.292545
3.alcohol	-0.0271734	0.973192	0.226259	35.23298	-0.1201	0.905087	-0.48639	0.432048
4.alcohol	-0.2400959	0.786552	0.305638	55.87853	-0.78556	0.435445	-0.85239	0.3722
1b.sm_status	0		0	0				
2.sm_status	0.1425377	1.153197	0.224733	29862.75	0.634254	0.52592	-0.29795	0.583023
3.sm_status	0.8787279	2.407835	0.159088	8136.828	5.523543	3.43E-08	0.566875	1.190581
antihyper_drup	0.4347132	1.54452	0.2346	594642.7	1.852994	6.39E-02	-0.0251	0.894523
diab	0.9404789	2.561208	0.277622	691879.2	3.38762	7.05E-04	0.396348	1.48461
lipid_drugs	-0.1977698	0.820559	0.338937	1965281	-0.5835	0.559557	-0.86208	0.466536
AF								
parm	estimate	HR	stderr	dof	t	p	min95	max95
0b.fib_category	0	1	0	0				
1.fib_category	0.1559858	1.16881	0.076271	1837667	2.04515	0.04084	0.006497	0.305474
age	0.0821449	1.085613	0.004212	2806021	19.50259	1.06E-84	0.07389	0.0904
sex	0.5082517	1.662382	0.066879	41492	7.599587	3.03E-14	0.377168	0.639336
BMI_imputed	0.0444802	1.045484	0.005279	1267.887	8.425747	9.60E-17	0.034124	0.054837
1b.alcohol	0	1	0	0				
2.alcohol	0.1352816	1.144859	0.20645	98.4807	0.655275	0.513818	-0.27439	0.54495
3.alcohol	-0.0218362	0.9784	0.104962	31.08767	-0.20804	0.836556	-0.23588	0.192211
4.alcohol	-0.069794	0.932586	0.153722	29.18858	-0.45403	0.653168	-0.3841	0.244515
1b.sm_status	0	1	0	0				
2.sm_status	0.2115716	1.235618	0.087631	70273.98	2.414339	0.015766	0.039815	0.383329
3.sm_status	0.1966313	1.217295	0.078055	12638.48	2.519126	0.011777	0.043631	0.349632
antihyper_drup	0.499619	1.648093	0.107498	648568.7	4.647689	3.36E-06	0.288926	0.710312
diab	0.3543981	1.425322	0.156872	2445096	2.259161	0.023873	0.046935	0.661861
lipid_drugs	-0.3718825	0.689435	0.188565	12792806	-1.97217	0.04859	-0.74146	-0.0023
HF								
parm	estimate	HR	stderr	dof	t	p	min95	max95
0b.fib_category	0	1	0	0				
1.fib_category	0.0410796	1.041935	0.127689	4390634	0.321715	0.747669	-0.20919	0.291346
age	0.0965591	1.101375	0.007243	3184722	13.33186	1.52E-40	0.082364	0.110755
sex	0.5694143	1.767232	0.112559	18263.06	5.058786	4.26E-07	0.348787	0.790041
BMI_imputed	0.0541374	1.05563	0.008476	2525.26	6.38744	2.00E-10	0.037518	0.070757

1b.alcohol	0		0	0				
2.alcohol	0.0738349	1.076629	0.367216	47.64544	0.201067	0.841502	-0.66465	0.812315
3.alcohol	-0.1818151	0.833756	0.164864	39.01467	-1.10282	0.276861	-0.51528	0.151649
4.alcohol	-0.2109504	0.809814	0.254876	31.72387	-0.82766	0.414046	-0.73029	0.308392
1b.sm_status	0		0	0				
2.sm_status	0.0429988	1.043937	0.162632	70735.58	0.264394	0.791477	-0.27576	0.361757
3.sm_status	0.5555741	1.742941	0.120439	60754.56	4.612918	3.98E-06	0.319514	0.791634
antihyper_drup	0.2998915	1.349712	0.177717	531662.9	1.687464	0.091515	-0.04843	0.648212
diab	0.6331747	1.883581	0.220156	906621.4	2.876034	0.004027	0.201677	1.064672
lipid_drugs	0.1638472	1.178034	0.240054	30114267	0.682542	0.494896	-0.30665	0.634345
PAD								
parm	estimate	HR	stderr	dof	t	p	min95	max95
0b.fib_category	0		0	0				
1.fib_category	-0.0341698	0.966407	0.145605	2514076	-0.23467	0.814461	-0.31955	0.251211
age	0.0915633	1.095886	0.007732	4968169	11.84161	2.38E-32	0.076408	0.106718
sex	0.4786095	1.613829	0.123869	346930.8	3.863847	0.000112	0.235831	0.721388
BMI_imputed	0.0078007	1.007831	0.011137	1135.201	0.700447	0.483792	-0.01405	0.029652
1b.alcohol	0		0	0				
2.alcohol	-0.1208799	1.128489	0.369456	47.9494	0.327184	0.744953	-0.62198	0.863741
3.alcohol	-0.1117967	0.894226	0.173074	52.44048	-0.64595	0.521129	-0.45902	0.235431
4.alcohol	-0.1371848	0.871809	0.289487	28.82572	-0.47389	0.639148	-0.72941	0.455037
1b.sm_status	0		0	0				
2.sm_status	0.512184	1.668932	0.201013	31055.62	2.548019	0.010838	0.118191	0.906177
3.sm_status	1.6065898	4.98578	0.139532	88006.52	11.51417	1.18E-30	1.333109	1.88007
antihyper_drup	0.6833735	1.980548	0.191458	5346762	3.569307	0.000358	0.308122	1.058625
diab	1.1243733	3.078287	0.233389	50333.02	4.817595	1.46E-06	0.666928	1.581818
lipid_drugs	-0.7685571	0.463682	0.337695	2403619	-2.27589	0.022852	-1.43043	-0.10669
MI								
parm	estimate	HR	stderr	dof	t	p	min95	max95
0b.fib_category	0		0	0				
1.fib_category	-0.0637795	0.938212	0.114986	11018697	-0.55467	0.579119	-0.28915	0.161589
age	0.0849869	1.088703	0.006144	11386341	13.83154	1.65E-43	0.072944	0.09703
sex	1.1717493	3.227634	0.110541	17464.09	10.60014	3.57E-26	0.955078	1.388421
BMI_imputed	0.0174683	1.017622	0.009185	525.2148	1.90183	0.05774	-0.00058	0.035512
1b.alcohol	0		0	0				
2.alcohol	-0.1357645	1.145412	0.36542	22.34555	0.371529	0.713743	-0.62139	0.892921
3.alcohol	-0.1295563	0.878485	0.132872	65.0303	-0.97505	0.333149	-0.39492	0.135804
4.alcohol	-0.2514687	0.777658	0.223773	32.10187	-1.12377	0.269446	-0.70722	0.204286
1b.sm_status	0		0	0				
2.sm_status	0.3824087	1.465811	0.139086	19124.29	2.749448	0.005975	0.109789	0.655029
3.sm_status	-0.8259272	2.283997	0.106233	41228.54	7.774712	7.74E-15	0.617709	1.034145
antihyper_drup	-0.0625007	0.939412	0.185385	297691	-0.33714	0.736012	-0.42585	0.300849
diab	0.5534542	1.73925	0.206386	838420.9	2.681645	0.007326	0.148944	0.957964
lipid_drugs	0.0620562	1.064022	0.234587	1052080	0.264533	0.791369	-0.39773	0.521839
Composite		1						

parm	estimate	HR	stderr	dof	t	p	min95	max95
0b.fib_category	0	1	0	0				
1.fib_category	0.0993833	1.10449	0.055847	1244933	1.779556	0.075149	-0.01008	0.208842
age	0.0801184	1.083415	0.002992	1323049	26.77749	6.52E-158	0.074254	0.085983
sex	0.6626983	1.94002	0.049267	16382.83	13.45128	5.00E-41	0.56613	0.759266
BMI_imputed	0.0345385	1.035142	0.004136	563.5147	8.350565	5.29E-16	0.026415	0.042662
1b.alcohol	0	1	0	0				
2.alcohol	0.1310782	1.140057	0.165514	35.00772	0.791947	0.433725	-0.20493	0.467086
3.alcohol	-0.0526968	0.948668	0.070328	41.92714	-0.7493	0.457861	-0.19463	0.089238
4.alcohol	-0.1207581	0.886248	0.140527	17.45342	-0.85933	0.40181	-0.41666	0.175141
1b.sm_status	0	1	0	0				
2.sm_status	0.2702641	1.310311	0.066865	72850.41	4.041926	5.31E-05	0.139209	0.40132
3.sm_status	0.6403608	1.897165	0.052733	16539.36	12.14353	8.65E-34	0.536999	0.743723
antihyper_drup	0.4152654	1.514773	0.07963	309020.3	5.214948	1.84E-07	0.259193	0.571338
diab	0.5329172	1.703896	0.106301	371284.1	5.013264	5.35E-07	0.32457	0.741265
lipid_drugs	-0.2043653	0.815165	0.122455	4464658	-1.66891	0.095136	-0.44437	0.035642

Table 7.3.2. Phenotypes and codelists derived from primary (CPRD, Read codes), secondary (HES, ICD10) and ONS (ICD 10) datasets; the vast majority of these phenotypes are now publicly available through the HDR UK Phenotype Library.[219]

Phenotype	Link to phenotype codes designed and made publicly available
Cardiovascular Outcomes	
Myocardial infarction	https://phenotypes.healthdatagateway.org/phenotypes/PH215/version/430/detail/#home
Ischaemic Stroke	https://phenotypes.healthdatagateway.org/phenotypes/PH56/version/112/detail/#home
Atrial Fibrillation	https://phenotypes.healthdatagateway.org/phenotypes/PH36/version/72/detail/#home
Peripheral Arterial Disease	https://phenotypes.healthdatagateway.org/phenotypes/PH236/version/472/detail/#home
Heart Failure	https://phenotypes.healthdatagateway.org/phenotypes/PH182/version/364/detail/#home
Liver outcomes	
Liver fibrosis, sclerosis and cirrhosis	https://phenotypes.healthdatagateway.org/phenotypes/PH140/version/280/detail/#home, https://www.dropbox.com/s/aiv36014mkz8n6u/lfts_nafld_200517_Kenan_redo.xlsx?dl=0
Hepatocellulcar carcinoma	https://www.dropbox.com/s/aiv36014mkz8n6u/lfts_nafld_200517_Kenan_redo.xlsx?dl=0
Covariates	
Diabetes	https://phenotypes.healthdatagateway.org/phenotypes/PH152/version/304/detail/#home
Antihypertensive medications	https://www.caliberresearch.org/portal/show/hypertension_and_heart_failure_gprdprod
Lipid regulating drugs	https://www.caliberresearch.org/portal/show/lipid_regulating_drugs_gprdprod

Steatogenic medication	https://www.dropbox.com/s/aiv36014mkz8n6u/lfts_nafld_200517_Kenan_redo.xlsx?dl=0
Smoking history	https://www.caliberresearch.org/portal/show/Smoker
Alcohol consumption	https://phenotypes.healthdatagateway.org/phenotypes/PH9/version/1510/detail/#home
Exclusion criteria	
Acutely unwell state	https://www.dropbox.com/sh/u6iyxac8v9a8kmc/AABjA7P8kMB8RCXsYJOLnlxha?dl=0
Alcohol excess	https://phenotypes.healthdatagateway.org/phenotypes/PH94/version/188/detail/#home, https://www.dropbox.com/s/aiv36014mkz8n6u/lfts_nafld_200517_Kenan_redo.xlsx?dl=0
Any cancer	https://www.caliberresearch.org/portal/show/cancer_gprd, https://www.caliberresearch.org/portal/show/cancer_hes
Steatosis	https://www.dropbox.com/s/aiv36014mkz8n6u/lfts_nafld_200517_Kenan_redo.xlsx?dl=0
Hepatopancreatobiliary disease	https://www.dropbox.com/s/aiv36014mkz8n6u/lfts_nafld_200517_Kenan_redo.xlsx?dl=0
Acutely unwell state	https://www.dropbox.com/sh/u6iyxac8v9a8kmc/AABjA7P8kMB8RCXsYJOLnlxha?dl=0

8 References

[1] Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes. Hepatology. 2016;64:73–84.

[2] Targher G, Byrne CD, Lonardo A, Zoppini G, Barbui C. Non-alcoholic fatty liver disease and risk of incident cardiovascular disease: A meta-analysis. J. Hepatol. 2016;65:589–600.

[3] Adams LA, Anstee QM, Tilg H, Targher G. Non-alcoholic fatty liver disease and its relationship with cardiovascular disease and other extrahepatic diseases. Gut [Internet]. 2017;Available from: http://dx.doi.org/10.1136/gutjnl-2017-313884

[4] Anstee QM, Targher G, Day CP. Progression of NAFLD to diabetes mellitus, cardiovascular disease or cirrhosis. Nat. Rev. Gastroenterol. Hepatol. 2013;10:330–344.

[5] Liver cirrhosis. Lancet. 2021;398:1359–1376.

[6] Friedman SL, Neuschwander-Tetri BA, Rinella M, Sanyal AJ. Mechanisms of NAFLD development and therapeutic strategies. Nat. Med. 2018;24:908–922.

[7] Newsome PN, Cramb R, Davison SM, Dillon JF, Foulerton M, Godfrey EM, et al. Guidelines on the management of abnormal liver blood tests. Gut. 2018;67:6–19.

[8] Pavlides M, Banerjee R, Tunnicliffe EM, Kelly C, Collier J, Wang LM, et al. Multiparametric magnetic resonance imaging for the assessment of non-alcoholic fatty liver disease severity. Liver Int. 2017;37:1065–1073.

[9] Pavlides M, Banerjee R, Sellwood J, Kelly CJ, Robson MD, Booth JC, et al. Multiparametric magnetic resonance imaging predicts clinical outcomes in patients with chronic liver disease. J. Hepatol. 2016;64:308–315.

[10] Banerjee R, Pavlides M, Tunnicliffe EM, Piechnik SK, Sarania N, Philips R, et al. Multiparametric magnetic resonance for the non-invasive diagnosis of liver disease. J. Hepatol. 2014;60:69–77.

[11] Cheung A, Neuschwander-Tetri BA, Kleiner DE, Schabel E, Rinella M, Harrison S, et al. Defining Improvement in Nonalcoholic Steatohepatitis for Treatment Trial Endpoints: Recommendations from the Liver Forum. Hepatology [Internet]. 2019;Available from: http://dx.doi.org/10.1002/hep.30672

[12] Harrison SA, Rossi SJ, Paredes AH, Trotter JF, Bashir MR, Guy CD, et al. NGM282 Improves Liver Fibrosis and Histology in 12 Weeks in Patients with Nonalcoholic Steatohepatitis. Hepatology [Internet]. 2019;Available from: http://dx.doi.org/10.1002/hep.30590

[13] Elder E. NAFLD – diagnosis, assessment and management [Internet]. 2020 [cited 2021 Jun 2];Available from: https://www.bsg.org.uk/clinical-articles-list/nafld-diagnosis-assessment-and-management/

[14] Armstrong MJ, Houlihan DD, Bentham L, Shaw JC, Cramb R, Olliff S, et al. Presence and severity of non-alcoholic fatty liver disease in a large prospective primary care cohort. J. Hepatol. 2012;56:234–240.

[15] Whitehead MW, Hawkes ND, Hainsworth I, Kingham JG. A prospective study of the causes of notably raised aspartate aminotransferase of liver origin. Gut. 1999;45:129–133.

[16] Patel YA, Gifford EJ, Glass LM, Turner MJ, Han B, Moylan CA, et al. Identifying Nonalcoholic Fatty Liver Disease Advanced Fibrosis in the Veterans Health Administration. Dig. Dis. Sci. [Internet]. 2018;Available from: http://link.springer.com/10.1007/s10620-018-5123-3

[17] Singh S, Allen AM, Wang Z, Prokop LJ, Murad MH, Loomba R. Fibrosis progression in nonalcoholic fatty liver vs nonalcoholic steatohepatitis: a systematic review and meta-analysis of paired-biopsy studies. Clin. Gastroenterol. Hepatol. 2015;13:643-54.e1–9; quiz e39-40.

[18] White DL, Kanwal F, El-Serag HB. Association between nonalcoholic fatty liver disease and risk for hepatocellular cancer, based on systematic review. Clin. Gastroenterol. Hepatol. 2012;10:1342-1359.e2.

[19] Lorbeer R, Bayerl C, Auweter S, Rospleszcz S, Lieb W, Meisinger C, et al. Association between MRI-derived hepatic fat fraction and blood pressure in participants without history of cardiovascular disease. J. Hypertens. 2017;35:737–744.

[20] Sorrentino P, Terracciano L, D'Angelo S, Ferbo U, Bracigliano A, Vecchione R. Predicting fibrosis worsening in obese patients with NASH through parenchymal fibronectin, HOMA-IR, and hypertension. Am. J. Gastroenterol. 2010;105:336–344.

[21] Palmer ND, Musani SK, Yerges-Armstrong LM, Feitosa MF, Bielak LF, Hernaez R, et al. Characterization of European ancestry nonalcoholic fatty liver disease-associated variants in individuals of African and Hispanic descent. Hepatology. 2013;58:966–975.

[22] Rotman Y, Koh C, Zmuda JM, Kleiner DE, Liang TJ. The association of genetic variability in patatin-like phospholipase domain-containing protein 3 (PNPLA3) with histological severity of nonalcoholic fatty liver disease. Hepatology. 2010;52:894–903.

[23] Liu Y-L, Reeves HL, Burt AD, Tiniakos D, McPherson S, Leathart JBS, et al. TM6SF2 rs58542926 influences hepatic fibrosis progression in patients with non-alcoholic fatty liver disease. Nat. Commun. 2014;5:4309.

[24] Anstee QM, Darlay R, Cockell S, Meroni M, Govaere O, Tiniakos D, et al. Genome-wide association study of non-alcoholic fatty liver and steatohepatitis in a histologically-characterised cohort. J. Hepatol. [Internet]. 2020;Available from: http://dx.doi.org/10.1016/j.jhep.2020.04.003

[25] Abul-Husn NS, Cheng X, Li AH, Xin Y, Schurmann C, Stevis P, et al. A Protein-Truncating HSD17B13 Variant and Protection from Chronic Liver Disease. N. Engl. J.

Med. 2018;378:1096–1106.

[26] Mancina RM, Dongiovanni P, Petta S, Pingitore P, Meroni M, Rametta R, et al. The MBOAT7-TMC4 Variant rs641738 Increases Risk of Nonalcoholic Fatty Liver Disease in Individuals of European Descent. Gastroenterology. 2016;150:1219-1230.e6.

[27] Emdin CA, Haas M, Khera AV, Aragam K, Chaffin M, Jiang L, et al. A missense variant in Mitochondrial Amidoxime Reducing Component 1 gene and protection against liver disease [Internet]. bioRxiv. 2019 [cited 2019 May 13];594523. Available from: https://www.biorxiv.org/content/10.1101/594523v1

[28] Regeneron and Alnylam Pharmaceuticals Announce Collaboration to Discover New Treatments for Nonalcoholic Steat [Internet]. [cited 2021 Jun 2];Available from: https://investors.alnylam.com/press-release?id=22511

[29] Neuschwander-Tetri BA. Hepatic lipotoxicity and the pathogenesis of nonalcoholic steatohepatitis: the central role of nontriglyceride fatty acid metabolites. Hepatology. 2010;52:774–788.

[30] Cusi K. Role of obesity and lipotoxicity in the development of nonalcoholic steatohepatitis: pathophysiology and clinical implications. Gastroenterology. 2012;142:711-725.e6.

[31] Boursier J, Mueller O, Barret M, Machado M, Fizanne L, Araujo-Perez F, et al. The severity of nonalcoholic fatty liver disease is associated with gut dysbiosis and shift in the metabolic function of the gut microbiota. Hepatology. 2016;63:764–775.

[32] Hirsova P, Ibrahim SH, Gores GJ, Malhi H. Lipotoxic lethal and sublethal stress signaling in hepatocytes: relevance to NASH pathogenesis. J. Lipid Res. 2016;57:1758–1770.

[33] Mota M, Banini BA, Cazanave SC, Sanyal AJ. Molecular mechanisms of lipotoxicity and glucotoxicity in nonalcoholic fatty liver disease. Metabolism. 2016;65:1049–1061.

[34] Mohamed M, Phillips J. Hereditary haemochromatosis. BMJ. 2016;353:i3128.

[35] Taher AT, Weatherall DJ, Cappellini MD. Thalassaemia. Lancet. 2018;391:155–167.

[36] Dongiovanni P, Fracanzani AL, Fargion S, Valenti L. Iron in fatty liver and in the metabolic syndrome: a promising therapeutic target. J. Hepatol. 2011;55:920–932.

[37] Ehrmann DA. Polycystic ovary syndrome. N. Engl. J. Med. 2005;352:1223–1236.

[38] Deugnier Y, Bardou-Jacquet É, Lainé F. Dysmetabolic iron overload syndrome (DIOS). Presse Med. 2017;46:e306–e311.

[39] Nelson JE, Wilson L, Brunt EM, Yeh MM, Kleiner DE, Unalp-Arida A, et al. Relationship between the pattern of hepatic iron deposition and histological severity in nonalcoholic fatty liver disease. Hepatology. 2011;53:448–457.

[40] Valenti L, Fracanzani AL, Bugianesi E, Dongiovanni P, Galmozzi E, Vanni E, et al. HFE genotype, parenchymal iron accumulation, and liver fibrosis in patients with nonalcoholic fatty liver disease. Gastroenterology. 2010;138:905–912.

[41] Lee KS, Buck M, Houglum K, Chojkier M. Activation of hepatic stellate cells by TGF alpha and collagen type I is mediated by oxidative stress through c-myb expression. J. Clin. Invest. 1995;96:2461–2468.

[42] Gabrielsen JS, Gao Y, Simcox JA, Huang J, Thorup D, Jones D, et al. Adipocyte iron regulates adiponectin and insulin sensitivity. J. Clin. Invest. 2012;122:3529–3540.

[43] Gandon Y, Olivié D, Guyader D, Aubé C, Oberti F, Sebille V, et al. Non-invasive assessment of hepatic iron stores by MRI. Lancet. 2004;363:357–362.

[44] Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. PLoS Med. 2015;12:e1001779.

[45] Adams LA, Crawford DH, Stuart K, House MJ, St Pierre TG, Webb M, et al. The impact of phlebotomy in nonalcoholic fatty liver disease: A prospective, randomized, controlled trial. Hepatology. 2015;61:1555–1564.

[46] St Pierre TG, Clark PR, Chua-Anusorn W. Measurement and mapping of liver iron concentrations using magnetic resonance imaging. Ann. N. Y. Acad. Sci. 2005;1054:379–385.

[47] Ernst O, Rose C, Sergent G, L'Herminé C. Hepatic iron overload: quantification with MR imaging at 1.5 T. AJR Am. J. Roentgenol. 1999;172:1141–1142.

[48] EASL clinical practice guidelines for HFE hemochromatosis. J. Hepatol. 2010;53:3–22.

[49] Rowe IA, Parker R. The Placebo Response in Randomized Trials in Nonalcoholic Steatohepatitis Simply Explained. Clin. Gastroenterol. Hepatol. [Internet]. 2021;Available from: http://dx.doi.org/10.1016/j.cgh.2021.05.059

[50] Pavlides M, Birks J, Fryer E, Delaney D, Sarania N, Banerjee R, et al. Interobserver Variability in Histologic Evaluation of Liver Fibrosis Using Categorical and Quantitative Scores. Am. J. Clin. Pathol. 2017;147:364–369.

[51] Chalasani N, Younossi Z, Lavine JE, Diehl AM, Brunt EM, Cusi K, et al. The diagnosis and management of non-alcoholic fatty liver disease: practice guideline by the American Gastroenterological Association, American Association for the Study of Liver Diseases, and American College of Gastroenterology. Gastroenterology. 2012;142:1592–1609.

[52] Mofrad P, Contos MJ, Haque M, Sargeant C, Fisher RA, Luketic VA, et al. Clinical and histologic spectrum of nonalcoholic fatty liver disease associated with normal ALT values. Hepatology. 2003;37:1286–1292.

[53] Schwimmer JB, Celedon MA, Lavine JE, Salem R, Campbell N, Schork NJ, et al. Heritability of Nonalcoholic Fatty Liver Disease. Gastroenterology. 2009;136:1585–1592.

[54] Eslam M, Valenti L, Romeo S. Genetics and epigenetics of NAFLD and NASH: Clinical impact. J. Hepatol. 2018;68:268–279.

[55] Caussy C, Soni M, Cui J, Bettencourt R, Schork N, Chen C-H, et al. Nonalcoholic fatty liver disease with cirrhosis increases familial risk for advanced fibrosis. J. Clin. Invest. 2017;127:2697–2704.

[56] Loomba R, Schork N, Chen C-H, Bettencourt R, Bhatt A, Ang B, et al. Heritability of Hepatic Fibrosis and Steatosis Based on a Prospective Twin Study. Gastroenterology. 2015;149:1784–1793.

[57] Wagenknecht LE, Scherzinger AL, Stamm ER, Hanley AJG, Norris JM, Chen Y-DI, et al. Correlates and heritability of nonalcoholic fatty liver disease in a minority cohort. Obesity . 2009;17:1240–1246.

[58] Eslam M, George J. Genetic contributions to NAFLD: leveraging shared genetics to uncover systems biology. Nat. Rev. Gastroenterol. Hepatol. 2019;17:40–52.

[59] Romeo S, Kozlitina J, Xing C, Pertsemlidis A, Cox D, Pennacchio LA, et al. Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. Nat. Genet. 2008;40:1461–1465.

[60] Valenti L, Al-Serri A, Daly AK, Galmozzi E, Rametta R, Dongiovanni P, et al. Homozygosity for the patatin-like phospholipase-3/adiponutrin I148M polymorphism influences liver fibrosis in patients with nonalcoholic fatty liver disease. Hepatology. 2010;51:1209–1217.

[61] Speliotes EK, Yerges-Armstrong LM, Wu J, Hernaez R, Kim LJ, Palmer CD, et al. Genome-Wide Association Analysis Identifies Variants Associated with Nonalcoholic Fatty Liver Disease That Have Distinct Effects on Metabolic Traits. PLoS Genet. 2011;7:e1001324.

[62] Valenti L, Dongiovanni P, Ginanni Corradini S, Burza MA, Romeo S. PNPLA3 I148M variant and hepatocellular carcinoma: a common genetic variant for a rare disease. Dig. Liver Dis. 2013;45:619–624.

[63] Liu Y-L, Patman GL, Leathart JBS, Piguet A-C, Burt AD, Dufour J-F, et al. Carriage of the PNPLA3 rs738409 C >G polymorphism confers an increased risk of non-alcoholic fatty liver disease associated hepatocellular carcinoma. J. Hepatol. 2014;61:75–81.

[64] Lauridsen BK, Stender S, Kristensen T, Fuglsang KK, Kober L, Nordestgaard B, et al. Liver fat content, NAFLD, and ischemic heart disease: Mendelian randomization and meta-analysis of 279,013 individuals. Eur. Heart J. 2017;IN PRESS.

[65] Emdin CA, Haas ME, Khera AV, Aragam K, Chaffin M, Klarin D, et al. A missense variant in Mitochondrial Amidoxime Reducing Component 1 gene and protection against liver disease. PLoS Genet. 2020;16:e1008629.

[66] Mahdessian H, Taxiarchis A, Popov S, Silveira A, Franco-Cereceda A, Hamsten A, et al. TM6SF2 is a regulator of liver fat metabolism influencing triglyceride secretion and hepatic lipid droplet content. Proc. Natl. Acad. Sci. U. S. A. 2014;111:8913–8918.

[67] Li T-T, Li T-H, Peng J, He B, Liu L-S, Wei D-H, et al. TM6SF2: A novel target for plasma lipid regulation. Atherosclerosis. 2018;268:170–176.

[68] Kahali B, Liu Y-L, Daly AK, Day CP, Anstee QM, Speliotes EK. TM6SF2: Catch-22 in the Fight Against Nonalcoholic Fatty Liver Disease and Cardiovascular Disease? Gastroenterology. 2015;148:679–684.

[69] Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat. Genet. 2010;42:937–948.

[70] Beer NL, Tribble ND, McCulloch LJ, Roos C, Johnson PRV, Orho-Melander M, et al. The P446L variant in GCKR associated with fasting plasma glucose and triglyceride levels exerts its effect through increased glucokinase activity in liver. Hum. Mol. Genet. 2009;18:4081–4088.

[71] Teo K, Abeysekera KWM, Adams L, Aigner E, Anstee QM, Banales JM, et al. rs641738C>T near MBOAT7 is associated with liver fat, ALT and fibrosis in NAFLD: A meta-analysis. J. Hepatol. 2021;74:20–30.

[72] Zhong G, Kirkwood J, Won K-J, Tjota N, Jeong H, Isoherranen N. Characterization of vitamin A metabolome in human livers with and without nonalcoholic fatty liver disease. J. Pharmacol. Exp. Ther. 2019;370:92–103.

[73] Bell S, Rigas AS, Magnusson MK, Ferkingstad E, Allara E, Bjornsdottir G, et al. A genome-wide meta-analysis yields 46 new loci associating with biomarkers of iron homeostasis. Commun. Biol. 2021;4:156.

[74] Benyamin B, Esko T, Ried JS, Radhakrishnan A, Vermeulen SH, Traglia M, et al. Novel loci affecting iron homeostasis and their effects in individuals at risk for hemochromatosis. Nat. Commun. 2014;5:4926.

[75] Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ. 2018;362:k601.

[76] Hingorani A, Humphries S. Nature's randomised trials. Lancet. 2005;366:1906–1908.

[77] Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. Int. J. Epidemiol. 2016;45:1866–1886.

[78] Pierce BL, Burgess S. Efficient Design for Mendelian Randomization Studies: Subsample and 2-Sample Instrumental Variable Estimators. Am. J. Epidemiol. 2013;178:1177–1184.

[79] Chen L, Davey Smith G, Harbord RM, Lewis SJ. Alcohol Intake and Blood Pressure: A Systematic Review Implementing a Mendelian Randomization Approach. PLoS Med. 2008;5:e52.

[80] Bastarache L, Denny JC, Roden DM. Phenome-Wide Association Studies. JAMA. 2022;327:75–76.

[81] Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife [Internet]. 2018;7. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5976434

[82] Null N, Aguet F, Anand S, Ardlie KG, Gabriel S, Getz GA, et al. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science. 2020;369:1318–1330.

[83] Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat. Commun. 2017;8:1826.

[84] GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. Nat. Genet. 2013;45:580–585.

[85] Targher G, Day CP, Bonora E. Risk of cardiovascular disease in patients with nonalcoholic fatty liver disease. N. Engl. J. Med. 2010;363:1341–1350.

[86] Gastaldelli A. Fatty liver disease: the hepatic manifestation of metabolic syndrome. Hypertens. Res. 2010;33:546–547.

[87] Kotronen A, Yki-Järvinen H. Fatty liver: a novel component of the metabolic syndrome. Arterioscler. Thromb. Vasc. Biol. 2008;28:27–38.

[88] Lonardo A, Nascimbeni F, Mantovani A, Targher G. Hypertension, diabetes, atherosclerosis and NASH: Cause or consequence? J. Hepatol. 2018;68:335–352.

[89] Anstee QM, Mantovani A, Tilg H, Targher G. Risk of cardiomyopathy and cardiac arrhythmias in patients with nonalcoholic fatty liver disease. Nat. Rev. Gastroenterol. Hepatol. 2018;1.

[90] Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. BMJ. 2017;357:j2099.

[91] Targher G, Chonchol M, Miele L, Zoppini G, Pichiri I, Muggeo M. Nonalcoholic Fatty Liver Disease as a Contributor to Hypercoagulation and Thrombophilia in the Metabolic Syndrome. Semin. Thromb. Hemost. 2009;35:277–287.

[92] Hagström H, Nasr P, Ekstedt M, Hammar U, Stål P, Hultcrantz R, et al. Fibrosis stage but not NASH predicts mortality and time to development of severe liver disease in biopsy-proven NAFLD. J. Hepatol. 2017;67:1265–1273.

[93] Taylor RS, Taylor RJ, Bayliss S, Hagström H, Nasr P, Schattenberg JM, et al. Association between fibrosis stage and outcomes of patients with nonalcoholic fatty liver disease: A systematic review and meta-analysis. Gastroenterology. 2020;158:1611-1625.e12.

[94] Sinn DH, Kang D, Chang Y, Ryu S, Cho SJ, Paik SW, et al. Non-alcoholic fatty liver disease and the incidence of myocardial infarction: A cohort study. J. Gastroenterol. Hepatol. 2020;35:833–839.

[95] Labenz C, Huber Y, Michel M, Nagel M, Galle PR, Kostev K, et al. Impact of NAFLD on the incidence of cardiovascular diseases in a primary care population in Germany. Dig. Dis. Sci. 2020;65:2112–2119.

[96] Zhou Y-Y, Zhou X-D, Wu S-J, Fan D-H, Van Poucke S, Chen Y-P, et al. Nonalcoholic fatty liver disease contributes to subclinical atherosclerosis: A systematic review and meta-analysis. Hepatol. Commun. 2018;2:376–392.

[97] Chang Y, Ryu S, Sung K-C, Cho YK, Sung E, Kim H-N, et al. Alcoholic and non-alcoholic fatty liver disease and associations with coronary artery calcification: evidence from the Kangbuk Samsung Health Study. Gut. 2019;68:1667–1675.

[98] Henson JB, Simon TG, Kaplan A, Osganian S, Masia R, Corey KE. Advanced fibrosis is associated with incident cardiovascular disease in patients with non-alcoholic fatty liver disease. Aliment. Pharmacol. Ther. 2020;51:728–736.

[99] Alexander M, Loomis AK, Van Der Lei J, Duarte-Salles T, Prieto-Alhambra D, Ansell D, et al. Non-alcoholic fatty liver disease and risk of incident acute myocardial infarction and stroke: Findings from matched cohort study of 18 million European adults. BMJ [Internet]. 2019;Available from: http://dx.doi.org/10.1136/bmj.l5367

[100] Sanyal AJ, Van Natta ML, Clark J, Neuschwander-Tetri BA, Diehl A, Dasarathy S, et al. Prospective Study of Outcomes in Adults with Nonalcoholic Fatty Liver Disease. N. Engl. J. Med. 2021;385:1559–1569.

[101] Fraser A, Harris R, Sattar N, Ebrahim S, Smith GD, Lawlor DA. Gamma-Glutamyltransferase Is Associated With Incident Vascular Events Independently of Alcohol Intake: Analysis of the British Women's Heart and Health Study and Meta-Analysis. Arterioscler. Thromb. Vasc. Biol. 2007;27:2729–2735.

[102] Kunutsor SK, Apekey TA, Khan H. Liver enzymes and risk of cardiovascular disease in the general population: A meta-analysis of prospective cohort studies. Atherosclerosis. 2014;236:7–17.

[103] Ghouri N, Preiss D, Sattar N. Liver enzymes, nonalcoholic fatty liver disease, and incident cardiovascular disease: A narrative review and clinical perspective of prospective

data. Hepatology. 2010;52:1156–1161.

[104] Liu J, Au Yeung SL, Lin SL, Leung GM, Schooling CM. Liver Enzymes and Risk of Ischemic Heart Disease and Type 2 Diabetes Mellitus: A Mendelian Randomization Study. Sci. Rep. 2016;6:38813.

[105] So-Armah KA, Lim JK, Lo Re V, Tate JP, Chung-Chou HC, Butt AA, et al. FIB-4 stage of liver fibrosis predicts incident heart failure among HIV infected and uninfected patients. Hepatology [Internet]. 2017;Available from: http://doi.wiley.com/10.1002/hep.29285

[106] Saito Y, Okumura Y, Nagashima K, Fukamachi D, Yokoyama K, Matsumoto N, et al. Impact of the Fibrosis-4 Index on Risk Stratification of Cardiovascular Events and Mortality in Patients with Atrial Fibrillation: Findings from a Japanese Multicenter Registry. J. Clin. Med. Res. [Internet]. 2020;9. Available from: http://dx.doi.org/10.3390/jcm9020584

[107] Lee J, Kim HS, Cho YK, Kim EH, Lee MJ, Bae IY, et al. Association between noninvasive assessment of liver fibrosis and coronary artery calcification progression in patients with nonalcoholic fatty liver disease. Sci. Rep. 2020;10:18323.

[108] Akuta N, Kawamura Y, Arase Y, Saitoh S, Fujiyama S, Sezaki H, et al. PNPLA3 genotype and fibrosis-4 index predict cardiovascular diseases of Japanese patients with histopathologically-confirmed NAFLD. BMC Gastroenterol. 2021;21:434.

[109] Hallsworth K, Hollingsworth KG, Thoma C, Jakovljevic D, MacGowan GA, Anstee QM, et al. Cardiac structure and function are altered in adults with non-alcoholic fatty liver disease. J. Hepatol. 2013;58:757–762.

[110] Fricker ZP, Pedley A, Massaro JM, Vasan RS, Hoffmann U, Benjamin EJ, et al. Liver fat is associated with markers of inflammation and oxidative stress in analysis of data from the Framingham Heart Study. Clin. Gastroenterol. Hepatol. 2019;17:1157-1164.e4.

[111] Hansson GK. Inflammation, Atherosclerosis, and Coronary Artery Disease. N. Engl. J. Med. 2005;352:1685–1695.

[112] The interleukin-6 receptor as a target for prevention of coronary heart disease: a mendelian randomisation analysis. Lancet. 2012;379:1214–1224.

[113] Harrison SC, Smith AJP, Jones GT, Swerdlow DI, Rampuri R, Bown MJ, et al. Interleukin-6 receptor pathways in abdominal aortic aneurysm. Eur. Heart J. 2013;34:3707–3716.

[114] Ferreira RC, Freitag DF, Cutler AJ, Howson JMM, Rainbow DB, Smyth DJ, et al. Functional IL6R 358Ala Allele Impairs Classical IL-6 Receptor Signaling and Influences Risk of Diverse Inflammatory Diseases. PLoS Genet. 2013;9:e1003444.

[115] Ridker PM, Everett BM, Thuren T, MacFadyen JG, Chang WH, Ballantyne C, et al.

Antiinflammatory Therapy with Canakinumab for Atherosclerotic Disease. N. Engl. J. Med. 2017;377:1119–1131.

[116] Li Z, Yang S, Lin H, Huang J, Watkins PA, Moser AB, et al. Probiotics and antibodies to TNF inhibit inflammatory activity and improve nonalcoholic fatty liver disease. Hepatology. 2003;37:343–350.

[117] Caussy C, Tripathi A, Humphrey G, Bassirian S, Singh S, Faulkner C, et al. A gut microbiome signature for cirrhosis due to nonalcoholic fatty liver disease. Nat. Commun. 2019;10:1406.

[118] Zhang Y, Xu J, Wang X, Ren X, Liu Y. Changes of intestinal bacterial microbiota in coronary heart disease complicated with nonalcoholic fatty liver disease. BMC Genomics. 2019;20:862.

[119] Staley JR, Blackshaw J, Kamat MA, Ellis S, Surendran P, Sun BB, et al. PhenoScanner: a database of human genotype–phenotype associations. Bioinformatics. 2016;32:3207–3209.

[120] Denaxas SC, George J, Herrett E, Shah AD, Kalra D, Hingorani AD, et al. Data Resource Profile: Cardiovascular disease research using linked bespoke studies and electronic health records (CALIBER). Int. J. Epidemiol. 2012;41:1625–1638.

[121] Gallagher AM, Puri S, van Staa TP. Linkage of the General Practice Research Database (GPRD) with other data sources. In: Pharmacoepidemiology and drug safety. WILEY-BLACKWELL COMMERCE PLACE, 350 MAIN ST, MALDEN 02148, MA USA; 2011. p. S230–S231.

[122] O'Neil M, Payne C, Read J. Read Codes Version 3: a user led terminology. Methods Inf. Med. 1995;34:187–192.

[123] George J, Mathur R, Shah AD, Pujades-Rodriguez M, Denaxas S, Smeeth L, et al. Ethnicity and the first diagnosis of a wide range of cardiovascular diseases: Associations in a linked electronic health record cohort of 1 million patients. PLoS One. 2017;12:e0178945.

[124] Bhaskaran K, Forbes HJ, Douglas I, Leon DA, Smeeth L. Representativeness and optimal use of body mass index (BMI) in the UK Clinical Practice Research Datalink (CPRD). BMJ Open. 2013;3:e003389.

[125] Mathur R, Bhaskaran K, Chaturvedi N, Leon DA, vanStaa T, Grundy E, et al. Completeness and usability of ethnicity data in UK-based primary care and hospital databases. J. Public Health . 2014;36:684–692.

[126] Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. Br. J. Clin. Pharmacol. 2010;69:4–14.

[127] Herbert A, Wijlaars L, Zylbersztejn A, Cromwell D, Hardelid P. Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC). Int. J. Epidemiol. 2017;46:1093–1093i.

[128] Herrett E, Smeeth L, Walker L, Weston C, MINAP Academic Group. The Myocardial Ischaemia National Audit Project (MINAP). Heart. 2010;96:1264–1267.

[129] Jordan H, Roderick P, Martin D. The Index of Multiple Deprivation 2000 and accessibility effects on health. J. Epidemiol. Community Health. 2004;58:250–257.

[130] Denaxas S, Gonzalez-Izquierdo A, Direk K, Fitzpatrick NK, Fatemifar G, Banerjee A, et al. UK phenomics platform for developing and validating electronic health record phenotypes: CALIBER. J. Am. Med. Inform. Assoc. 2019;26:1545–1559.

[131] Williams R, Aspinall R, Bellis M, Camps-Walsh G, Cramp M, Dhawan A, et al. The Lancet Commissions Addressing liver disease in the UK: a blueprint for attaining excellence in health care and reducing premature mortality from lifestyle issues of excess consumption of alcohol, obesity, and viral hepatitis The Lancet Commissions. 2014;Available from: http://dx.doi.org/10.1016/

[132] Rinella ME. Nonalcoholic fatty liver disease: a systematic review. JAMA. 2015;313:2263–2273.

[133] Pais R, Charlotte F, Fedchuk L, Bedossa P, Lebray P, Poynard T, et al. A systematic review of follow-up biopsies reveals disease progression in patients with non-alcoholic fatty liver. J. Hepatol. 2013;59:550–556.

[134] Sorrentino P, D'Angelo S, Ferbo U, Micheli P, Bracigliano A, Vecchione R. Liver iron excess in patients with hepatocellular carcinoma developed on non-alcoholic steato-hepatitis. J. Hepatol. 2009;50:351–357.

[135] Koivula RW, Heggie A, Barnett A, Cederberg H, Hansen TH, Koopman AD, et al. Discovery of biomarkers for glycaemic deterioration before and after the onset of type 2 diabetes: rationale and design of the epidemiological studies within the IMI DIRECT Consortium. Diabetologia. 2014;57:1132–1142.

[136] McKay A, Wilman HR, Dennis A, Kelly M, Gyngell ML, Neubauer S, et al. Measurement of liver iron by magnetic resonance imaging in the UK Biobank population. PLoS One. 2018;13:e0209340.

[137] Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat. Genet. 2012;44:821–824.

[138] Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7.

[139] Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PC, et al. LD Hub: a centralized database and web interface to perform LD score regression that

maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics. 2017;33:272–279.

[140] de Leeuw CA, Stringer S, Dekkers IA, Heskes T, Posthuma D. Conditional and interaction gene-set analysis reveals novel functional pathways for blood pressure. Nat. Commun. 2018;9:3768.

[141] Koivula RW, Forgie IM, Kurbasic A, Viñuela A, Heggie A, Giordano GN, et al. Discovery of biomarkers for glycaemic deterioration before and after the onset of type 2 diabetes: an overview of the data from the epidemiological studies within the IMI DIRECT Consortium [Internet]. bioRxiv. 2018 [cited 2019 Feb 7];300244. Available from: https://www.biorxiv.org/content/10.1101/300244v1

[142] Dongiovanni P, Lanti C, Gatti S, Rametta R, Recalcati S, Maggioni M, et al. High fat diet subverts hepatocellular iron uptake determining dysmetabolic iron overload. PLoS One. 2015;10:e0116855.

[143] Paisant A, d'Assignies G, Bannier E, Bardou-Jacquet E, Gandon Y. MRI for the measurement of liver iron content, and for the diagnosis and follow-up of iron overload disorders. Presse Med. 2017;46:e279–e287.

[144] Bowden J, Del Greco M F, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic. Int. J. Epidemiol. 2016;45:1961–1974.

[145] Burgess S, Bowden J, Fall T, Ingelsson E, Thompson SG. Sensitivity Analyses for Robust Causal Inference from Mendelian Randomization Analyses with Multiple Genetic Variants. Epidemiology. 2017;28:30–42.

[146] Rees JMB, Wood AM, Burgess S. Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured pleiotropy. Stat. Med. [Internet]. 2017;Available from: http://doi.wiley.com/10.1002/sim.7492

[147] Canela-Xandri O, Rawlik K, Tenesa A. An atlas of genetic associations in UK Biobank. Nat. Genet. 2018;50:1593–1599.

[148] Nuttall KL, Palaty J, Lockitch G. Reference limits for copper and iron in liver biopsies. Ann. Clin. Lab. Sci. 2003;33:443–450.

[149] Nikpay M, Goel A, Won H-H, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat. Genet. 2015;47:1121–1130.

[150] Paternoster L, Standl M, Waage J, Baurecht H, Hotze M, Strachan DP, et al. Multi-ancestry genome-wide association study of 21,000 cases and 95,000 controls identifies new risk loci for atopic dermatitis. Nat. Genet. 2015;47:1449–1456.

[151] Lu Y, Day FR, Gustafsson S, Buchkovich ML, Na J, Bataille V, et al. New loci for body fat percentage reveal link between adiposity and cardiometabolic disease risk. Nat. Commun. 2016;7:10495.

[152] Nemeth E, Tuttle MS, Powelson J, Vaughn MB, Donovan A, Ward DM, et al. Hepcidin regulates cellular iron efflux by binding to ferroportin and inducing its internalization. Science. 2004;306:2090–2093.

[153] Nai A, Pagani A, Silvestri L, Campostrini N, Corbella M, Girelli D, et al. TMPRSS6 rs855791 modulates hepcidin transcription in vitro and serum hepcidin levels in normal individuals. Blood. 2011;118:4459–4462.

[154] Barton JC, Edwards CQ, Acton RT. HFE gene: Structure, function, mutations, and associated iron abnormalities. Gene. 2015;574:179–192.

[155] Lumeng CN, Bodzin JL, Saltiel AR. Obesity induces a phenotypic switch in adipose tissue macrophage polarization. J. Clin. Invest. 2007;117:175–184.

[156] Orr JS, Kennedy A, Anderson-Baucum EK, Webb CD, Fordahl SC, Erikson KM, et al. Obesity alters adipose tissue macrophage iron content and tissue iron distribution. Diabetes. 2014;63:421–432.

[157] Khan FA, Fisher MA, Khakoo RA. Association of hemochromatosis with infectious diseases: expanding spectrum. Int. J. Infect. Dis. 2007;11:482–487.

[158] Fonseca-Nunes A, Jakszyn P, Agudo A. Iron and cancer risk--a systematic review and meta-analysis of the epidemiological evidence. Cancer Epidemiol. Biomarkers Prev. 2014;23:12–31.

[159] Lv Y-F, Chang X, Hua R-X, Yan G-N, Meng G, Liao X-Y, et al. The risk of new-onset cancer associated with HFE C282Y and H63D mutations: evidence from 87,028 participants. J. Cell. Mol. Med. 2016;20:1219–1233.

[160] Elliott LT, Sharp K, Alfaro-Almagro F, Shi S, Miller KL, Douaud G, et al. Genome-wide association studies of brain imaging phenotypes in UK Biobank. Nature. 2018;562:210–216.

[161] Brar S, Henderson D, Schenck J, Zimmerman EA. Iron accumulation in the substantia nigra of patients with Alzheimer disease and parkinsonism. Arch. Neurol. 2009;66:371–374.

[162] Burgetova A, Seidl Z, Krasensky J, Horakova D, Vaneckova M. Multiple sclerosis and the accumulation of iron in the Basal Ganglia: quantitative assessment of brain iron using MRI t(2) relaxometry. Eur. Neurol. 2010;63:136–143.

[163] The Lancet Gastroenterology Hepatology. Ironing out unmet need in genetic haemochromatosis. Lancet Gastroenterol Hepatol. 2019;4:1.

[164] Ellervik C, Tybjaerg-Hansen A, Appleyard M, Ibsen H, Nordestgaard BG. Haemochromatosis genotype and iron overload: association with hypertension and left ventricular hypertrophy. J. Intern. Med. 2010;268:252–264.

[165] Cai H. Hydrogen peroxide regulation of endothelial function: origins, mechanisms, and consequences. Cardiovasc. Res. 2005;68:26–36.

[166] Pilling LC, Tamosauskaite J, Jones G, Wood AR, Jones L, Kuo C-L, et al. Common conditions associated with hereditary haemochromatosis genetic variants: cohort study in UK Biobank. BMJ. 2019;364:k5222.

[167] Gill D, Del Greco M F, Walker AP, Srai SKS, Laffan MA, Minelli C. The Effect of Iron Status on Risk of Coronary Artery Disease: A Mendelian Randomization Study-Brief Report. Arterioscler. Thromb. Vasc. Biol. 2017;37:1788–1792.

[168] Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. Am. J. Epidemiol. 2017;186:1026–1034.

[169] Gilmore I, Williams R. Alcohol policy in the UK: where next? Lancet [Internet]. 2019;Available from: https://doi.org/10.1016/S0140-6736(19)31396-0

[170] Younossi Z, Tacke F, Arrese M, Sharma BC, Mostafa I, Bugianesi E, et al. Global Perspectives on Non-alcoholic Fatty Liver Disease and Non-alcoholic Steatohepatitis. Hepatology [Internet]. 2018;Available from: http://dx.doi.org/10.1002/hep.30251

[171] Harris R, Harman DJ, Card TR, Aithal GP, Guha IN. Prevalence of clinically significant liver disease within the general population, as defined by non-invasive markers of liver fibrosis: a systematic review. The Lancet Gastroenterology & Hepatology. 2017;2:288–297.

[172] Day CP, Saksena S. Non-alcoholic steatohepatitis: definitions and pathogenesis. J. Gastroenterol. Hepatol. 2002;17 Suppl 3:S377-84.

[173] Williams CD, Stengel J, Asike MI, Torres DM, Shaw J, Contreras M, et al. Prevalence of nonalcoholic fatty liver disease and nonalcoholic steatohepatitis among a largely middle-aged population utilizing ultrasound and liver biopsy: a prospective study. Gastroenterology. 2011;140:124–131.

[174] Mojtahed A, Kelly CJ, Herlihy AH, Kin S, Wilman HR, McKay A, et al. Reference range of liver corrected T1 values in a population at low risk for fatty liver disease-a UK Biobank sub-study, with an appendix of interesting cases. Abdom Radiol (NY). 2019;44:72–84.

[175] Petersen SE, Matthews PM, Bamberg F, Bluemke DA, Francis JM, Friedrich MG, et al. Imaging in population science: cardiovascular magnetic resonance in 100,000 participants of UK Biobank - rationale, challenges and approaches. J. Cardiovasc. Magn. Reson. 2013;15:46.

[176] Wilman HR, Kelly M, Garratt S, Matthews PM, Milanesi M, Herlihy A, et al. Characterisation of liver fat in the UK Biobank cohort. PLoS One. 2017;12:e0172921.

[177] Wilman HR, Parisinos CA, Atabaki-Pasdar N, Kelly M, Louise Thomas E, Neubauer S, et al. Genetic studies of abdominal MRI data identify genes regulating hepcidin as major determinants of liver iron concentration. J. Hepatol. [Internet]. 2019;Available from: http://www.sciencedirect.com/science/article/pii/S0168827819303538

[178] Irving B, Hutton C, Dennis A, Vikal S, Mavar M, Kelly M, et al. Deep Quantitative Liver Segmentation and Vessel Exclusion to Assist in Liver Assessment. In: Medical Image Understanding and Analysis. Springer International Publishing; 2017. p. 663–673.

[179] Wood JC, Enriquez C, Ghugre N, Tyzka JM, Carson S, Nelson MD, et al. MRI R2 and R2* mapping accurately estimates hepatic iron concentration in transfusion-dependent thalassemia and sickle cell disease patients. Blood. 2005;106:1460–1465.

[180] Mojtahed A, Gee MS. Magnetic resonance enterography evaluation of Crohn disease activity and mucosal healing in young patients. Pediatr. Radiol. 2018;48:1273–1279.

[181] Frayling TM, Beaumont RN, Jones SE, Yaghootkar H, Tuke MA, Ruth KS, et al. A Common Allele in FGF21 Associated with Sugar Intake Is Associated with Body Shape, Lower Total Body-Fat Percentage, and Higher Blood Pressure. Cell Rep. 2018;23:327–336.

[182] Chambers JC, Zhang W, Sehmi J, Li X, Wass MN, Van der Harst P, et al. Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. Nat. Genet. 2011;43:1131–1138.

[183] Non-invasive Rapid Assessment of NAFLD Using Magnetic Resonance Imaging With LiverMultiScan - Full Text View - ClinicalTrials.gov [Internet]. [cited 2019 May 23];Available from: https://clinicaltrials.gov/ct2/show/NCT03289897

[184] Younossi ZM. Non-Alcoholic Fatty Liver Disease-A Global Public Health Perspective. J. Hepatol. [Internet]. 2018;Available from: http://dx.doi.org/10.1016/j.jhep.2018.10.033

[185] Ji Y, Yiorkas AM, Frau F, Mook-Kanamori D, Staiger H, Thomas EL, et al. Genome-Wide and Abdominal MRI Data Provide Evidence That a Genetically Determined Favorable Adiposity Phenotype Is Characterized by Lower Ectopic Liver Fat and Lower Risk of Type 2 Diabetes, Heart Disease, and Hypertension. Diabetes. 2019;68:207–219.

[186] Giri A, Hellwege JN, Keaton JM, Park J, Qiu C, Warren HR, et al. Trans-ethnic association study of blood pressure determinants in over 750,000 individuals. Nat. Genet. 2019;51:51–62.

[187] Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. Hum. Mol. Genet. 2018;27:3641–3649.

[188] Cantor RM, Lange K, Sinsheimer JS. Prioritizing GWAS results: A review of statistical methods and recommendations for their application. Am. J. Hum. Genet. 2010;86:6–22.

[189] Pavlides M, Banerjee R, Kelly CJ, Robson MD, Neubauer S, Barnes E. Reply to: "Multiparametric magnetic resonance imaging to predict clinical outcomes in patients with chronic liver disease: A cautionary note on a promising technique." J. Hepatol. 2017;66:457–458.

[190] Harrison SA, Dennis A, Fiore MM, Kelly MD, Kelly CJ, Paredes AH, et al. Utility and variability of three non-invasive liver fibrosis imaging modalities to evaluate efficacy of GR-MD-02 in subjects with NASH and bridging fibrosis during a phase-2 randomized clinical trial. PLoS One. 2018;13:e0203054.

[191] McDonald N, Eddowes PJ, Hodson J, Semple SIK, Davies NP, Kelly CJ, et al. Multiparametric magnetic resonance imaging for quantitation of liver disease: a two-centre cross-sectional observational study. Sci. Rep. 2018;8:9189.

[192] Ostovaneh MR, Ambale-Venkatesh B, Fuji T, Bakhshi H, Shah R, Murthy VL, et al. Association of Liver Fibrosis With Cardiovascular Diseases in the General Population: The Multi-Ethnic Study of Atherosclerosis (MESA). Circ. Cardiovasc. Imaging. 2018;11:e007241.

[193] Liu C-Y, Liu Y-C, Wu C, Armstrong A, Volpe GJ, van der Geest RJ, et al. Evaluation of age-related interstitial myocardial fibrosis with cardiac magnetic resonance contrast-enhanced T1 mapping: MESA (Multi-Ethnic Study of Atherosclerosis). J. Am. Coll. Cardiol. 2013;62:1280–1287.

[194] Mealer RG, Jenkins BG, Chen C-Y, Daly MJ, Ge T, Lehoux S, et al. A schizophrenia risk locus alters brain metal transport and plasma glycosylation [Internet]. bioRxiv. 2019 [cited 2019 Sep 26];757088. Available from: https://www.biorxiv.org/content/10.1101/757088v1

[195] Bachtiar V, Kelly MD, Wilman HR, Jacobs J, Newbould R, Kelly CJ, et al. Repeatability and reproducibility of multiparametric magnetic resonance imaging of the liver. PLoS One. 2019;14:e0214921.

[196] Liu M-J, Bao S, Gálvez-Peralta M, Pyle CJ, Rudawsky AC, Pavlovicz RE, et al. ZIP8 regulates host defense through zinc-mediated inhibition of NF-κB. Cell Rep. 2013;3:386–400.

[197] Haller G, McCall K, Jenkitkasemwong S, Sadler B, Antunes L, Nikolov M, et al. A missense variant in SLC39A8 is associated with severe idiopathic scoliosis. Nat. Commun. 2018;9:4171.

[198] Lin W, Vann DR, Doulias P-T, Wang T, Landesberg G, Li X, et al. Hepatic metal ion transporter ZIP8 regulates manganese homeostasis and manganese-dependent enzyme activity. J. Clin. Invest. 2017;127:2407–2417.

[199] Mousavi SN, Faghihi A, Motaghinejad M, Shiasi M, Imanparast F, Amiri HL, et al. Zinc and Selenium Co-supplementation Reduces Some Lipid Peroxidation and Angiogenesis Markers in a Rat Model of NAFLD-Fed High Fat Diet. Biol. Trace Elem. Res. 2018;181:288–295.

[200] Kang X, Zhong W, Liu J, Song Z, McClain CJ, Kang YJ, et al. Zinc supplementation reverses alcohol-induced steatosis in mice through reactivating hepatocyte nuclear factor-4alpha and peroxisome proliferator-activated receptor-alpha. Hepatology. 2009;50:1241–1250.

[201] Liu L, Geng X, Cai Y, Copple B, Yoshinaga M, Shen J, et al. Hepatic ZIP8 deficiency is associated with disrupted selenium homeostasis, liver pathology, and tumor formation. Am. J. Physiol. Gastrointest. Liver Physiol. 2018;315:G569–G579.

[202] Mukhopadhyay S. Familial manganese-induced neurotoxicity due to mutations in SLC30A10 or SLC39A14. Neurotoxicology. 2018;64:278–283.

[203] Tuschl K, Clayton PT, Gospe SM Jr, Gulab S, Ibrahim S, Singhi P, et al. Syndrome of hepatic cirrhosis, dystonia, polycythemia, and hypermanganesemia caused by mutations in SLC30A10, a manganese transporter in man. Am. J. Hum. Genet. 2012;90:457–466.

[204] Li D, Achkar J-P, Haritunians T, Jacobs JP, Hui KY, D'Amato M, et al. A Pleiotropic Missense Variant in SLC39A8 Is Associated With Crohn's Disease and Human Gut Microbiome Composition. Gastroenterology. 2016;151:724–732.

[205] Costas J. The highly pleiotropic gene SLC39A8 as an opportunity to gain insight into the molecular pathogenesis of schizophrenia. Am. J. Med. Genet. B Neuropsychiatr. Genet. 2018;177:274–283.

[206] Hommes FA, Bendien K, Elema JD, Bremer HJ, Lombeck I. Two cases of phosphoenolpyruvate carboxykinase deficiency. Acta Paediatr. Scand. 1976;65:233–240.

[207] Gill D, Brewer CF, Monori G, Trégouët D-A, Franceschini N, Giambartolomei C, et al. Effects of Genetically Determined Iron Status on Risk of Venous Thromboembolism and Carotid Atherosclerotic Disease: A Mendelian Randomization Study. J. Am. Heart Assoc. 2019;8:e012994.

[208] Yaghootkar H, Lotta LA, Tyrrell J, Smit RAJ, Jones SE, Donnelly L, et al. Genetic Evidence for a Link Between Favorable Adiposity and Lower Risk of Type 2 Diabetes, Hypertension, and Heart Disease. Diabetes. 2016;65:2448–2460.

[209] Younossi Z, Anstee QM, Marietti M, Hardy T, Henry L, Eslam M, et al. Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. Nat. Rev. Gastroenterol. Hepatol. 2017;nrgastro.2017.109.

[210] Söderberg C, Stål P, Askling J, Glaumann H, Lindberg G, Marmur J, et al. Decreased survival of subjects with elevated liver function tests during a 28-year follow-up. Hepatology. 2010;51:595–602.

[211] Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. Available from: https://qrisk.org/2017/BMJ-QRISK2.pdf

[212] Vilar-Gomez E, Chalasani N. Non-invasive assessment of non-alcoholic fatty liver disease: Clinical prediction rules and blood-based biomarkers. J. Hepatol. 2018;68:305–315.

[213] George J, Rapsomaniki E, Pujades-Rodriguez M, Shah AD, Denaxas S, Herrett E, et al. How Does Cardiovascular Disease First Present in Women and Men?CLINICAL PERSPECTIVES. Circulation. 2015;132:1320–1328.

[214] Pujades-Rodriguez M, George J, Shah AD, Rapsomaniki E, Denaxas S, West R, et al. Heterogeneous associations between smoking and a wide range of initial presentations of cardiovascular disease in 1937360 people in England: lifetime risks and implications for risk prediction. Int. J. Epidemiol. 2015;44:129–141.

[215] Rapsomaniki E, Timmis A, George J, Pujades-Rodriguez M, Shah AD, Denaxas S, 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. Lancet. 2014;383:1899–1911.

[216] Pujades-Rodriguez M, Timmis A, Stogiannis D, Rapsomaniki E, Denaxas S, Shah A, et al. Socioeconomic Deprivation and the Incidence of 12 Cardiovascular Diseases in 1.9 Million Women and Men: Implications for Risk Prediction and Prevention. PLoS One. 2014;9:e104671.

[217] Shah AD, Langenberg C, Rapsomaniki E, Denaxas S, Pujades-Rodriguez M, Gale CP, et al. Type 2 diabetes and incidence of cardiovascular diseases: a cohort study in 1·9 million people. The Lancet Diabetes & Endocrinology. 2015;3:105–113.

[218] Shah AD, Denaxas S, Nicholas O, Hingorani AD, Hemingway H. Neutrophil Counts and Initial Presentation of 12 Cardiovascular Diseases: A CALIBER Cohort Study. J. Am. Coll. Cardiol. 2017;69:1160–1169.

[219] Phenotype Library [Internet]. [cited 2021 Oct 31];Available from: https://phenotypes.healthdatagateway.org/home/

[220] Stensrud MJ, Hernán MA. Why test for proportional hazards? JAMA. 2020;323:1401–1402.

[221] Zhang Z. Multiple imputation with multivariate imputation by chained equation (MICE) package. Annals of translational medicine. 2016;4:30.

[222] Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the tidyverse. J. Open Source Softw. 2019;4:1686.

[223] Therneau TM. Survival Analysis [R package survival version 3.4-0]. 2022 [cited 2022 Aug 30];Available from: https://CRAN.R-project.org/package=survival

[224] Shah AG, Lydecker A, Murray K, Tetri BN, Contos MJ, Sanyal AJ, et al. Use of the Fib4 Index for Non-Invasive Evaluation of Fibrosis in Nonalcoholic Fatty Liver Disease. Clin. Gastroenterol. Hepatol. 2009;7:1104–1112.

[225] Targher G, Valbusa F, Bonapace S, Bertolini L, Zenari L, Rodella S, et al. Non-Alcoholic Fatty Liver Disease Is Associated with an Increased Incidence of Atrial Fibrillation in Patients with Type 2 Diabetes. PLoS One. 2013;8:e57183.

[226] Markus MRP, Baumeister SE, Stritzke J, Dörr M, Wallaschofski H, Völzke H, et al. Hepatic steatosis is associated with aortic valve sclerosis in the general population: the Study of Health in Pomerania (SHIP). Arterioscler. Thromb. Vasc. Biol. 2013;33:1690–1695.

[227] Bonapace S, Valbusa F, Bertolini L, Pichiri I, Mantovani A, Rossi A, et al. Nonalcoholic fatty liver disease is associated with aortic valve sclerosis in patients with type 2 diabetes mellitus. PLoS One. 2014;9:e88371.

[228] Mantovani A, Pernigo M, Bergamini C, Bonapace S, Lipari P, Valbusa F, et al. No Title. 2015;64. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0026049515001080

[229] Gong H, Liu X, Cheng F. Relationship between non-alcoholic fatty liver disease and cardiac arrhythmia: a systematic review and meta-analysis. J. Int. Med. Res. 2021;49:3000605211047074.

[230] Le Couteur DG, Blyth FM, Creasey HM, Handelsman DJ, Naganathan V, Sambrook PN, et al. The Association of Alanine Transaminase With Aging, Frailty, and Mortality. J. Gerontol. A Biol. Sci. Med. Sci. 2010;65A:712–717.

[231] Lazarus JV, Mark HE, Anstee QM, Arab JP, Batterham RL, Castera L, et al. Advancing the global public health agenda for NAFLD: a consensus statement. Nat. Rev. Gastroenterol. Hepatol. [Internet]. 2021;Available from: http://dx.doi.org/10.1038/s41575-021-00523-4

[232] Ward LD, Tu H-C, Quenneville CB, Tsour S, Flynn-Carroll AO, Parker MM, et al. GWAS of serum ALT and AST reveals an association of SLC30A10 Thr95Ile with hypermanganesemia symptoms. Nat. Commun. 2021;12:1–14.

[233] Hardy T, Wonders K, Younes R, Aithal GP, Aller R, Allison M, et al. The European NAFLD Registry: A real-world longitudinal cohort study of nonalcoholic fatty liver disease. Contemp. Clin. Trials. 2020;98:106175.

[234] Home [Internet]. HDR UK. 2019 [cited 2022 Aug 31];Available from: https://www.hdruk.ac.uk/

[235] Better, broader, safer: using health data for research and analysis [Internet]. GOV.UK. 2022 [cited 2022 Aug 31];Available from: https://www.gov.uk/government/publications/better-broader-safer-using-health-data-for-

research-and-analysis

[236] Ferris T. Data saves lives: reshaping health and social care with data [Internet]. GOV.UK. [cited 2022 Aug 31];Available from: https://www.gov.uk/government/publications/data-saves-lives-reshaping-health-and-social-care-with-data/data-saves-lives-reshaping-health-and-social-care-with-data

[237] Zhou W, Kanai M, Wu K-HH, Humaira R, Tsuo K, Hirbo JB, et al. Global Biobank Meta-analysis Initiative: powering genetic discovery across human diseases [Internet]. bioRxiv. 2021;Available from: http://medrxiv.org/lookup/doi/10.1101/2021.11.19.21266436

[238] Graham SE, Clarke SL, Wu K-HH, Kanoni S, Zajac GJM, Ramdas S, et al. The power of genetic diversity in genome-wide association studies of lipids. Nature [Internet]. 2021;Available from: http://dx.doi.org/10.1038/s41586-021-04064-3

[239] Noncommunicable diseases [Internet]. [cited 2021 Dec 19];Available from: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases