GENOMIC, METABOLOMOMIC AND CLINICAL PROFILING OF DYSLIPIDEMIA IN FAMILIES

Joel Rämö

Institute for Molecular Medicine Finland (FIMM)
Helsinki Institute of Life Science (HiLIFE)
Research Programs Unit
MD-PhD Programme
Faculty of Medicine
Doctoral Programme in Population Health
Doctoral School in Health Sciences
University of Helsinki
Helsinki, Finland

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Medicine of the University of Helsinki, for public examination in Lecture Hall 1, Haartman Institute, on the 6th of November, 2019 at 12 noon.

Helsinki 2019
The Faculty of Medicine uses the Urkund system (plagiarism recognition) to examine all doctoral dissertations.

Unigrafia
Helsinki, Finland 2019
Supervisors

Prof. Aarno Palotie, M.D., Ph.D.
Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Finland, and
The Broad Institute of MIT and Harvard, and
Massachusetts General Hospital, Boston, Massachusetts, USA

Prof. Samuli Ripatti, Ph.D.
Institute for Molecular Medicine Finland (FIMM), HiLIFE and Department of Public Health, University of Helsinki, Finland, and
The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA

Prof. Kirsi Pietiläinen, M.D., Ph.D.
Obesity Research Unit, Research Program for Clinical and Molecular Metabolism, University of Helsinki, Finland, and
Obesity Center, Abdominal Center, Endocrinology, Helsinki University Hospital, Finland

Reviewers

Prof. Johanna Kuusisto, M.D., Ph.D.
Centre for Medicine and Clinical Research, University of Eastern Finland, and
Kuopio University Hospital, Kuopio, Finland

Prof. Emma Ahlqvist, Ph.D.
Lund University Diabetes Centre, Department of Clinical Sciences, Lund University, Skåne University Hospital, Malmö, Sweden

Opponent

Prof. Katriina Aalto-Setälä, M.D., Ph.D.
Heart Group, Faculty of Medicine and Health Technology, Tampere University, and
Heart Hospital, Tampere University Hospital, Tampere, Finland
To my dear family
Despite decades of progress in primary prevention and treatment of acute coronary syndromes and strokes, cardiovascular disease (CVD) remains the leading cause of death and loss of disability-adjusted life years in Western countries. In this thesis, we sought to identify risk stratifying factors beyond the traditional measures of body mass index (BMI) and dyslipidemia. In particular, we focused on the role of hepatic steatosis in obesity, and on the family history of patients with hyperlipidemia.

We quantified circulating metabolites during an oral glucose tolerance test in BMI-discordant (ΔBMI ≥ 3 kg/m²) monozygotic twin pairs based on the presence of concomitant liver fat discordance. Liver fat -discordant cotwins exhibited greater putatively atherogenic differences in metabolomic parameters across a wide range of molecular classes, including lipoproteins, fatty acids, amino acids and glycoproteins. We also observed several putatively atherogenic differences between liver fat -concordant twin pairs, suggesting that increased BMI without concomitant liver fat accumulation may not be entirely neutral with respect to CVD risk.

We performed the first comprehensive genotyping analysis of familial combined hyperlipidemia (FCH), a common familial hyperlipidemia typically characterized by elevations in total cholesterol or triglycerides. We observed rare high-impact variants in the APOE or APOA5 genes in only a few (3%) of hyperlipidemic family members. Almost a third of hyperlipidemic family members had elevated polygenic burden for LDL-C or triglycerides, similar to population samples with comparable lipid levels.

Next, we estimated incident CVD risk in an overlapping cohort of hyperlipidemic families. We focused on common familial hyperlipidemias characterized by high LDL cholesterol or triglyceride levels after excluding individuals with monogenic FH. In our study, such familial hyperlipidemias conferred increased coronary artery disease and CVD risk, but the elevation in risk was similar to that observed in population-ascertained hyperlipidemias. Additionally, we observed highly similar lipidomic profiles consisting of 151 circulating lipid species between individuals with familial and population-ascertained hyperlipidemias.

Our results add to existing support for hepatic steatosis as a more discerning stratifying CVD risk factor among individuals with increased BMI. Our findings on FCH and familial aggregation of high LDL cholesterol or triglyceride levels suggest that they share similar and overlapping pathophysiology with common population-ascertained hyperlipidemias, and may not confer differential CVD risk.
Sydän- ja verisuonisairaudet ovat yhä länsimaiden yleisin kuolinsyy, vaikka niiden ennaltaehkäisy ja akuuttihoito ovat kehittynyt merkittävästi viime vuosikymmeninä. Tässä väitöskirjassa pyrimme löytämään uusia, tarkempia osatekijöitä kohonneeseen kehon painoindeksiin (BMI) ja dyslipidemiaan liittyvän sydän- ja verisuonitauringon arvioimiseksi. Keskitymme erityisesti lihavuuden yhteydessä esiintyvään maksarasvaan sekä dyslipidemian sukuhistoriaan.


Kartoitimme geneettisiä riskitekijöitä suomalaisissa perheissä, joissa esiintyy familiäalista hyperlipidemiaa (FKH), jota luonnehtivat kohonneet veren totaalikolesterolin ja/triglyseridien arvot. Totesimme vain pienellä osalla (3%) perheisissä suurivaikutteisia harvinaisia mutaatioita, jotka selittivät heidän taudinkuvaansa. Noin kolmasosalla hyperlipidemiaa sairastavista perheisissä oli kohonnut polygeeninen (useiden mutaatiojen yhteisvaikutuksesta koostuva) tautiriski. Perheenjäsenten geneettinen tausta oli hyvin samanlainen kuin väestönliikehyperlipideemialisillä kontrollihenkilöillä.

Totesimme myös, että harvinaisen monogeenisen familiäalisen hyperkolesterolemiain poissulun jälkeen suvuittaiseen hyperlipidemiaan liittyvä sepelvaltimotautiriski ei eronnut merkittävästi väestöaineistossa todettavasta vastaavan tausojen hyperlipidemiaan liittyvää tautiriskistä. Suvuittaisessa ja väestöaineistossa esiintyvä hyperlipidemiaan liittyvä riski oli lisäksi todettavissa samanlainen veren 151:stä rasvamolekylijistä koostuva profiili.

Tulkemme mukaan maksan rasvapitoisuutta piirteenä, joka voi mahdollistaa tarkemman sydän- ja verisuonitauringon riskiarvon lihavilla henkilöillä. Suvuittaista hyperlipidemiaa koskevat tuloksemme viittaavat näiden hyperlipidemoiden olevan hyvin samankaltaisia tauta ja väestössä muutenkin esiintyvät hyperlipidemiat. Tutkimuksemme on otettu huomioon suomalaisessa dyslipidemioihin Käypä Hoito -suositussa.
CONTENTS

ABSTRACT ..........................................................................................................................5
THIVISTELMÄ ..................................................................................................................6
CONTENTS ..........................................................................................................................7
LIST OF ORIGINAL PUBLICATIONS .................................................................10
ABBREVIATIONS ..............................................................................................................11

1 INTRODUCTION ...........................................................................................................13

2 REVIEW OF THE LITERATURE ...................................................................................16

2.1 Atherosclerotic cardiovascular disease: a modern epidemic 17
  2.1.1 Coronary artery disease, cerebrovascular disease and peripheral artery disease 17
  2.1.2 Trends in ASCVD prevalence and incidence ......................................................... 19
  2.1.3 Pathophysiology of atherosclerosis ..................................................................... 21
  2.1.4 Individual risk factors for ASCVD ....................................................................... 24
  2.1.5 Familial predisposition ....................................................................................... 31
  2.1.6 Predictive models ............................................................................................... 32
  2.1.7 More discernible tools for understanding individual risk ..................................... 33

2.2 Detailed phenotyping of circulating ASCVD risk markers .................................... 34
  2.2.1 Lipoprotein particle count and size ................................................................. 34
  2.2.2 Circulating lipid species ............................................................................... 36
  2.2.3 Other circulating metabolites ........................................................................ 39
  2.2.4 Fasting and postprandial measures ............................................................... 41

2.3 Stratification of obesity ............................................................................................ 42
  2.3.1 The concept of healthy obesity ....................................................................... 42
  2.3.2 Distribution and measurement of fat tissue ...................................................... 44
  2.3.3 Abdominal obesity and ASCVD risk .............................................................. 45
  2.3.4 Visceral fat and ASCVD risk ....................................................................... 46
  2.3.5 Liver fat and ASCVD risk ........................................................................... 47
  2.3.6 Proposed mechanisms underlying liver fat and cardiometabolic changes 49
  2.3.7 On timing and the duration of exposure ......................................................... 50

2.4 Genetic regulation of obesity and dyslipidemia ..................................................... 50
  2.4.1 Classification and consequences of genetic variation ..................................... 50
  2.4.2 The effect of rare high-impact variants on obesity and dyslipidemia .......... 52
  2.4.3 Effect of common variants and polygenic burden on obesity and dyslipidemia 54

3.1 Aims of the studies .................................................................................................. 58
4 METHODS ................................................................. 60

4.1 Study subjects ......................................................................... 60
  4.1.1 BMI-discordant twins ................................................................. 60
  4.1.2 Dyslipidemic family collection (EUFAM) ...................................... 60
  4.1.3 FINRISK population survey ...................................................... 61

4.2 Clinical measurements and laboratory methods ...................... 62
  4.2.1 Clinical MEASURES and blood sampling ...................................... 62
  4.2.2 Generation of NMR metabolomics data ....................................... 63
  4.2.3 Generation of lipidomics data .................................................... 64
  4.2.4 Genotyping and imputation ....................................................... 64
  4.2.5 HEALTH Registry data .......................................................... 65

4.3 Statistical methods .................................................................... 65
  4.3.1 Twin methodology .................................................................... 65
  4.3.2 Changes during an OGTT and Area Under the Curve measures ....... 66
  4.3.3 Variant enrichment analysis ....................................................... 66
  4.3.4 Polygenic scores ....................................................................... 66
  4.3.5 Proportional hazards model ....................................................... 67
  4.3.6 Lipidomics analyses ................................................................. 67
  4.3.7 Correction for relatedness ......................................................... 67
  4.3.8 Statistical software ................................................................... 68

5 RESULTS AND DISCUSSION ................................................. 69

5.1 Metabolomic profiles of acquired changes in BMI and liver fat
(Study I) ..................................................................................... 69
  5.1.1 Reactivity of circulating biomarkers to oral glucose tolerance test .... 69
  5.1.2 Acquired BMI-discordance with or without LF discordance: two distinct subgroups 72
  5.1.3 Few metabolomic differences between BMI-discordant MZ cotwins concordant for LF .......................................................... 72
  5.1.4 Widespread metabolomic differences between BMI-discordant MZ cotwins discordant for LF ..................................................... 74
  5.1.5 Associations of metabolomic measures with indices of insulin sensitivity 76

5.2 The Genetic architecture of FCH (Study II) .............................. 78
  5.2.1 Lipid-elevating variants are enriched in individuals with FCH across the allele frequency spectrum ........................................... 78
  5.2.2 High-impact variants in a minority of subjects with FCH or population-ascertained hyperlipidemia ................................................. 80
  5.2.3 High polygenic scores in up to a third of subjects with FCH or population-ascertained hyperlipidemia ................................................. 81

5.3 Coronary artery disease risk and lipidomic profiles in familial and population-ascertained hyperlipidemias (Study III) ............. 83
| 5.3.1 | Coronary artery disease risk in familial and population-ascertained hyperlipidemias | 84 |
| 5.3.2 | Assessing the biological validity of two lipidomics platforms | 86 |
| 5.3.3 | Lipidomic profiles of high LDL-C | 91 |
| 5.3.4 | Lipidomic profiles of high TGs | 92 |
| 5.3.5 | Adjustment for combined hyperlipidemia | 94 |

6 **CONCLUSIONS** ........................................................................................................... 98

6.1 Acquired BMI and metabolomic profiles – the role of liver fat 98

6.2 Familial and population-ascertained hyperlipidemias – similar genomic, lipidomic and clinical characteristics ........................................... 99

7 **CONCLUDING REMARKS AND FUTURE PROSPECTS** ...... 101

8 **REFERENCES** .............................................................................................................. 107

9 **ORIGINAL PUBLICATIONS** ..................................................................................... 140
LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following publications:


The publications are referred to in the text by their roman numerals. The original publications are reproduced with the permission of the copyright holders.
## ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>-C</td>
<td>cholesterol</td>
</tr>
<tr>
<td>ApoB</td>
<td>apolipoprotein B</td>
</tr>
<tr>
<td>ASCVD</td>
<td>atherosclerotic cardiovascular disease</td>
</tr>
<tr>
<td>AUC</td>
<td>area under the curve</td>
</tr>
<tr>
<td>BMI</td>
<td>body mass index</td>
</tr>
<tr>
<td>CAD</td>
<td>coronary artery disease</td>
</tr>
<tr>
<td>CE</td>
<td>cholesteryl ester</td>
</tr>
<tr>
<td>Cer</td>
<td>ceramide</td>
</tr>
<tr>
<td>CI</td>
<td>confidence interval</td>
</tr>
<tr>
<td>CVD</td>
<td>cardiovascular disease</td>
</tr>
<tr>
<td>DG</td>
<td>diacylglyceride</td>
</tr>
<tr>
<td>DEXA</td>
<td>dual-energy X-ray absorptiometry</td>
</tr>
<tr>
<td>DZ</td>
<td>dizygotic</td>
</tr>
<tr>
<td>EUFAM</td>
<td>European Multicenter Study on Familial Dyslipidemias in Patients with Premature Coronary Heart Disease</td>
</tr>
<tr>
<td>FA</td>
<td>fatty acid</td>
</tr>
<tr>
<td>FCH</td>
<td>familial combined hyperlipidemia</td>
</tr>
<tr>
<td>FH</td>
<td>familial hypercholesterolemia</td>
</tr>
<tr>
<td>GWAS</td>
<td>genome-wide association study</td>
</tr>
<tr>
<td>HDL</td>
<td>high-density lipoprotein</td>
</tr>
<tr>
<td>HR</td>
<td>hazard ratio</td>
</tr>
<tr>
<td>ICD</td>
<td>international classification of diseases</td>
</tr>
<tr>
<td>IDL</td>
<td>intermediate-density lipoprotein</td>
</tr>
<tr>
<td>LDL</td>
<td>low-density lipoprotein</td>
</tr>
<tr>
<td>LDLR</td>
<td>low-density lipoprotein receptor</td>
</tr>
<tr>
<td>LF</td>
<td>liver fat</td>
</tr>
<tr>
<td>LPC</td>
<td>lysophosphatidylcholine</td>
</tr>
<tr>
<td>LPE</td>
<td>lysophosphatidylethanolamine</td>
</tr>
<tr>
<td>MI</td>
<td>myocardial infarction</td>
</tr>
<tr>
<td>MHO</td>
<td>metabolically healthy obesity</td>
</tr>
<tr>
<td>MRI</td>
<td>magnetic resonance imaging</td>
</tr>
<tr>
<td>MZ</td>
<td>monozygotic</td>
</tr>
<tr>
<td>NAFLD</td>
<td>non-alcoholic fatty liver disease</td>
</tr>
<tr>
<td>NASH</td>
<td>non-alcoholic steatohepatosis</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>OGTT</td>
<td>oral glucose tolerance test</td>
</tr>
<tr>
<td>OR</td>
<td>odds ratio</td>
</tr>
<tr>
<td>PC</td>
<td>phosphatidylcholine</td>
</tr>
<tr>
<td>PCO</td>
<td>phosphatidylcholine-ether</td>
</tr>
<tr>
<td>PE</td>
<td>phosphatidylethanolamine</td>
</tr>
<tr>
<td>PEO</td>
<td>phosphatidylethanolamine-ether</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>PI</td>
<td>phosphatidylinositol</td>
</tr>
<tr>
<td>PRS</td>
<td>polygenic risk score</td>
</tr>
<tr>
<td>RR</td>
<td>risk ratio</td>
</tr>
<tr>
<td>SAT</td>
<td>subcutaneous adipose tissue</td>
</tr>
<tr>
<td>SM</td>
<td>sphingomyelin</td>
</tr>
<tr>
<td>ST</td>
<td>sterol</td>
</tr>
<tr>
<td>TG</td>
<td>triacylglyceride</td>
</tr>
<tr>
<td>VAT</td>
<td>visceral adipose tissue</td>
</tr>
<tr>
<td>VLDL</td>
<td>very low-density lipoprotein cholesterol</td>
</tr>
<tr>
<td>WC</td>
<td>waist circumference</td>
</tr>
</tbody>
</table>
1 INTRODUCTION

Atherosclerotic cardiovascular disease (ASCVD) rose in the 1900s to replace communicable diseases as the most common cause of death in Western countries.\(^1\) Despite decades of advancement in the treatment of coronary syndromes since, as well as public measures to curb lifestyle factors associated with increased risk, cardiovascular disease still accounts for 30-40\% of deaths in Western countries.\(^2\)\(^-\)\(^4\) A large subset of acute coronary syndrome patients never reach advanced care; improving their prognosis requires more effective and targeted preventive measures to halt the chronic disease early.\(^5\)\(^-\)\(^7\) Personalized medicine approaches address these unmet needs, aiming to dissect and act on individual risk beyond traditional population-level risk factors.

Dyslipidemia and obesity are traditional, often coexisting, risk factors for ASCVD. At the individual level they are, however, imperfect predictors of risk. Traditionally, a BMI over 30 kg/m\(^2\) – easily measurable in clinical practice – is used for the diagnosis of obesity.\(^8\) A subset of obese individuals – sometimes called the metabolically healthy obese (MHO) – are, however, free of dyslipidemia and other risk factors such as dysglycemia and elevated blood pressure.\(^9\)\(^-\)\(^12\) Many studies have reported lower risk of ASCVD in MHO than in the metabolically unhealthy obese.\(^13\)\(^-\)\(^16\) To explain why the same BMI translates to varying cardiometabolic outcomes, the location of fat has been raised as a key defining feature. It has long been recognized, for example, that for a given BMI, subcutaneous fat may accumulate preferentially in an abdominal (android) or femorogluteal (gynoid) distribution, with the android distribution associating with more pronounced ASCVD risk.\(^17\)

More recent imaging studies have also revealed differences in compartmental adiposity. Instead of subcutaneous adipose tissue, some individuals accumulate considerable adipose tissue within the visceral cavity.\(^18\) Further, some exhibit fat in non-adipose tissue, such as the liver. Accumulating data suggests that the content of liver fat – which unlike BMI follows a highly skewed distribution in the population – may delineate many of the cardiometabolic states associated with obesity and serve as a more precise marker of actual ASCVD risk.\(^10\), \(^19\) The phenotypes associated with liver fat are complex, however, and may depend on etiology. Several genetic variants predisposing to liver fat do not confer a similar lipid phenotypes or ASCVD risk as observed in population studies. To further dissect the detailed metabolic changes in obesity-associated liver fat accumulation, studies must control for the genetic makeup of individuals.

Twin studies allow for the evaluation of acquired changes: as the genetic makeup of monozygotic twins is identical, differences between the twins
acquired later in life must follow from lifestyle and environmental exposure. In a twin study of acquired obesity, the content of liver fat strikingly separated those who developed dyslipidemia, dysglycemia, and high blood pressure, from those who remained metabolically healthy.10

Detailed profiling of circulating lipid and nonlipid metabolites may uncover pathways relevant to disease. In addition to conventional measures of dyslipidemia, in population studies, liver fat has been associated with widespread changes in circulating and hepatic lipid species, suggesting that dysfunctional hepatic fatty acid metabolism may bridge obesity-associated fatty liver disease and dyslipidemia.20-23 Most studies, however, have been unable to adjust for genetic background, and it is still poorly understood to what extent these changes are modifiable.

Although inherited susceptibility to liver fat does not seem to confer classical atheroregic lipid phenotypes, dyslipidemias – especially extreme forms – appear highly heritable. In individuals with familial hypercholesterolemia (FH), the concentration of the atherogenic low-density lipoprotein cholesterol (LDL-C) is considerably elevated due to mutations in a small number of genes such as LDLR, APOB, and PCSK9, often resulting in considerably increased ASCVD risk despite adherence to a healthy lifestyle. Familial hypercholesterolemia is a classical example of a series of rare monogenic diseases, which predispose to an extreme phenotype due to individual high-impact genetic variants. Knowledge of a hypercholesterolemic individual’s genetic background improves ASCVD prediction: those with monogenic FH have higher lifelong cumulative exposure to high LDL-C levels and elevated risk compared with population controls exhibiting similar LDL-C levels in adulthood. These individuals, then, seem genetically destined for persistently higher ASCVD risk, and guidelines recommend more intensive treatment for them, including novel lipid-lowering therapies targeting the PCSK9 protein.24,25

Considerable efforts have been made to discover similar monogenic bases for other familial dyslipidemias, which have traditionally been considered similarly Mendelian disorders. Findings have, however, been scarce. A locus near the gene USF1 was linked with familial combined hypercholesterolemia (FCH) – a disorder characterized by familial segregation of high total cholesterol and triglycerides, more common in the population than FH and present in 11-14% of individuals with premature coronary artery disease.26-29 Other studies have noted high-impact variants in LPL in a minority of individuals with FCH.30-32 These findings, despite initial optimism, have mostly not translated to clinically actionable risk stratification, and no targeted medication exists for reducing ASCVD risk in individuals with FCH. Whereas 2011 ESC/EAS guidelines recommended screening of FCH relatives and follow-up of affected individuals in lipid clinics, the most recent 2019
guidelines refrain from screening recommendations and, partly citing results from Study II, state that “FCH has no monogenic component and is not linked to a single genetic cause.”33-34

One explanation for the difficulty of discovering new high-impact genetic variants in many familial diseases is that they may not exist or may be exceedingly rare, requiring very large sample sizes for detection. Instead of rare variation, some phenotypes presenting with familial segregation patterns may actually reflect more common polygenic burden: the accumulation of many low-to intermediate variants across the whole genome.35 In a paradigm shift, Talmud and colleagues showed in 2013 that some individuals with clinical FH but carrying no identified high-impact variant exhibited high polygenic burden for LDL-C, proposing a novel designation of “polygenic hypercholesterolemia”.36 These individuals may have lower ASCVD risk than those with monogenic FH, although conclusive data is still lacking.37

Polygenic burden has also been shown to contribute to extreme levels of high-density lipoprotein cholesterol.38 Prior to this thesis, it remained an open question to which extent FCH also reflects such polygenic influences. Further, if FCH is a polygenic disease – instead of the result of a disruption of a single pathway as observed in monogenic FH – would it differ phenotypically from hyperlipidemia observed at random in the population? And would it confer different risk of ASCVD compared with population-ascertained hyperlipidemia?
2 REVIEW OF THE LITERATURE

On a cold day late in the winter of 2018, in an emergency department of the McGill University Health Centre in Montréal, Canada, the author of this thesis was interviewing a patient who had presented to the emergency department with a complaint of vague weakness, dyspnea and palpitations. She was in her 50s and otherwise healthy save for having hypercholesterolemia and being overweight. Walking downtown some weeks before, she had noticed becoming out of breath more easily than usual. Back at home, the symptoms had passed. Over the following weeks, however, the symptoms had periodically recurred – varied, but increasingly disabling – especially during brisk walks. Finally, that morning the weakness had been so pronounced that she had had to lie down on the street and wait for it to pass. Passers-by had called an ambulance, and paramedics had brought her to the hospital for an urgent evaluation.

Her presentation was concerning for a coronary origin – a problem in the vessels that nourished her heart muscle. She was not particularly surprised about this. Heart disease “ran” in her family. Her father and sibling had been diagnosed with it. Her grandmother had died suddenly; relatives had said it was due to her heart. Many family members were also overweight or obese, and – the patient recalled – had had trouble with cholesterol.

The patient was scheduled for a coronary angiography, which found diffuse atherosclerosis in her coronary arteries. One artery, in particular, had a focal high-degree stenosis. This presumed culprit lesion was opened with a drug-eluting stent and the patient was started on an array of medications. Some of these, including a statin to control her high cholesterol, she was advised to remain on for life, to stabilize her newly diagnosed coronary artery disease.

Unlike her grandmother, she had escaped her acute attack without life-threatening consequences. She would, nevertheless, be at heightened risk of a recurrent attack in the future. Although treatment of acute coronary syndromes has improved in recent decades, many patients with an acute syndrome never make it to the hospital, and many of those who do never make it back to independent living at home. The only way to prevent such mortality and morbidity is to prevent the very development of the underlying disease in susceptible individuals. To achieve this, we must first understand who is at risk, and why.
2.1 Atherosclerotic Cardiovascular Disease: A Modern Epidemic

2.1.1 Coronary Artery Disease, Cerebrovascular Disease and Peripheral Artery Disease

No human organ can survive without perfusion; all tissues are connected – both in function and in anatomy – by the vascular system. The beating of the heart propels an average of 5 litres of blood through this system every minute, allowing for a constant flow of nutrients and oxygen to tissues, as well as the removal of byproducts of tissue metabolism.\textsuperscript{39, 40} It is no surprise, then, that major disruptions of this system cause immediate symptoms, and if not resolved, permanent tissue damage.

Several mechanisms can cause obstruction of blood flow. These include local stenoses of the vessels, embolism of a clot from elsewhere in the body, or mechanical dissection of the layers of the arterial wall.\textsuperscript{41-44} The most common origins of obstruction are atherosclerotic lesions of the vessel wall, which capable of giving rise to both local stenosis or thrombosis and distal emboli.\textsuperscript{41}

Atherosclerosis does not appear suddenly. Despite a propensity for producing catastrophic life-threatening syndromes, in contrast with the sudden event of a vascular dissection, an atherosclerotic lesion is often present for years or decades before the onset of any symptoms.\textsuperscript{45}

Atherosclerotic lesions may manifest in any artery in the human body, with drastically different sequelae depending on their site and characteristics. Atherosclerotic cardiovascular disease (ASCVD) is thus commonly divided into regional subtypes, including coronary artery disease (CAD), cerebrovascular disease, and peripheral artery disease (PAD).\textsuperscript{46}

In CAD, atherosclerosis is present in the coronary arteries, the vessels which nourish the heart muscle itself. In its stable form, CAD can cause chest pain and dyspnea with exertion, as well as a heart failure due to ischemic cardiomyopathy.\textsuperscript{47} Unstable CAD lesions, characterized by a susceptibility for thrombosis within the vessel lumen at the lesion site, may provoke myocardial infarction (MI): a sudden subtotal or total obstruction of a coronary artery, which can cause catastrophic failure of the heart’s capacity to support the whole circulation (“circulatory shock”), or sudden death due to fatal arrhythmias.\textsuperscript{41, 42}

In those who survive the initial event of a myocardial infarction, modern treatment can drastically change the natural course of the disease. In the 1960s, in-hospital mortality was still very high at 29\%.\textsuperscript{48} Since the 1950s,
direct visualization of the coronary arteries has been possible through the intra-arterial injection of radiographic contrast material to the roots of the main coronary arteries via specialized catheters. In individuals with acute coronary syndromes, such imaging typically shows either a local narrowing or complete blockage of the lumen of the artery. After localization of the culprit lesion, reperfusion can be pursued through the use of intra-arterial balloons inflated at the site of the lesion to push the obstructing material outwards, re-establishing flow. Stenting material can be inserted to further keep the artery open. In many cases, dramatic recovery from the presenting state is possible, with patients returning to independent daily life. However, even with improved diagnosis and treatment, in-hospital mortality remains at 4-7%. Among those who are discharged, 17-28% have suffered from heart failure in contemporary case cohorts. Perhaps most concerning, approximately 20-29% of those with a first incidence of MI die before reaching the hospital. No improvements in hospital care can save such patients; they highlight a dire need to prevent the disease in the first place.

It was previously thought that screening for stable CAD, and subsequent dilatation of the lesions might prevent future MIs. However, though effective in the setting of acute MI, the benefit of catheter-based procedures in stable CAD remains limited to selected situations. Instead, lifestyle and pharmacological management constitute the mainstay of treatment. Despite considerable advances in recent decades, however, treatment is able to halt the progression of CAD or prevent MIs in only a subset of patients.

CAD, though common, is not the only sequela of atherosclerotic disease. Atherosclerosis is also a major contributor to cerebrovascular disease. The brain is among the organs least resistant to persistent ischemia, tolerating obstruction of perfusion for only minutes. In contrast with MI, in which thrombosis and obstruction typically occur near the lesion site, many strokes originate from more proximal atherosclerotic lesions – e.g. in the carotid arteries – from which clot material may travel to a more distal site in the intracranial vasculature. A stroke can also occur due to other causes, such as blood clots formed within the heart’s chambers due to atrial fibrillation.

Reflecting both different pathophysiology and procedural risks, the success rates of modern treatments for severe stroke remain lower than those for acute coronary syndromes. If the benefits of intervention are deemed higher than the risks, treatment typically includes the injection of fibrinolytic agents to promote breakdown of the clot, and in selected cases, intra-arterial thrombectomy with the aid of specialized catheters and suction devices to
remove the clot. Still, even with the best treatment and rehabilitation, mortality and morbidity are high in the immediate period after a stroke, and remain elevated in the following years in those who survive the initial presentation.\textsuperscript{61}

In addition to the heart and the brain, atherosclerosis may affect the arteries perfusing any other organ in the human body. For example, atherosclerosis in the lower extremities – often simply termed peripheral artery disease – may affect 13\% of adults over 50 years, and is symptomatic in approximately 5\% of individuals between 55 and 74 years.\textsuperscript{62, 63} Early lower extremity PAD may present with symptoms of claudication (muscle pain at exercise which is relieved with rest).\textsuperscript{64} Later and unstable forms of the disease may cause rest pain and eventually - similarly to acute myocardial infarctions – necrosis of limb tissues.

ASCVD, bridging organ systems and medical specialties, is a significant contributor to worldwide mortality and morbidity. Yet, it does not seem to affect all equally, nor has its prevalence remained stable in recent times. Individual risk factors, both modifiable and fixed, modulate cardiovascular risk. To put their significance in context, what follows is first an overview of changing temporal trends in global ASCVD burden. The pathophysiology of atherosclerotic plaques will then be reviewed before a more detailed focus on individual risk factors – including the potential for improved risk stratification and targeted prevention with a focus on dyslipidemia.

\subsection*{2.1.2 TRENDS IN ASCVD PREVALENCE AND INCIDENCE}

The landscape of ASCVD has been in flux ever since the late 1800s. In the US, deaths due to heart disease rose to peak in the 1960s, when every third American died due to heart disease.\textsuperscript{1} Since then, developed countries have seen a notable and consistent decline in cardiovascular mortality in the later half of the 1900s and early 2000s.\textsuperscript{2} These benefits are not limited to developing countries. The age-standardized mortality from CAD has also decreased globally from 2005 to 2015.

Nevertheless, the local and global burden of ASCVD is still tremendous. In fact, driven partly by aging and growing populations, the absolute incidence of deaths due to CAD remains on the rise. Depending on the definition, ischemic heart disease is currently the leading individual cause of death worldwide, followed by stroke as the third leading cause.\textsuperscript{2-4} ASCVD is also the leading cause of loss of disability-adjusted life years.\textsuperscript{65} Far from being a solely Western disease, ASCVD is now prevalent in both developed and developing countries, with 80\% of global ASCVD deaths occurring in developing nations.\textsuperscript{4} Driven
mostly by CAD and cerebrovascular disease, ASCVD accounts for approximately 30% of deaths globally and 40% of deaths in the EU.\textsuperscript{4, 66}

Changes in national ASCVD burdens have been closely associated with changing environmental and lifestyle factors, most notably increasing energy intake, diets high in sugar and salt content, decreasing physical activity and increasing tobacco consumption.\textsuperscript{67-74} Lifestyle changes have presented in conjunction with measurable risk factors such as high BMI, high blood pressure, dyslipidemia, and diabetes. Such ASCVD risk factors used to be more prominent in developed countries, but today their contribution to ASCVD is more pronounced in developing countries, likely reflecting changes in lifestyle.\textsuperscript{2, 73, 75, 76}

Finland has a unique place in the history of ASCVD prevention by public lifestyle interventions. In the 1960s and 1970s, international statistics such as the Seven Countries Study revealed that mortality due to CAD was higher in Finland than in any other participating country.\textsuperscript{77} Further, Finns had the highest average blood pressure and serum cholesterol values among the countries, and half of Finnish men reported smoking. The prevalence of CAD was particularly high in Eastern Finland, where a far-reaching population-based primary prevention programme (The North Karelia Project) was instituted with a concomitant 82-84% decrease in premature CAD mortality from 1972 to 2012.\textsuperscript{78} Most of the observed reduction could be explained by changes in three main risk factors: smoking, serum cholesterol and systolic blood pressure.

Overall, the age-standardized mortality due to CAD has more than halved in Finland from 1971 to 2014.\textsuperscript{79} Simultaneously, the age of those who die due to CAD has risen from 65 years for men and 73 years for women 1971 to 78 or 87 years, respectively, in 2014. The incidence of CAD remains higher in Eastern Finland than in other parts of the country. In addition to lifestyle factors, differences in the polygenic burden of CAD risk variants have been proposed to contribute to this geographical distribution.\textsuperscript{80}

Predictions of ASCVD prevalence in the future vary. In Western countries, young individuals aged 18-50 are increasingly affected by ASCVD risk factors, most notably overweight.\textsuperscript{81} The prevalence of ASCVD has not decreased among young individuals, and has even increased in some Western countries, leading to concerns of a new wave of incident ASCVD along with their aging. Worryingly, the most recent estimates from the US indicate a slowing down of the decline in CVD mortality, suggesting that current prevention strategies are insufficient to halt this disease burden.\textsuperscript{82}

In addition to causing individual suffering, ASCVD exerts tremendous economic pressures. The global costs of cardiovascular disease in 2010 were
estimated at 863 billion USD, 55% of which were direct and 45% of which were indirect costs (modelled as productivity loss due to disability or death or time loss from work). These costs are unevenly distributed. The total costs have been estimated at over 150 billion USD in Europe and at over 500 billion USD in the US. Globally, the cost of ASCVD is predicted to rise by over 20% by 2030. Prevention of ASCVD has become a major health policy focus of both developed and developing countries, the latter of which have in many cases experienced rapid shifts from a focus on undernourishment to a focus on lifestyle risk factors. Preventive measures, including lifestyle changes and use of common medication to control measurable risk factors, are on average more cost effective than treatment after ASCVD diagnosis; further, their benefit is greatest when they are targeted to those at highest risk.

2.1.3 PATHOPHYSIOLOGY OF ATHEROSCLEROSIS

The earliest treatment modalities for ASCVD viewed the disease as primarily a mechanical (“plumbing”) issue to be solved with bypassing and dilatation. Currently, ASCVD is viewed as a multistage (Figure 1) and system-wide process, in which the interplay of dyslipidemia, endothelial dysfunction, inflammation, and thrombosis act to produce either slow occlusion or abrupt disruption of the arterial lumen.

![Figure 1. Progression of an atherosclerotic lesion. Reprinted from Cell, Vol 145, Moore and Tabas, Macrophages in the Pathogenesis of Atherosclerosis, Pages 341-355, Copyright (2011) with permission from Elsevier.](image)

Development of fatty streaks as precursor lesions

The first manifestations of atherosclerotic lesions are fatty streaks, in which lipid molecules (mostly cholesterol esters, but also free cholesterol, phospholipids and triacylglycerols) accumulate in the intimal layer of
arteries. Circulating apolipoprotein B (ApoB) containing lipoproteins (including low-density lipoprotein [LDL], intermediate-density lipoprotein [IDL], and very low-density lipoprotein [VLDL]) are thought to have a key role as the initiators of these lesions, passing through the vascular endothelium into the intima.

Early mechanistic studies noted that atherosclerotic lesions began to develop if the endothelium was mechanically removed, and postulated endothelial desquamation as an origin for the disease. Later, it was recognized that the endothelium in atherosclerotic lesions was actually intact but dysfunctional, mediating for example a paradoxical vasoconstriction in response to acetylcholine, a stimulus which normally leads to arterial dilatation. A dysfunctional endothelium may facilitate the entry of ApoB-containing lipoproteins. However, the timing and role of endothelial dysfunction still remain incompletely understood.

Lipoproteins may exit the intima back to the circulation, but this is less likely if they aggregate together, bind to extracellular matrix molecules (primarily proteoglycans), and undergo enzymatic and chemical modifications, such as oxidation via reactive oxygen species produced by the endothelium and by inflammatory cells. Oxidized lipoproteins, particularly LDL, may in turn stimulate cytokine production by endothelial and smooth muscle cells and subsequently attract further inflammatory cells, including monocytes, from the circulation.

Monocytes circulating in the bloodstream adhere to adhesion molecules present on the endothelial cells and subsequently move through the endothelial barrier. Within the arterial wall, they are transformed to locally acting macrophages. Exhibiting scavenger receptors, macrophages take in LDL and are often transformed into foam cells, which may have additional proinflammatory properties.

**Plaque progression**

As fatty streaks mature, they can attract smooth muscle cells from the muscle layer – media – of the artery to the intimal layer. Smooth muscle cells proliferate locally and produce fibrous extracellular matrix (including collagen and elastin), increasing the size of the lesion and consequently narrowing the arterial lumen. In the later stages of this process, the lesion may undergo partial calsification and apoptosis of the smooth muscle cells, resulting in a hardened, acellular fibrous cap over a nucleus enriched in lipids and foam cells. The plaque core may also contain necrotic debris associated with apoptosis and defective clearance of macrophages.
Plaque disruption

In some progressed plaques, the fibrous cap is thin, reflecting a balance between secretion of fibrotic material and its degradation by proteolytic enzymes secreted by foam cells. Although often producing a smaller degree of stenosis, these thin-capped lesions are prone to mechanical stress and are at higher risk of rupture, leading to an acute prothrombotic state and sudden occlusion of the lumen. A thrombotic occlusion may also be produced due to superficial plaque or calcium nodule erosion even without rupture of the lesion core.

In the coronary arteries, these events can result in unstable angina (in the case of a resolving occlusion), or acute myocardial infarction with severe myocardial necrosis (in the case of a permanent occlusion). From carotid arteries, thrombotic material may also embolize to more distant intracranial arteries.

Several mechanisms have been proposed to explain the susceptibility to thrombosis in plaque disruption. These include platelet activation due to contact with collagen, and activation of the coagulation cascade due to tissue factor production by the macrophages and smooth muscle cells. Further, a normal coronary endothelium secretes vasodilators and suppresses platelet activation and aggregation by promoting the breakdown of the prothrombotic mediators adenosine 5′-triphosphate (ATP) and adenosine diphosphate (ADP) into a precursor of the antithrombotic mediator adenosine. This function can be impaired at the site of an unstable plaque.

Circulating (“fluid state”) factors can also promote thrombosis; for example, the risk of sudden ASCVD events appears to be higher in individuals with systemic procoagulant states. Plaque thrombosis can thus be understood as an interplay of local and systematic mechanisms which can further provoke each other.

Regression of plaques through pharmacological intervention

Although typically progressive throughout life, atherosclerotic lesions can also regress. Highlighting the key role of dyslipidemia in plaque maturation, pharmacological lowering of ApoB-containing lipoproteins is associated not only with reduction in plaque lipid content, but also in foam cells, necrosis and fibrosis. In addition, endothelial dysfunction may be ameliorated with LDL-C-lowering therapy. The extent of such plaque regression may be higher in young individuals with early atherosclerotic lesions than in those with more mature disease. Whether, and how, the development of ASCVD could be prevented – as opposed to merely slowed down – by identifying and acting on early plaques is still debated and a subject for ongoing studies.
2.1.4 INDIVIDUAL RISK FACTORS FOR ASCVD

Unlike many diseases contributing to global mortality and morbidity, ASCVD risk is elevated by risk factors that are often highly modifiable. The World Health Organization has estimated that 80% of premature heart disease and stroke is preventable. Some risk factors are often intertwined (such as dyslipidemia, obesity and diabetes), while others (such as smoking) are more isolated in their effects.

Dyslipidemia

Lipids, including cholesterol, triglycerides (TGs), and phospholipids, are mainly transported in the circulation in lipoproteins – particles composed of apolipoprotein molecules, an outer amphiphilic lipid monolayer, and a lipophilic inner lipid core. Lipoproteins are produced in the liver and partly in the gut. The metabolism of different lipoprotein particles is closely intertwined, and as such it is no surprise that the understanding of the contribution of various lipoprotein parameters to ASCVD still continues to evolve. To provide context for this connectedness, the metabolism of lipoproteins is reviewed briefly below for both the exogenous (chylomicron-producing) and endogenous pathways (Figure 2).

After a meal, fatty acids and monoacylglycerol are absorbed into endothelial cells in the small intestine, within which they can be converted to triacylglycerols and assembled into chylomicrons, large lipid particles consisting of an inner lipid core and an outer lipid monolayer, as well as the structural protein apolipoprotein B48. Chylomicrons are released into lymphatic vessels and eventually to the circulation, where they transport their mostly triglyceride content to peripheral tissues and the liver.

The liver produces lipoproteins in two main pathways, beginning with either very low-density lipoprotein (VLDL) or high-density lipoprotein (HDL) particles. Compared with chylomicrons, VLDL particles carry a different structural apolipoprotein (apolipoprotein B100). They can be separated into two main forms: the larger and more TG-rich VLDL₁, and the smaller VLDL₂ particles. The content of both chylomicron and VLDL particles in the circulation is mediated by enzymes such as lipoprotein lipase (LPL), which acts to hydrolize triglycerides on the luminal side of the vessel wall within adipose, muscle and heart tissue. This can also occur in atherosclerotic plaques, where TGs are cleaved by LPL near the endothelium or within the intima, generating fatty acids and monoacylglycerol.
Figure 2. Simplified schematic of lipoprotein metabolism. In the exogenous pathway, dietary fats are processed in the intestine and secreted as triglyceride-rich chylomicrons which eventually enter the circulation. In the endogenous pathway, the liver produces triglyceride-rich VLDL lipoproteins the content of which is released to peripheral tissues by the action of lipoprotein lipase (LPL). VLDL lipoproteins may subsequently undergo conversion to IDL and cholesterol-rich LDL lipoproteins. The cholesterol content of LDL lipoproteins can be transferred to peripheral cells or the liver via the LDL receptor. The liver also produces HDL lipoproteins, which function in reverse cholesterol transport from peripheral tissues back to the liver. Reprinted by permission from Springer Nature: Nature Reviews Nephrology. Dyslipidaemia in nephrotic syndrome: mechanisms and treatment. Agrawal et al. Copyright (2017).

LPL, however, does not affect the cholesterol content of these lipoproteins. As the TG-content of VLDL particles is reduced by the action of LPL, they are transformed to intermediate-density lipoprotein (IDL) particles and eventually to low-density lipoprotein (LDL) particles. Other enzymes, such as the hepatic lipase, also play a role in this transformation. LDL is rich in cholesterol and relatively poor in TG content compared with VLDL. This cholesterol content is transported to peripheral cells. Cholesterol from LDL,
IDL and chylomicrons may also be reabsorbed into the via LDL receptor (LDLR) mediated endocytosis.124

On the contrast, high-density lipoprotein (HDL) particles secreted by the liver function in reverse cholesterol transport (RCT), transporting cholesterol from peripheral tissues and tissue macrophages to the liver.125 The structural lipoprotein of HDL is apolipoprotein A1, and to a lesser extent, apolipoprotein A2. HDL particles take in cholesterol from peripheral cells mediated by the SR-B1 receptor and the ABCA1 (for nascent HDL particles) and ABCG1 (for mature HDL particles) transporters. In the liver, SR-B1 also participates in the hepatic intake of cholesterol, while the intake of triglycerides is mediated by hepatic lipase.126 Several other enzymes also act in RCT. After intake from peripheral tissues to HDL particles, unesterified cholesterol is transformed to cholesteryl esters by the action of lecithin-cholesterol acyltransferase (LCAT), allowing for relocation into the lipophilic core.125 In the circulation, HDL particles also engage in an exchange of cholesterol for triglycerides with VLDL, LDL and chylomicron particles, mediated by the cholesterylester transfer protein (CETP).127, 128

Multiple lines of evidence suggest that lipoprotein metabolism is associated with both preclinical atherosclerosis as well as ASCVD events. In particular, evidence from genetic studies and clinical trials has pointed to a causal role for Apolipoprotein B-containing particles (reflected in measures such as LDL-C and total circulating TGs), although their precise contribution relative to each other continues to be debated.129-131 The central role of hyperlipidemias is also highlighted by findings that individuals with low concentrations of circulating ApoB-carrying lipoproteins have low ASCVD risk despite other risk factors.132,

Dyslipidemias are common in the population, with the prevalence of having any of several common dyslipidemias ranging from 21-37% between different ethnicities.134 They are also are highly modifiable ASCVD risk factors. Several groups of medications, including statins, ezetimibe and PCSK9 inhibitors, act to lower LDL-C and reduce ASCVD risk in a dose-dependent manner.135, 136 Recently, icosapent ethyl was shown to reduce TGs and ASCVD risk in certain high-risk individuals.137 Recent genetic studies have also highlighted APOC3 and ANGPTL4 as promising targets for therapeutic inhibition to lower TG concentrations.138, 139

In contrast, low HDL-C, though associated with increased ASCVD risk, may not be causally related to ASCVD. Mendelian randomization studies have not supported a causal role for HDL-C when adjusted for triglyceride levels, and therapeutic trials of CETP inhibition to increase HDL-C have not shown significant HDL-C-associated reductions in ASCVD risk.129, 130, 140 It still
remains unclear, however, whether ASCVD risk could be reduced by improving the function of HDL in RCT.

**Obesity**

The World Health Organization defines obesity as a BMI over 30 kg/m², and overweight as a BMI over 25 kg/m². Numerous epidemiological studies have established obesity as a risk factor for intermediate phenotypes such as dyslipidemia and high blood pressure, as well as chronic diseases, including ASCVD, type 2 diabetes, chronic kidney disease, and nonalcoholic fatty liver disease.141-145

The association of cause-specific mortality and body-mass index was estimated in 900,000 adults by the Prospective Studies Collaboration in 2009.145 In the collaborative analyses, overall mortality was lowest for those with a BMI of 22 to 25 kg/m², and was increased by 30% for each 5 kg/m² increase in BMI. The increase in vascular mortality was even higher at 40% for each 5 kg/m² increase in BMI. An even larger meta-analysis of 10,600,000 participants in 2016 observed the lowest ASCVD risk at a BMI of 20-25 kg/m² and estimated a 1.3-1.4-fold HR per each 5 kg/m² increase across regions.144

Recent genetic studies have also supported a potential causal association between obesity and CAD, stroke, and type 2 diabetes.146-149 This risk is thought to be modifiable: a meta-analysis of 54 randomized weight loss intervention trials concluded that moderate quality evidence exists of a protective effect of weight loss on cardiovascular mortality.150

The prevalence of obesity has increased in the last decades, doubling in over 70 countries since 1980.142 In the last decade, this increase has showed signs of slowing down in developed countries, but future predictions remain uncertain. In 2013, 37-38% of adults worldwide were classified as overweight or obese.143 In Finland in 2017, almost 3/4 of men and 2/3 of women were classified as overweight.151 The prevalence of obesity rose from 24% to 27% in men and from 22% to 26% in women between 2011 and 2017. The prevalence of abdominal obesity (waist circumference >100cm in men or >90cm in women) also increased from 39% to 43% in men and from 36% to 40% in women.

Concerningly, obesity is not restricted to old age. Many countries have seen faster increased in childhood obesity than in adult obesity.142 In 2009-2010, 16.9% of US children and adolescents were obese and 31.8% were either overweight or obese.152 The public health impact of such early obesity may differ from that of adulthood obesity, as not just the extent, but also the duration of exposure to obesity, appears to increase ASCVD risk.153
However, BMI has been criticized as a crude measure of obesity, in that it does not account for fat distribution, muscle mass or fitness. Further, a subset of obese individuals appears to remain free of dyslipidemia and insulin resistance, suggesting considerable differences in pathophysiology. A later chapter of this thesis will discuss more precise risk stratification in individuals with high BMI in more detail.

**Dysglycemia, insulin resistance and diabetes**

Diabetes is a well recognized risk factor for ASCVD, increasing risk of CAD by 2-fold and risk of of stroke by 1.5-1.8-fold. Worldwide, up to a third of diabetic individuals are affected by CVD. Increased risk is also seen in individuals with insulin resistance or elevated fasting glucose even without a diabetes diagnosis.

Diabetes is often associated with other cardiometabolic risk factors including dyslipidemia, obesity and high blood pressure. Some of the associated cardiometabolic states have been proposed to represent direct and causal links between diabetes and CVD, although estimation of the direction causality has proven difficult. Proposed mechanisms include direct effects of hyperglycemia and insulin resistance, diabetes-provoked dyslipidemia, increased vascular inflammation, increased susceptibility to thrombosis, and endothelial injury. Diabetes appears to be associated with both hepatic overproduction and decreased clearance of TG-rich lipoproteins (in particular, VLDL₁ particles), as well as a relative increase in small dense LDL particles and reduced HDL-cholesterol. Diabetes is also associated with increased hepatic de novo lipogenesis and steatosis, which will be discussed in more detail in a later chapter of this review. Finally, diabetes has also been proposed to exacerbate cardiac myopathy independent of vascular changes.

**Diet and exercise**

Among directly controllable lifestyle factors, diet has long been proposed to affect individual risk for ASCVD, both through direct effects on obesity and mechanisms unrelated to measures of adiposity. Changes in dietary patterns, such as increases in the consumption of processed foods, red meat and sweetened drinks, and low intake of fruits, vegetables and nuts, have been proposed to underlie changes in ASCVD risk in both developed and developing countries. Complicating strong conclusions, however, many studies have been observational with limited ability to adjust for confounding factors, and many prominent findings have failed to replicate consistently.

Nevertheless, based on a synthesis of available evidence, to reduce ASCVD risk, the European Society of Cardiology currently advocates a Mediterranean-
type diet rich in vegetables, fruit and wholegrain products and moderate intake of nuts, fish, poultry, legumes and low-fat dairy products, while also advising against red meat and sweetened drinks. The American Heart Association and American College of Cardiology recommends a diet rich in “vegetables, fruits, nuts, whole grains, lean vegetable or animal protein, and fish”, and recommends reducing intake of trans fat, sweetened drinks, processed carbohydrates and whole grains.

A sedentary lifestyle has also been associated with increased risk of CVD. Measures of cardiorespiratory fitness have been proposed to differentiate ASCVD risk in obese individuals (emphasizing a focus on fitness under the concept of “fat but fit”), although these results have also been questioned. Both European and American guidelines for the management of dyslipidemias recommend regular physical activity to reduce CVD risk.

**High blood pressure**

Hypertension – persistently elevated blood pressure – affects up to a third of adults worldwide. Despite decreases in recent decades, the prevalence of hypertension in Finland has remained relatively high among Western countries. Hypertension can accelerate the development of atherosclerosis and increases the risk of a spectrum of ASCVDs, including stable CAD, MI, and ischemic and hemorrhagic stroke.

The precise mechanisms of hypertension-associated ASCVD are still debated. Hypertension leads to vessel size -dependent hypertrophy or hyperplasia of arterial walls with proliferation of smooth muscle. It may, however, also alter the metabolism of connective tissue proteins and acid mucopolysaccharides within the arterial wall. In addition, hypertension may exert proinflammatory effects, increasing leukocyte adhesion and migration through the vascular endothelium.

The proliferation of smooth muscle, in addition to sharing features with smooth muscle proliferation in atherosclerosis, may decrease oxygen diffusion and directly promote free radical generation. Oxidative stress also seems to increase even with acute hypertension. The shared milieu of oxidative stress may conceivably underlie the interaction of hypertension and hypercholesterolemia, highlighted by the finding that the proatherosclerotic effects of hypertension are greatly reduced in the absence of hypercholesterolemia.

Lowering of blood pressure through lifestyle change or pharmacological therapy leads to a reduction in cardiovascular morbidity and mortality. Our understanding still continues to evolve regarding optimal blood pressure targets, as recently highlighted by the SPRINT trial, which suggested benefit
from more intensive treatment than previously recommended in high-risk populations.\textsuperscript{195}

**Cigarette smoking**

Cigarette smoking is a relatively recent habit in human history, becoming popularized since the late 1800s with the advent of mechanized production and mass marketing.\textsuperscript{196} In the 1940s and 1950s, evidence began to mount for a link between cigarette smoking and lung cancers. Later studies also established a dose-dependent association with increased risk of atherosclerosis, CAD, stroke and PAD.\textsuperscript{197-199} In addition, passive exposure to cigarette smoking is associated with ASCVD even in nonsmokers.\textsuperscript{200} Current evidence overwhelmingly supports a causal role between cigarette smoking and ASCVD, potentially mediated by endothelial dysfunction and proinflammatory effects.\textsuperscript{201, 202} In the US, cigarette smoking has been estimated to be the single leading modifiable risk factor for mortality and morbidity. Smoking cessation, if successful, leads to considerable decrease in CAD, stroke and overall mortality risk.\textsuperscript{203-205} In Finland, smoking prevalence has decreased considerably from the 1980s, plateauing somewhat in the recent years at approximately 13\% for men and 10\% for women.\textsuperscript{206}

From 1980 to 2012, the prevalence of cigarette smoking decreased by a quarter in men from 41\% to 31\%, and even more in relative terms in women from 11\% to 6\%.\textsuperscript{196} Nevertheless, cigarette smoking remains a major burden on global health, with the absolute numbers of smokers still rising due to population growth. Additionally, many countries have seen increasing popularity of electronic cigarettes especially among young individuals, the cardiovascular effects of which have not yet been firmly established.\textsuperscript{207}

**Inflammation**

In addition to local plaque characteristics, the role of systemic measures of inflammation in ASCVD has been established by early studies showing that elevated circulating levels of the acute-phase reactant C-reactive protein and interleukin-6 are associated with CVD events independently of cholesterol levels.\textsuperscript{208-210} Genetic studies have also supported a causal role for interleukin-6 receptor pathways in coronary heart disease.\textsuperscript{211, 212} Individuals with rheumatoid arthritis have been recognized to have increased risk of ASCVD events.\textsuperscript{213} Interestingly, in the population, the incidence of coronary events appears to peak during episodes of severe infection, such as pneumonia or sepsis, and this increased risk may last for years after the infection.\textsuperscript{214}

Whether inflammation-related risk is modifiable has been of considerable interest. Statin use appears to be associated with a reduction in CRP levels independently of LDL-C lowering. Individuals with elevated CRP may also
derive greater benefit from statin and aspirin therapy. In a recent proof-of-concept study, Ridker et al showed that canakinumab, an antibody against interleukin-1β, reduced cardiovascular events in individuals with a previous MI and elevated CRP levels. However, due to concerns of an increased risk of fatal infections associated with canakinumab, studies are ongoing to discover other, safer means of reducing inflammation-associated ASCVD risk.

### 2.1.5 Familial Predisposition

It has long been recognized that CAD, with an estimated narrow-sense heritability of 40-60%, has an inherited predisposition. Individuals with a family history of CAD have elevated CAD risk which increases even further with family history of premature CAD (CAD before the age of 55-60 in men, and 60-65 in women, depending on definition). However, family history of CAD alone does not reveal the pathogenic mechanism. Some familial risk may be transmitted by shared environmental (as opposed to inherited) factors, such as smoking or diet. Dissecting how familial risk is transmitted may allow for more precise targeting of interventions.

Familial dyslipidemias represent entities in which both ASCVD risk and measurable ASCVD risk factors appear to segregate in the same individuals within pedigrees. The best characterized of such familial dyslipidemias is familial hypercholesterolemia (FH). Considered to be an autosomal dominant disease, the prevalence of its heterozygous form has been estimated at 1:200-1:500 in many Western countries. FH is most typically characterized by familial aggregation of considerably elevated LDL-C concentrations, often present from early childhood onwards, and an increased likelihood of tendon xanthomas. Most worryingly, if not treated with LDL-C-lowering medication, individuals with FH have considerably elevated risk of ASCVD, with many patients experiencing ASCVD events before the age of 50-60. Although LDL-C levels, which are presumed to mediate ASCVD risk in FH, can be measured directly, many guidelines (including the 2016 European Society of Cardiology (ESC) / European Atherosclerosis Society (EAS) Guidelines for the Management of Dyslipidaemias and the 2017 Finnish “Käypä hoito” practice guidelines on dyslipidemias) assign a higher risk classification to individuals with FH and often recommend more intensive lipid-lowering therapy than would be warranted based on a single LDL-C measurement alone.

Another frequently observed hyperlipidemia is familial combined hyperlipidemia (FCH). In its classical form, FCH is characterized by a combination of elevations of total cholesterol (or LDL-C) and triglycerides within a single pedigree. The measurement of Apolipoprotein B concentration has also been suggested to aid in the diagnosis of FCH, but
precise definitions vary between studies, and there is considerable inter- and intraperson variability within families.\textsuperscript{24} Many diagnostic criteria additionally require premature coronary disease in at least one family member with hyperlipidemia.

FCH has been reported to be associated with considerably elevated ASCVD risk, including 5-fold elevated CAD risk in first- and second-degree relatives of probands compared with married-in spouses.\textsuperscript{230-232} However, the analyses have generally not adjusted for the degree of hyperlipidemia or isolated the effect of family history. While present in approximately 1% of the general population, up to 11-14% of individuals with premature CAD may meet criteria for FCH diagnosis.\textsuperscript{26-28} Currently, similarly to FH, Finnish practice guidelines recommend designation of high ASCVD risk status and a lower threshold for initiating lipid-lowering therapy for individuals with FCH.\textsuperscript{229} In the recent decade, EAS/ESC guidelines have moved away from precise screening and treatment recommendations for individuals with FCH.\textsuperscript{34} The 2016 EAS/ESC Guidelines still stated that “the concept of FCH is also valuable clinically in assessing CV risk”, whereas the most recent 2019 guidelines, under the section on FCH, state that “the concept of mixed dyslipidaemia is also valuable clinically in assessing CV risk” (italics added).\textsuperscript{24, 33} The revised guidelines also state that FCH has no monogenic component and is not linked to a single genetic cause, citing Study II and other research.\textsuperscript{33} Whether FCH represents a distinct disease entity similarly to FH, and to which extent it increased CAD risk, remain questions of considerable public health importance.

Many other familial dyslipidemias, including hyperlipidemias (e.g. familial high lipoprotein(a)) and hypolipoproteinemias (e.g. familial low HDL-C) have been extensively covered elsewhere, and are not within the scope of this thesis. A later chapter will discuss the genetics of FH and FCH in more detail. In Study II, we aimed to better characterize the genetic architecture of FCH. In Study III, we analyzed the CAD risk and lipidomic profiles of high LDL-C and TG traits in hyperlipidemic families, most of which had originally met criteria for FCH.

\textbf{2.1.6 PREDICTIVE MODELS}

Individuals rarely present with only one ASCVD risk factor. To collate information across risk categories, predictive models have been developed to estimate individual ASCVD risk. A 2016 systematic review catalogued 363 such models for ASCVD in the general population, a fifth of which had been externally validated by independent investigators.\textsuperscript{233} More than 100 individual predictors were included across the models, with a median number of predictors of 7 (range 2-80). The most common predictors were age, smoking, blood pressure, and blood cholesterol. Other common predictors were
diabetes, BMI, and family history of ASCVD. The most frequent prediction horizons were 5 and 10 years. National guidelines recommend the use of specific risk calculators to classify risk and aid in decisions regarding treatment initiation.\textsuperscript{24, 229}

The predictive accuracies of the scores vary, often depending on the cohort under study. A common approach for estimating discriminatory ability is measuring the area under the receive operating characteristics (ROC) curve or the C-statistic. A C-statistic of 0.5 indicates a poor model and a value of 1.0 indicates a model with perfect discrimination. In different validation cohorts, the lowest observed C-statistics for the three most commonly used score versions (Framingham (Wilson 1998), Framingham (Anderson 1991), and SCORE (Conroy 2003)) have varied between 0.57 and 0.62, and the highest observed C-statistics varied between 0.91 and 0.99, likely reflecting differences in the participants and outcomes under study.\textsuperscript{233-236}

Similar to ASCVD prediction, models have also been developed for predicting ASCVD events, such as acute coronary syndromes, in those who have already developed ASCVD. Such scores typically place a higher emphasis on clinical characteristics during hospital admissions as well as imaging measures.\textsuperscript{237, 238}

### 2.1.7 MORE DISCERNIBLE TOOLS FOR UNDERSTANDING INDIVIDUAL RISK

Despite considerable improvements in recent decades, even the best prediction tools are limited in their accuracy. It is evident that better understanding of ASCVD risk factors is needed to limit the disease burden of what still constitutes the most common cause of death in most Western countries.

Potential avenues for more detailed understanding and risk stratification of ASCVD are numerous, and not amenable to comprehensive review in a single thesis. Here, I focus on three approaches: more precise phenotyping of dyslipidemia and related circulating ASCVD risk markers, stratification of obesity based on fat distribution, and familial segregation of dyslipidemia.

The purpose of the the work included in this thesis was not to produce new risk models, but rather to provide fundamental knowledge regarding potential stratifying factors. Such understanding is a prerequisite for developing new approaches to stratify individuals with respect to risk, and possibly, for the development of targeted therapeutic approaches.
Although clinical ascertainment of ASCVD risk relies on a small subset of circulating risk markers, the human circulation contains thousands of metabolites, many of which have been implicated in ASCVD pathogenesis. On a weight basis, the vast proportion of these metabolites are lipids, but carbohydrates, amino acids and nucleic acids also constitute a large fraction (Figure 3).

**Figure 3.** Relative Distribution of Biologic Molecules in Human Plasma. Distributions are shown on a weight basis. Reproduced with permission from Quehenberger and Dennis. 2011. The Human Plasma Lipidome. The New England Journal of Medicine. Copyright Massachusetts Medical Society (2011).

### 2.2 DETAILED PHENOTYPING OF CIRCULATING ASCVD RISK MARKERS

In addition to traditionally used lipoprotein parameters such as total cholesterol (TC), LDL-C, HDL-C and total TGs, several other lipoprotein markers have been studied with respect to ASCVD risk. These include measures of apolipoproteins, lipoprotein particle number, lipoprotein size and the content of different lipoprotein fractions.

Lipoprotein particle number is typically quantified using nuclear magnetic resonance (NMR) spectroscopy or differential ion mobility analysis, whereas several methods, including biochemical assays and NMR, exist for the quantitation of apolipoprotein concentration.
In a study comparing the heritability of various lipoprotein sizes and content as measured by NMR, these parameters appeared highly heritable, and the variance explained by loci identified in genetic screens of traditional lipids differed considerably, suggesting that measurement of detailed parameters can provide biologically distinct information. For example, a similar LDL-C concentration may be observed in individuals with very different LDL particle contenttions but differing LDL particle size. The measurement of ApoB 100, the main apolipoprotein of TG-enriched lipoprotein particles secreted from the liver, reflects the particle count of VLDL, IDL and LDL lipoproteins. For HDL particles, ApoAI can be similarly measured.

The extent to which detailed lipoprotein parameters can improve ASCVD risk stratification in clinical practice, however, is uncertain. In the Emerging Risk Factors Collaboration’s prospective analysis of 165,544 individuals without baseline ASCVD, measuring ApoB and ApoAI provided only slight benefit. In a prospective study of 25,663 European individuals with elevated LDL-C, LDL particle count (LDL-P) had additional predictive value for CAD compared with LDL-C, but not when adjusted with HDL-C and TGs. Similarly, in a follow-up study of 27,673 Women’s Health Study participants, NMR measurements of lipoproteins provided comparable, but not superior, classification ability compared with traditional risk factors. Later, it was suggested that in cases of discordance between detailed lipoprotein parameters including LDL-P, the use of LDL-C alone may under- or overestimate CAD risk.

Interestingly, a recent analysis of the Women’s Health Study found that detailed lipoprotein parameters (including LDL particle count, medium and very large VLDL particle count, and HDL particle count) improved prediction of incident PAD, in contrast with poorer performance for CAD or cerebrovascular disease. In an analysis of the population-based Cardiovascular Risk in Young Finns Study, NMR-based measurement of LDL-C and medium-sized HDL particles, in conjunction with other NMR-derived parameters, modestly improved prediction of subclinical atherosclerosis beyond traditional lipid measures. Finally, several studies have suggested that among HDL parameters, HDL particle count (HDL-P) may largely explain the association between HDL-C and CAD risk; this contrasts with the aim of previous drug trials e.g. with CETP inhibitors, which have aimed to lower HDL-C.

In addition to potential improvements in prognostic accuracy, metabolic panels also allow for the profiling and contrasting of disease states. Observational cohort studies have found weight change to correlate positively with VLDL particle size, and negatively with LDL and HDL particle size. Estimates from observational studies closely match genetic effects estimated
using the method of Mendelian randomization, suggesting widespread causal influences of BMI.\textsuperscript{257} The causal effects of obesity on NMR metabolic profiles have also been estimated in the twin cohort analyzed for Study I of this thesis, showing a widespread atherogenic lipid profile associated with acquired obesity.\textsuperscript{258} The study’s findings pertaining to different adipose deposits are discussed in more detail below.

The use of novel metabolomics measures is not without limitations. Despite extensive use in population-based and patient cohorts, the biological validity of many popular NMR metabolomics measures has not been conclusively established. Databases of circulating metabolites are still not comprehensive, and interlaboratory standards are relatively lacking compared with many traditional biochemical measures.\textsuperscript{259} Although the use of replication cohorts in metabolomics studies is increasing similarly to the field of genetics, many studies – including Study I of this thesis – continue to be limited by a lack of cross-platform replication for validation of findings.\textsuperscript{260}

### 2.2.2 CIRCULATING LIPID SPECIES

The human plasma has been estimated to contain thousands of distinct lipid species, both bound and unbound to lipoproteins.\textsuperscript{239, 261} Per weight, lipids comprise the largest proportion of molecules transported in the circulation. In addition to being structural components of cellular and lipoprotein membranes, they serve to store and transport energy (most importantly between fat tissue and and energy-expending tissues in the form of fatty acids), and can also act as signaling or regulatory molecules relevant for cardiometabolic pathways. Although many studies have examined overall circulating concentrations (Figure 4), considerable differences in lipid composition exist between different lipoprotein fractions (Table 1).

Glycerophospholipids form, on a weight basis, the most abundant circulating lipid group, consisting mostly of phosphatidylcholines (PCs) and phosphatidylethanolamines (PEs).\textsuperscript{239} PCs are major constituents of HDL (> 30% of total lipid amount), and to a lesser extent of LDL (12%) and VLDL (8.5%) particles.

Sterol lipids are the second most abundant lipid group on a weight basis, and the most abundant on a molar basis. The vast majority of the sterol lipid fraction is cholesterol (mostly in esterified and to a smaller extent in free form), but other sterols also exist in plasma at low quantities (e.g. intrinsically synthesized lathosterol and desmosterol as well as the diet-derived sitosterol, which is elevated in sitosterolemia, a sterol transporter related inherited dyslipidemia predisposing to atherosclerosis).\textsuperscript{262} Cholesterol esters are a
Sphingolipids are mostly comprised of sphingomyelins, but also include ceramides and free sphingoid bases such as sphingosine-1-phosphate, a signaling molecule whose receptor is the target for fingolimod, a therapeutic for multiple sclerosis which acts by sequestering lymphocytes in lymphoid tissue. Sphingolipids are also enriched in atherosclerotic plaques compared with normal artery walls. Within the intima, they (with the exception of S1P) are associated with inflammation, and may be directly implicated in inflammation and cellular apoptosis, suggesting a role in plaque instability.\textsuperscript{265, 266}

Glycerolipids without a phosphate group consist mostly of TGs, and, at lower concentrations, of diacylglycerols (DGs). Unlike TGs which form a large part of chylomicrons and VLDL particles, DGs are mostly carried by VLDL particles, acting both as structural and signaling molecules, with potential roles in diabetes (via protein kinase C activation and consequent changes in insulin signalling), cancer and immunological disorders.\textsuperscript{267, 268} They may also be converted further to TGs or phospholipids.\textsuperscript{269}
Table 1. Lipid composition of four lipoprotein fractions. TAG = triacylglyceride. Reprinted with permission from Nicolas Christinat and Mojgan Masoodi. J. Proteome Res., 2017, 16 (8), pp 2947–2953. DOI: 10.1021/acs.jproteome.7b00236. Copyright 2017 American Chemical Society.

LPS = lipopolysaccharide, LPA = lysophosphatic acid, LPE = lysophosphatidylethanolamine, LPG = lysoglycerophosphoglycerol, LPI = lysophosphatidylinositol, LPC = lysophosphatidylcholine, PI = phosphatidylcholine, PC = phosphatidylcholine, DAG = diacylglyceride, TAG = triacylglyceride, SE = sterol ester.

<table>
<thead>
<tr>
<th></th>
<th>CM (%)</th>
<th>VLDL (%)</th>
<th>LDL (%)</th>
<th>HDL (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPS</td>
<td>0.1</td>
<td></td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>LPA</td>
<td>0.1</td>
<td></td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>LPE</td>
<td>&lt;0.1</td>
<td></td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>LPG</td>
<td>&lt;0.1</td>
<td></td>
<td>&lt;0.1</td>
<td>&lt;0.1</td>
</tr>
<tr>
<td>LPI</td>
<td>0.3</td>
<td></td>
<td>&lt;0.1</td>
<td>&lt;0.1</td>
</tr>
<tr>
<td>LPC</td>
<td>2.3</td>
<td>0.4</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>PI</td>
<td></td>
<td></td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>PC</td>
<td>3.0</td>
<td>8.5</td>
<td>11.6</td>
<td>31.5</td>
</tr>
<tr>
<td>DAG</td>
<td>0.8</td>
<td>8.3</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>TAG</td>
<td>56.9</td>
<td>59.0</td>
<td>10.0</td>
<td>6.3</td>
</tr>
<tr>
<td>cholesterol</td>
<td>2.3</td>
<td></td>
<td>2.9</td>
<td>1.4</td>
</tr>
<tr>
<td>SE</td>
<td>37.1</td>
<td>21.6</td>
<td>74.5</td>
<td>54.7</td>
</tr>
</tbody>
</table>

In addition to esterified fatty acids (FAs), free FAs (FFAs) also comprise a small fraction of circulating lipids, arising mostly from adipose tissue and bound largely to albumin. The most common species (together comprising roughly 80% of all FFAs) are the monounsaturated oleic acid (18:1) (a fatty acid with 18 carbon atoms and one double bond), palmitic acid (16:0), and stearic acid (18:0). FFAs also include polyunsaturated FAs and eicosanoids, the latter of which function as signaling molecules involved in inflammation, pain sensation and thrombosis.

To what extent measuring individual lipid species in addition to traditional lipoprotein parameters improves ASCVD prediction has not been firmly established. Studies have highlighted several lipid species – ratios of ceramide species being the most consistent predictors – as modestly improving prediction of ASCVD events over traditional risk factors.

Lipidomics has also been employed to interrogate potential disease pathophysiology in hypolipidemia. For example, a study of individuals with low HDL-C observed reduced HDL plasmalogen content and increased ceramide and sphingomyelin content within HDL particles, suggesting altered...
HDL quality in addition to low quantities of cholesterol in the HDL fraction.\textsuperscript{275} Prior to Study III, however, the relationships of circulating lipid species with common forms of hyperlipidemia were not well characterized.

Many lipidomics studies have focused on easily ascertainable blood samples, but targeted analyses in tissues of interest are also needed. For example, Borén et al have recently suggested that the sphingomyelin content of LDL may affect the aggregation of LDL particles within atherosclerotic plaques.\textsuperscript{276} The study provides a potential explanation for the association of circulating sphingomyelins with ASCVD risk.

Unlike the lipoprotein parameters discussed above, circulating lipid species are typically measured using mass spectrometry (MS) based platforms. Complicating the interpretation of individual lipidomics studies, similarly to NMR metabolomics, concerns have been raised over the validity and replicability of lipidomics measures. A 2017 National Institute of Standards and Technology (NIST) study compared analyses of a standard reference material by 31 different lipidomics laboratories, highlighting several areas in need of improvement.\textsuperscript{277} Among 1,527 unique lipids measured across the laboratories, only 339 lipids were reported by at least 5 laboratories. Considerable interlaboratory variation was observed for many lipids, with phospholipids being measured more consistently compared with other lipid classes. The authors of the NIST study recognized a need for better internal standards, and in general, quality assurance and control. These discrepancies may partly explain why different studies have highlighted different lipid species as improving ASCVD prediction.

\textbf{2.2.3 OTHER CIRCULATING METABOLITES}

In addition to lipoproteins, human plasma also contains numerous other metabolites, including but not limited to amino acids, fatty acids and carbohydrates. Many of these metabolites are associated with cardiometabolic disease states, some with predictive utility even when adjusted for conventional risk factors. These metabolites can be measured using individually standard biochemical assays or, increasingly commonly, simultaneously using high-throughput mass spectrometry (MS) or NMR profiling. On average, MS-based platforms offer better sensitivity for metabolite detection, while NMR-based platforms allow faster and more cost-effective quantitation, making analysis of large cohorts more feasible. Excepting lipids, MS-based platforms typically require chromatographic separation, while NMR can be used without separation and does not destroy the sample under study. Reflecting their relative strengths, NMR is particularly useful for discovery of novel biomarkers in large cohorts, while
Metabolite measures beyond lipoprotein parameters hold promise for improving ASCVD prediction. In a longitudinal study of 7,602 participants from diverse nationalities (European, Indian Asian and African), four NMR metabolite measures remained associated with cardiovascular risk even after adjusting for traditional lipids: phenylalanine and monounsaturated fatty acids were associated with increased risk, and docosahexaenoic acids and omega-6 fatty acids were associated with decreased risk. The study also validated its NMR measurements by comparison with MS and gas chromatography in smaller cohorts.

Metabolite profiling has been used to uncover differences in ASCVD subtypes. Unlike lipoprotein parameters, many of which are strongly associated with MI and IS but not strongly with ICH, NMR-based measures of glycoprotein acetyls appear similarly associated with ICH, MI and IS. Glycoproteins, potentially reflecting inflammation and the acute phase response, have also been associated with all-cause mortality in addition to citrate, albumin, VLDL size.

Fatty acids, amino acids and other small metabolites appear correlated with obesity (discussed in more detail below) and diabetic traits. In the Finnish METSIM study, a set of amino acids were associated with 2-hour plasma glucose and Matsuda ISI, and some (alanine, leucine, isoleucine, tyrosine, phenylalanine, and glutamine) were predictive of incident T2D. In another study of the same cohort, glycerol, FFAs, MUFAs, saturated FAs and n-7 and n-9 FAs were associated with increased risk of T2D and worsening hyperglycemia during follow-up, while omega-6 FAs were associated with reduced risk. However, adjustment for Matsuda ISI and/or glucose measurements resulted in considerable reductions in these associations, suggesting that the additive value of these biomarkers in individual diabetes risk prediction is limited.

Although metabolomics measures – quantified either by NMR or MS-based platforms – have seen increasing use in large studies of disease prediction and characterization, their ultimate clinical utility has not yet been defined. Many of the gains in prediction accuracy have been small and clinically negligible. Only few studies have replicated their findings using other metabolomics platforms. The number of biologically informative variables reported by metabolomics panels may be surprisingly low: for example, although one of the most commonly used NMR metabolomics panels reports >120 metabolite measures, over 95% of their variation can be explained by some 20 principal components, reflecting considerable intercorrelations.
Finally, even if individual biomarkers turn out to be consistently useful in prediction, and technical challenges related to their reliable quantitation are overcome, further studies are still needed to assess their effect on patient-centered outcomes. A parallel can be drawn with the high-sensitivity CRP measures as a marker of inflammation to stratify individuals with respect to ASCVD risk. Despite first having been reported to improve risk prediction over traditional risk factors over 20 years ago, the measure is still not widely used in clinical practice in Europe.\textsuperscript{284} The path to clinical translation can be longer than initially expected. Thus, at least in the short-term, the benefit of metabolomics studies may rather lie in enabling incremental advances in pathophysiological understanding.

### 2.2.4 FASTING AND POSTPRANDIAL MEASURES

Lipoprotein parameters are not static throughout the day. Most notably, they experience short-term changes in response to diet. Triglycerides are elevated for several hours after complex meals (meals composed of various nutrient groups, such as carbohydrates, fatty acids and amino acids), initially in chylomicron particles.\textsuperscript{123} These elevations appear to be markedly pronounced in the carriers of rare variants in hypertriglyceridemia-associated genes such as \textit{LPL} or \textit{APOA5}.\textsuperscript{285} This atherogenic exposure would not be captured by the standard measurement of fasting TGs alone.

The levels of many other circulating metabolites also react to meals. Reflecting known actions of insulin, the concentrations of several metabolite groups such as amino acids, fatty acids, and ketone bodies, are transiently decreased after glucose ingestion. In contrast, the concentrations of many carbohydrate metabolites (including many glycolysis products) and some bile acids are elevated.\textsuperscript{286-290} Interestingly, the changes in some metabolites (including the ketone body beta-hydroxybutyrate, the amino acid isoleucine, and lactate) appear to be altered (mostly delayed or diminished) in insulin resistant or obese individuals.\textsuperscript{289, 290}

Although the reactivity of many non-lipid metabolites to meals or pure glucose has been ascertained using extensive metabolite screens, the changes in detailed lipoprotein parameters were not well characterized prior to Study I.
2.3 STRATIFICATION OF OBESITY

2.3.1 THE CONCEPT OF HEALTHY OBESITY

It has long been recognized that not all obese individuals exhibit the atherogenic metabolic sequelae associated with BMI in population studies. To explain this finding, the concept of metabolically healthy obesity (MHO) has been proposed, and its significance and existence has since been widely debated.\textsuperscript{16} Complicating its interpretation is the large variance of employed criteria. Most commonly used criteria incorporate normal values of several or all of the following measures: blood pressure, HDL-C, TGs and fasting plasma glucose.\textsuperscript{291, 292} Often, the clinical definition of the metabolic syndrome is used. Definitions of the metabolic syndrome also vary, and often include the simultaneous presence of several cardiometabolic risk factors such as abdominal obesity, hypertriglyceridemia, low HDL-C, high blood pressure, or high blood glucose.\textsuperscript{293} Indices of insulin resistance and inflammatory markers (e.g. CRP) have been investigated as additional defining features. Depending on the criteria used, the prevalence of MHO among obese individuals is typically reported at approximately 30-40\%, but varies widely between 3\% and 57\%.\textsuperscript{11, 291}

Recent years have seen an evolving understanding of ASCVD risk in putative MHO. In a 2013 meta-analysis of eleven studies ($n = 66,556$ individuals), those with investigator-defined MHO (mostly defined by whether participants fulfilled criteria for the metabolic syndrome) had similar risk of a composite outcome of death or cardiovascular events compared with metabolically healthy normal-weight individuals.\textsuperscript{294} However, when only studies with a longer follow-up time of 10 years were considered ($n = 61,386$ individuals), MHO individuals had somewhat increased risk ($\text{risk ratio} \ (\text{RR}) = 1.24$). Still, in contrast, all groups of metabolically unhealthy individuals (normal weight, overweight, and obese) had considerably higher risks ($\text{RR} 2.65-3.14$). As an important caveat to this and many following studies, the use of the metabolic syndrome to define MHO may not adequately discriminate between those with only few cardiometabolic risk factors and those with no risk factors.

A more recent and larger meta-analysis of 22 studies ($n = 584,799$ participants) concluded that even the MHO phenotype (defined mostly by the presence of the metabolic syndrome) was associated with risk of CVD events in particular, but not with all-cause mortality.\textsuperscript{13} In a report of the European Prospective Investigation into Cancer and Nutrition study comprising 520,000 individuals with a follow-up of 12.2 years, metabolically healthy obese and overweight individuals had increased CVD event risk compared with metabolically healthy normal weight individuals ($\text{HR} 1.26-1.28$),
although this risk was again lower than in metabolically unhealthy obesity (defined as the presence of three or more cardiometabolic risk factors).\textsuperscript{14}

Whether the reported modest elevation in ASCVD risk affects all MHO patients similarly is uncertain. The stability of the phenotype may be important for long term prognosis. In the Multi-Ethnic Study of Atherosclerosis \textit{(n = 6,809 individuals)} almost half of the individuals with baseline MHO (defined based on the metabolic syndrome) developed metabolic syndrome during a median follow-up of 12.2 years.\textsuperscript{295} Only this unstable MHO phenotype was associated with increased ASCVD risk.

However, an analysis of 30 years of follow-up based on self-reported data on BMI and metabolic illnesses (updated every 2 years) among 90,257 individuals from the Nurses’ Health Study reported contradictory results.\textsuperscript{15} The study found that even women who maintained MHO (defined as no self-reported hypertension, diabetes or hypercholesterolemia) throughout the study had higher ASCVD risk than women with stable normal-weight metabolically healthy phenotypes (HR 1.57).\textsuperscript{15} Further, during the long follow-up, most (84\%) women with MHO converted to metabolically unhealthy phenotypes. The investigators noted that this conversion was not highly correlated with change in BMI during follow-up, and suggested examining more precise measures of fat distribution – in particular liver fat – as a more discerning stratifying factor. Finally, the highest ASCVD risk was observed in metabolically unhealthy obese individuals, and metabolically unhealthy normal weight individuals had similar risk compared with individuals with MHO.

These results are open to several interpretations. Firstly, there is considerable variation in the definitions of MHO, which complicates the interpretation of results across different studies. Secondly, MHO (based on most commonly used criteria) in itself does not seem to provide a warranty against future metabolic sequelae and ASCVD, and is for many a temporary phenotype. Thirdly, such MHO, if stable, may provide some, if not complete, protection from ASCVD. Fourthly, absence of obesity does not protect from the atherogenic metabolic changes more typically observed in obese individuals.

Finally, although the definition and significance of MHO remain to be disputed, it has highlighted the question of what is driving differences in metabolic health in obesity. It has been proposed that BMI alone is too crude a measure to fully capture ASCVD risk in obesity. Similarly, the metabolic syndrome, comprising many different cardiometabolic risk factors, may not best reflect individual differences in pathophysiology. To understand why some individuals suffer more severe metabolic consequences and higher risk of cardiovascular sequelae, many studies have focused not just on the \textit{quantity}, but also on the \textit{location} of fat tissue.
2.3.2 DISTRIBUTION AND MEASUREMENT OF FAT TISSUE

Several approaches have been proposed for measuring the distribution of human adiposity. Most commonly, these can be separated between regional distribution (typically quantified based on waist and hip diameter, or by using imaging technology), and compartmental distribution (quantified using imaging technology). It is no surprise that earlier studies focused on the former, as regional distribution is easily measurable in the clinical setting in large cohorts of individuals. In recent years, the increasing availability of imaging has cast more light on the significance of compartmental distribution.

It has long been recognized that adiposity can accumulate preferentially in the upper body (android fat distribution) or the lower body (gynoid distribution). These preferential localizations are more common to men and women, respectively, with less than 10% of men and women exhibiting the opposite type of distribution. An easily ascertained measure reflecting this distribution is the waist-to-hip ratio (WHR), based on circumferential measurements at the level of the waist and the hip.

Modern imaging – especially computed tomography and magnetic resonance imaging – has also allowed for localization by compartment in addition to region. Fat tissue in healthy humans is primarily located in subcutaneous adipose tissue (SCAT), which forms approximately 80% of total adipose tissue. SCAT functions as an energy reservoir, storing FAs and glycerol in the form of triglycerides and releasing them to the circulation to meet energy needs. SCAT is mostly located in three regions: the femorogluteal area, back, and the anterior abdominal wall.

A majority of the remaining fat is located in the visceral compartment, at approximately 10-20% of adipose tissue in healthy individuals. Visceral adipose tissue (VAT) is most prominently located in the omentum, the mesentery and the retroperitoneal space. Compared with SCAT, VAT appears to harbor more large adipose cells, to exhibit higher fat turnover, and to contain more inflammatory cells. Additionally, in contrast with SCAT, the majority of VAT is drained by the portal venous system through the liver.

Fat can also be present ectopically in non-adipose tissue, such as the liver, skeletal muscle and the heart. When steatosis involves more than 5% of liver parenchyma (hepatocytes) without alcohol as a predisposing factor, the term non-alcoholic fatty liver disease (NAFLD) is used. Worldwide, 25% of the population may be affected by NAFLD, with the highest prevalences (30-32%) observed in South America and the Middle East and lowest prevalences (14%) observed in in Africa, although estimates vary depending on imaging techniques. NAFLD is more common in the obese, but can also affect normal-weight individuals. NAFLD confers considerably increased
morbidity. Within the liver, the disease may progress to non-alcoholic steatohepatitis (NASH), cirrhosis and eventually hepatocellular carcinoma.\(^\text{306}\) Even without local progression, NAFLD may confer increased cardiometabolic risk, the nature of which is discussed in more detail in later chapters.

Measures of fat distribution are intercorrelated. In Western populations, BMI has been reported to correlate very strongly \((r = 0.92-0.93)\) with total fat mass and waist circumference (WC), and almost as strongly with SCAT \((r = 0.86-0.93)\).\(^\text{307, 308}\) Surprisingly, BMI and WC both appear to explain only approximately half of variation in VAT \((r_{\text{BMI}} = 0.61-0.72, r_{\text{WC}} = 0.73-0.77)\) and are thus unreliable indicators of visceral adiposity. BMI is also positively but nonlinearly associated with liver fat (LF) content, which presents in a highly skewed distribution in the population.\(^\text{309}\)

Fat distribution appears to be highly heritable. The heritability estimates for WHR range between 20-60%.\(^\text{310}\) Heritability for metrics of trunk-to-lower body distribution appear even more heritable at 60-85% (notably, with a heritability of 0.55 even after adjusting for whole body fat mass).\(^\text{311-313}\) In an analysis of the Quebec family study, the heritability of VAT was 56%, higher than that of SCAT at 42%.\(^\text{314}\) The heritability of the liver fat fraction appears somewhat lower at 38%, although differences in methodology make precise comparisons difficult.\(^\text{315}\) Genetic variation predisposing to different fat distributions have also been identified; these are discussed in a later part of the review.

Despite a heritable component, fat distribution is not static throughout life. Age is associated with an increasing volume of VAT, even after adjusting for SCAT.\(^\text{297}\) As discussed above, sex predisposes to differences in regional distribution, although these differences are reduced after menopause in women suggesting a role for hormonal influences.\(^\text{316}\) Diet, such as sweetened beverages in particular, has been proposed contribute to increased VAT and NAFLD.\(^\text{317, 318}\) Various other environmental influences, including exercise, infections, and medications, may also explain variability in LF content.\(^\text{319-321}\)

### 2.3.3 ABDOMINAL OBESITY AND ASCVD RISK

Many studies have reported that abdominal obesity (based on measures of waist and hip circumference) correlates with dyslipidemia, blood pressure and many diabetic traits independently of BMI, although not all reports have been conclusive.\(^\text{322-326}\) Use of the waist-to-hip ratio has also been reported to improve prediction of mortality and ASCVD even when adjusted for BMI, although the total benefits in prediction accuracy are modest, possibly reflecting the high correlatedness of the measures.\(^\text{327-329}\)
Given the prognostic value and easy ascertainment of WC, it is no surprise that it is often included among the diagnostic constellation of findings that comprise the metabolic syndrome.\textsuperscript{330} However, it has been argued that while useful for ASCVD risk stratification, the metabolic syndrome does not represent a distinct disease entity, and that focus should be on its constituent traits which appear independent though intercorrelated.\textsuperscript{331, 332}

In an analysis contrasting the predictive utility of traditional ASCVD risk factors in 220,000 individuals from 17 countries, both BMI and waist-to-hip ratio were no longer associated with ASCVD when adjusted for circulating lipids, systolic blood pressure, and diabetes.\textsuperscript{17} These findings suggest that ASCVD risk may be transmitted through the latter risk factors.

The association between detailed circulating metabolite profiles and abdominal obesity was explored in an earlier analysis of the twins included in Study I.\textsuperscript{258} In the study, unfavorable serum NMR profiles were more strongly associated with WC and the android to gynoid fat ratio than BMI, total fat percentage or subcutaneous fat volumes. The altered metabolites included putatively atherogenic lipoprotein measures, branched-chain amino acids (BCAAs), aromatic amino acids, glycoproteins, and the ratio of MUFAs to all FAs. The study also assessed the contributions of imaging-based SCAT, VAT and LF indices, but different sample sizes impeded direct comparison with abdominal obesity. Modelling of genetic correlation suggested that genetic variation predisposing to abdominal obesity partly overlaps with genetic variation regulating the metabolite levels. The metabolic associations were also largely overlapping with those observed for measures of insulin resistance.

\subsection*{2.3.4 VISCERAL FAT AND ASCVD RISK}

Reflecting the difficulty of imaging-based ascertainment in large cohorts, few well-powered studies have examined the association of VAT with ASCVD. In a small longitudinal study of 291 American men, visceral fat (quantified by computed tomography) was reported to predict mortality over not only BMI, WC, and SAT, but also LF.\textsuperscript{333}

In a larger longitudinal study of 1,116 American men and 1,387 American women, VAT was independently associated with MI, but only in women. This association was independent of HDL-C, hypertension and diabetes.\textsuperscript{334} A recent larger follow-up study of 2,899 older men also did not find a significant association.\textsuperscript{335} It is however unclear how studies of older individuals generalize to younger populations; for example, even LDL-C is not positively associated with incident ASCVD in some elderly cohorts.\textsuperscript{336}
Many studies have explored potential intermediate phenotypes that might contribute to ASCVD risk individuals with elevated VAT. These studies have generally concluded that VAT is associated with dyslipidemia, blood pressure, insulin sensitivity, and diabetes independently of BMI and SCAT, and that these atherogenic risk factors seem to track changes in VAT.\textsuperscript{337-345}

In addition to traditional risk factors, some studies have interrogated detailed metabolomic profiles associated with VAT. An analysis of the twins included in Study I showed that especially within monozygotic twin pairs discordant for BMI, VAT was associated with ApoB, acetoacetate, and the ratio of MUFAs to all FAs.\textsuperscript{258} In the small sample, only these associations reached statistical significance while other measures including lipoprotein parameters did not. In a larger study of 2,401 women from the TwinsUK cohort, DXA-based measurement of visceral fat was associated with several metabolite measures including BCAAs, lactate, 3-methyl-2-oxovalerate, and 4-methyl-2-oxopentanoate, the peptide HWESASXX, the steroid hormone androstenedione, and palmitate.\textsuperscript{346} Associations with the first four metabolite types were shared with T2D, and the authors postulated links to central energy metabolism as a common factor.

### 2.3.5 LIVER FAT AND ASCVD RISK

Most reported studies examining the association of LF and ASCVD are limited either by small sample sizes or by imprecise ascertainment of LF. A meta-analysis of 16 observational prospective and retrospective studies (total $n = 34,043$ individuals) found a significant association between NAFLD and a combined endpoint of non-fatal or fatal ASCVD events without adjusting for other risk factors (OR 1.64 compared with no NAFLD).\textsuperscript{347} Most studies employed ultrasound to ascertain LF and a minority used non-contrast enhanced computed tomography or liver biopsies. When only “severe” NAFLD grades (defined as steatosis with elevated serum gamma-glutamyltransferase, high hepatic 18F-fluoro-2-deoxyglucose uptake, or by high fibrosis stage on histology) were included, NAFLD remained associated with the endpoint of fatal ASCVD events.

The meta-analysis could not ascertain whether the additional ASCVD risk was mediated by other known ASCVD risk factors, or whether NAFLD provides additional information when added to current risk stratification models. Some individual studies have sought to answer this question. In a longitudinal study of 988 Finnish individuals, the association of LF with incident ASCVD was only partly diminished after adjusting for BMI and risk factors including LDL-C and SBP.\textsuperscript{348} After additionally adjusting for the QUICKI index (a measure of insulin sensitivity based on fasting glucose and insulin), LF and ASCVD were
no longer significantly associated. However, the small sample size precludes strong conclusions.

An earlier meta-analysis of 36 studies analyzing subclinical atherosclerosis contrasted NAFLD and conventional risk factors. To ascertain NAFLD, eighteen studies had employed ultrasound, seven had used liver biopsies, and five studies had used computed tomography to ascertain NAFLD. The meta-analysis found strong evidence of an association between NAFLD and measures of subclinical atherosclerosis (carotid intima-media thickness, coronary calcification, endothelial dysfunction and arterial stiffness). Further, most studies comparing NAFLD with metabolic syndrome and conventional risk factors found NAFLD to be an independent predictor of subclinical atherosclerosis. Nevertheless, the causality of NAFLD for ASCVD, as well as its relative role with respect to other measures of adiposity, remain uncertain.

The excess ASCVD risk in individuals with NAFLD has been proposed to be mediated by traditional and measurable risk factors. Multiple studies have reported that LF is associated with dyslipidemia, dysglycemia and insulin resistance independently of SAT and VAT, although LF does not explain all of the association between these risk factors and SAT. The association of LF with systolic and diastolic blood pressure appears diminished by adjustment for other traditional risk factors.

To identify potential links to ASCVD beyond traditional risk factors, studies have examined the association of LF with detailed metabolite profiles. In a prior study of the twin cohort analysed in Study I, we characterized the associations of LF with circulating NMR metabolite profiles. LF was strongly associated with LDL-C and LDL particle concentration, the ApoB/ApoA1 ratio, the ratio of MUFAs to all FAs, and acetoacetate. For LDL-related measures, the associations with LF were stronger than for intra-abdominal fat.

An earlier analysis of a small sample of the twins analysed in Study I proposed that LF, a measure with a conspicuously right-skewed distribution, could differentiate between healthy and unhealthy obesity in a dichotomous fashion. Based on conventional risk factor profiles, discordance for LF within MZ pairs appeared to strikingly delineate the metabolic and clinical characteristics of acquired obesity. Acquired obesity without a concomitant increase in LF appeared largely free of metabolic sequelae, whereas obesity with a concomitant increase in LF was associated with aberrations in traditional lipid markers, blood pressure, insulin sensitivity, and CRP. However, prior to Study I, it was unknown whether these differences would extend to more comprehensive metabolite panels.
Despite accumulating data highlighting links between LF and ASCVD risk, it remains unknown whether the quantitation of liver fat can translate to improved disease prevention and treatment. Parallels can be drawn for example with the noninvasive measurement of coronary artery calcium, which allows for reclassification of future ASCVD risk in primary prevention. Despite considerable differences in risk between those with no coronary artery calcium and those with more calcium burden, the practical benefit of calcium scanning still remains debated.24, 171 Thus, even if LF is shown to improve ASCVD risk prediction over other risk factors, the patient-centered benefit must still be established.

### 2.3.6 PROPOSED MECHANISMS UNDERLYING LIVER FAT AND CARDIOMETABOLIC CHANGES

A key feature of NAFLD is accumulation of triglycerides in hepatocytes. This appears to reflect both increased intake of free FAs from the circulation, as well as increased de novo lipogenesis in the liver. The latter appears to be increased with a diet rich in fructose; conversely, a low-carbohydrate and high-protein diet may reduce LF content and improve cardiometabolic risk profiles.355, 356 Overfeeding with saturated fatty acids (SFAs) as opposed to unsaturated FAs also appears to increase hepatic de novo lipogenesis and decrease adipose tissue lipolysis, in addition to inducing insulin resistance and increasing concentrations of circulating plasma ceramides.357 Overfeeding with simple sugars was also associated with LF content and hepatic de novo lipogenesis in the same study.

Regional differences in FA flux may also play a role even when diet is controlled. One hypothesis bridging the association of VAT with LF is that VAT leads to a “constitutively fed state” where metabolites released from mesenteric fat turnover are carried through the portal vein, where they exert more direct influences on hepatic metabolism.358

LF content and and insulin resistance are closely intertwined. The prevalence of NAFLD in individuals with T2D has been reported to range between 40% and 75%.359, 360 Even small increases in LF content are associated with hepatic insulin resistance.361, 362 Plasma insulin concentrations and adipose tissue insulin sensitivity are also associated with LF content, measured with magnetic resonance spectroscopy, in a consistent gradient.362 Individuals with NAFLD exhibit reduced suppression of free FAs and endogenous glucose production, and reduced glucose disposal during hyperinsulinemic–euglycemic clamping, a pattern similar to that observed in diabetic individuals.363 Proposed mechanisms linking LF content and insulin resistance include direct effects due to hepatic fatty acid accumulation in the liver, as well as intermediate metabolites such as DAG and ceramide species.
with putative signaling roles; however, definite mechanisms and directions of causality are yet to be established.\textsuperscript{364, 365}

2.3.7 ON TIMING AND THE DURATION OF EXPOSURE

This chapter has dealt with manifestations of obesity, which, to an extent, are modifiable risk factors throughout life. However, acquired influences interact with a person’s inherited liability for obesity and its metabolic sequelae. In many cases, several phenotypes (such as obesity, diabetes, NALFD and dyslipidemia) coexist, and the root cause is not easily traceable.

Whether knowledge of the cause of metabolic disturbances aids in ASCVD risk stratification is incompletely understood. For example, does a risk factor that appears inherited – and presumably exerts its effects starting early in childhood – confer higher risk than a similar risk factor acquired later in life? This, starting with a review of mechanisms of inheritance, is the focus of the next chapter.

2.4 GENETIC REGULATION OF OBESITY AND DYSLIPIDEMIA

2.4.1 CLASSIFICATION AND CONSEQUENCES OF GENETIC VARIATION

The human genome is both tremendously complex and varied. According to the latest statistics from the 1000 Genomes project, the average human genome has 4-5 million sites of variation compared with the reference genome.\textsuperscript{366} The vast majority of these variants are single-nucleotide polymorphisms (SNPs) or short deletions or insertions. Additionally, the average individual carries over 2000 structural variants (larger deletions, insertions and copy number variants).

The importance of any genetic variant is dependent on both its location in the genomic sequence and its functional consequence. In the coding regions of genes – regions whose base pairs translate to amino acids incorporated to protein structure – variants can be synonymous (no effect on amino acid sequence of the protein due to redundancy in the amino acid code) or nonsynonymous. Missense variant are nonsynonymous variants that incorporate different amino acids. The alternative amino acids may be similar to or highly different from each other.
A third type of coding region variant, a nonsense variant, causes a premature termination of translation and often destruction of the protein product in a process called nonsense-mediated decay. In terms of impact, nonsense variants belong to the larger group of protein-truncating variants (PTVs), which also include frameshift variants which change the reading frame of the genetic code, and variants that interfere with the splicing together of exons within a gene. Both missense variants and PTVs can have mild or moderate biological effects, but PTVs are more rare and more likely to predispose to severe disease phenotypes. Compared with only 120-180 PTVs, the average individual carries some 10,000 to 12,000 missense variants.\textsuperscript{366, 367}

Although many classical genetic studies have focused on coding variants, the vast majority of variants fall in non-coding or intergenic regions. Such variation is not necessarily neutral with respect to disease, but may affect the binding of regulatory factors or epigenetic modifications. These variants are also frequently used in genetic association studies as markers of genetic haplotypes – groups of genetic variants likely inherited together.

![Figure 5. Feasibility of identifying genetic variants by risk allele frequency and strength of genetic effect (odds ratio). Reprinted by permission from Springer Nature: Nature, Finding the missing heritability of complex diseases. TA Manolio et al. Copyright (2009). Reprinted figure originally adapted from McCarthy et al (2008).](image)

In a given population, variants can be present at different frequencies. Typical classifications used in the era of genome-wide association studies include common (minor allele frequency [MAF] > 5%), low-frequency (0.5% < MAF <
5%), rare (0.1% < MAF < 0.5%) and very rare (MAF < 0.1%) variants.\textsuperscript{368, 369} According to the 1000 Genomes project, most (over 75%) of discovered variants are rare, but per individual genome, 96-99% of variants are low-frequency or common with a MAF > 0.5%.\textsuperscript{366} Overall, the effect size of variants tends to decrease as a function of increasing variant frequency (Figure 5). Most common variants modulate disease risk only modestly. Truly high-impact variants are rare or very rare in the general population, often segregating in affected families in a Mendelian fashion.

The frequencies of disease-associated variants are highly dependent on the population under study. For example, due to relatively recent population bottlenecks, Finns have on average fewer variable loci and longer regions of linkage disequilibrium, but more loss of function variants compared with non-Finnish Europeans.\textsuperscript{370} Such differences may predispose to different disease burden but can also aid in gene discovery.\textsuperscript{371} As a point of caution, differences between populations may translate to biases when genetic models gleaned from one population are used to predict disease or trait susceptibility in another population.\textsuperscript{372}

### 2.4.2 THE EFFECT OF RARE HIGH-IMPACT VARIANTS ON OBESITY AND DYSLIPIDEMIA

Mendelian diseases, though rare, often cause such distinct phenotypes that their discovery has been feasible even with very small sample sizes. For example, although BMI is close to normally distributed in the population, rare cases of extreme obesity may have a monogenic origin. Variants in genes encoding leptin or the leptin receptor, as well as in the SH2B adaptor protein 1, a regulator of leptin signaling, may predispose to severe childhood-onset obesity via changes in the regulation of eating behavior.\textsuperscript{373-376} In addition, several genetically caused syndromes (e.g. Prader–Willi syndrome and Bardet–Biedl syndrome) exhibit obesity as one of many clinical features.\textsuperscript{377} However, these forms of obesity do not explain a meaningful proportion of obesity observed in the general population.\textsuperscript{375}

Monogenic causes have also been identified for extreme dyslipidemias. Early genetic studies were often motivated by an observation of a lipid trait’s segregation in families in patterns suggestive of Mendelian diseases. Such studies have implicated variants in over 20 genes as potential causes of extreme dyslipidemias.\textsuperscript{378}

The best genetically characterized type of familial dyslipidemias is FH. Caused mainly by variants in 3 genes (\textit{LDLR}, \textit{APOB}, and \textit{PCSK9}), monogenic FH frequently presents with elevated LDL-C and, sometimes, in external manifestations such as xanthomas. Reflecting the unique genetic history of
Finland, most putative FH variants discovered elsewhere have not been found to be polymorphic in Finns; in contrast, 5-7 locally enriched LDLR variants have accounted for most (>80%) of the genetic diagnoses within the country.\textsuperscript{379} The prevalence of these variants has been estimated at 1:600 – considerably lower than the up to 1:200 prevalence of monogenic FH reported in other countries.\textsuperscript{223, 380-382} Monogenic FH may thus be less common in Finland, but it is also possible that some Finnish FH-predisposing variants have yet to be discovered.

Untreated, the elevated LDL-C translates to accelerated atherosclerosis and increased risk of cardiovascular morbidity and mortality. Importantly, the increase in ASCVD risk in individuals with monogenic FH appears to be greater than expected from the elevation in LDL-C alone, potentially reflecting lifelong exposure to high LDL-C levels.\textsuperscript{383, 384}

Based on segregation patterns, FCH was also presumed to be a monogenic disease. Linkage studies in the 1990s and early 2000s identified several potential causative loci, as well as lipoprotein lipase deficiency caused by variants in LPL in a minority of individuals with FCH.\textsuperscript{30-32} A study in Finnish and Dutch families identified the 1q21-q23 locus as strongly linked with the condition, later refined to the region of the USF1 gene in Finnish families.\textsuperscript{385, 386} None of the single-nucleotide polymorphisms (SNPs) identified in the studies changed the amino acid code, but the risk haplotype was associated with altered expression profiles in fat tissue, including genes implicated in glucose and lipid metabolism. However, with the exception of LPL, these early findings have not translated to clinically actionable tools for monogenic diagnosis in most FCH patients.\textsuperscript{387, 388}

With improvements in sequencing technologies, the genetic diagnosis rate of putative monogenic familial dyslipidemias was expected to improve. For FH, exome sequencing has increased the frequency of actionable genetic findings in hyperlipidemic cohorts.\textsuperscript{383, 389} For other dyslipidemias the benefit has been more modest. An exome sequencing study of extreme LDL-C and HDL-C levels discovered a genetic diagnosis in only 22% of cases, identifying no novel causative genes.\textsuperscript{38} Among individuals with severe hypertriglyceridemia (TGs > 10 mmol/l), 1% were recently reported to have homozygous or compound heterozygous combinations of high-impact variants, and 14% were heterozygous for high-impact variants.\textsuperscript{390} Interestingly, both studies noted elevated polygenic burden in a large subset of the cases without monogenic diagnosis, the importance of which will be further discussed below.
In contrast with analyses of candidate loci harbouring putative high-impact variants, genome-wide association (GWA) studies seek to interrogate variant-disease associations across the whole genome, adjusting the statistical significance threshold based on the number of variants under study. Most GWA studies have focused on common variation, identifying genetic loci associated with disease. In some cases, one of the highlighted disease-associated variants is causal for the disease. Often, however, due to the non-random pattern of association between SNPs in a given loci (termed linkage disequilibrium), GWA studies highlight SNPs that serve as markers of other, more directly causal variants. Even if the precise causal variant is not identified, the association loci can highlight genes and pathways that are potentially relevant to disease pathophysiology. Results from GWA studies can also be used to estimate SNP-based heritability and genetic correlation between disease states, and in Mendelian randomization studies to assess causality.

As the power of GWA studies has increased, so too has the number of identified risk loci. Many of these loci have only small and incremental effects, but in aggregate, they can have considerable impact on individual disease susceptibility. Polygenic risk scores (PRSs), in their simplest forms, aggregate these incremental effects by summing up the number of risk alleles for each individual weighted by their effect size. Improvements in the 2010s in sample sizes, numbers of examined variants and statistical methods have yielded PRSs with increasing predictive – and potentially clinical – utility for such biologically varied phenotypes and disease as obesity, CAD, T2D, atrial fibrillation, breast and prostate cancer, and Alzheimer’s disease. For example, in recent studies, among individuals with early-onset MI, 9-10 times more individuals were noted to harbour high polygenic burden for CAD compared with those with an identifiable variant predisposing to monogenic FH (Figure 6).

Although genetic studies of obesity have not yet translated to therapeutic advances, they have cast light on the pathophysiology of obesity-related states. Association studies have identified 141 loci associated with BMI-related traits, 97 loci associated with waist-to-hip ratio adjusted for BMI, and 2 loci associated with visceral adiposity. The effect of an increased polygenic burden for BMI can already be seen in childhood; in adulthood, a BMI difference of over 10 kilograms has been observed between the extreme tails of a BMI PRS, similar in magnitude to some forms of monogenic obesity. Interestingly, genes near loci associated with BMI appear to exhibit preferential expression in the central nervous system, suggesting a role of central regulation of energy intake. In contrast, genes near loci associated
with waist-to-hip-ratio adjusted for BMI are mostly non-overlapping with BMI-associated loci and exhibit preferential expression in adipose tissue, suggesting a role in peripheral energy metabolism. Many loci associated with waist-to-hip-ratio also show sexual dimorphism, potentially underlying some of the observed sex differences in fat distribution. However, strong inferences based on these early data should be drawn with care. The precise effects of most association loci are not yet known, and future studies are likely to discover novel loci with potentially differing mechanisms.


In addition to BMI and abdominal obesity, association studies have yielded several genetic variants predisposing to NAFLD. The most robust reported association is for the gene *PNPLA3*, in which the exonic I148M missense variant (isoleucine to methionine substitution of the 148th amino acid) increases the likelihood of NAFLD manifold in homozygous carriers. The mechanism of action may be related to reduced enzymatic activity within hepatocytes, leading to impaired VLDL formation and accumulation of intracellular triglycerides. Another recently discovered risk variant is the E167K missense variant (glutamic acid to lysine substitution) in *TM6SF2*, a gene which may be necessary for normal VLDL secretion, particularly through the incorporation of TGs to ApoB-containing lipoproteins. Interestingly, these two variants do not appear to confer a dyslipidemic phenotype similar to that classically observed in NAFLD, and the *TM6SF2* E167K substitution is even associated with decreased circulating TG concentrations and reduced ASCVD risk, albeit with the tradeoff of susceptibility to liver damage. Increased BMI appears to correlate with the risk of NAFLD in the carriers of
the PNPLA3 I148M and TM6SF2 E167K variants. Conversely, at least the PNPLA3 I148M variant seems to confer sensitivity to weight loss-associated reductions in LF content.

Other risk loci have also been reported for NAFLD, including a MBOAT7 variant implicated in phosphatidylinositol metabolism. A recent study highlighted variation in the MARC1 gene that simultaneously protects against hepatic steatosis and cirrhosis and, unlike many previous genetic loci, atherogenic dyslipidemia. Although present in only a subset of individuals with NAFLD, such genetic variants may eventually highlight pathways relevant for LF accumulation in the general population.

GWA studies have identified over 150 loci affecting traditional lipid levels. Studies exploiting these associations, mainly through Mendelian Randomization analyses, have strengthened the evidence for the causality of LDL-C and TGs in ASCVD, but have not implicated HDL-C as a causal factor in ASCVD. GWA studies have also identified over 60 loci associated with detailed circulating NMR metabolite profiles (including lipoprotein parameters, amino acids, FFAs and carbohydrates) and several loci for individual lipid species, only a subset of which overlap with loci associated with traditional lipid measures.

The understanding of familial dyslipidemias was also impacted by the use of PRSs constructed based on GWAS findings. In a paradigm shift, Talmud et al. showed high polygenic burden in patients in the UK with clinically apparent FH without known genetic diagnosis. Individuals with the FH phenotype and high polygenic burden were designated as having “polygenic hypercholesterolemia”, a designation which has been proposed to enable cessation of further genetic screening. The clinical utility of the designation has not yet been fully established. Talmud et al have suggested that identifying polygenic hyperlipidemia would enable cessation of screening for high-impact variants in the family; however, some individuals exhibit both high-impact variants and polygenic burden, complicating this approach. A preclinical study suggested less severe carotid atherosclerosis in polygenic compared with monogenic hypercholesterolemia. It remains unknown to what extent CVD risk is elevated in polygenic hypercholesterolemia.

In addition to extreme LDL-C levels, high polygenic burden for lipid traits has also been implicated in other extreme and/or familial hyperlipidemias, including hypertriglyceridemia and extreme HDL-C levels. Prior to our study, however, the contribution of polygenic burden to FCH, which was initially considered to be a monogenic disease and which has components of both hypercholesterolemia and hypertriglyceridemia, had not been quantified.
Parallel to familial hyperlipidemias, polygenic burden has been found to contribute to other familial traits, such as migraine. Polygenic burden may also underlie the familial aggregation of some cancer types. Multiple studies are ongoing to assess the practical utility of polygenic scores in the prediction and risk stratification of these diseases.

Simultaneously, singular explanations are increasingly being replaced with models that account for multiple interacting influences. To be comprehensive, risk classification models would ideally integrate data across a wide range of traditional biochemical and anthropometric measures, genetic variants, other omics measures, and environmental risk factors. However, risk models must also consider the complexity of the tools used in clinical practice. Decision support systems – such as the Finnish CardioCompass tool – may help bridge this complexity and integrate data in a user-friendly format that could translate to improved care. Prediction models integrating genetic data must also carefully account for differences due to population stratification. Considerable work remains to be done in this area to improve the utility of genetic tools and to avoid the exacerbation of health disparities. Finally, the cost-effectiveness of novel diagnostic and therapeutic approaches must be evaluated and weighed against existing methods.
3  AIMS AND STUDY DESIGN

3.1  AIMS OF THE STUDIES

Despite decades of progress, dyslipidemias pose a major cardiovascular burden. Currently available prediction tools are insufficient for estimating risk for all individuals. We hypothesized that more detailed stratification of established risk factors would aid in identifying those at highest risk, who might consequently derive most benefit from preventive interventions.

In Study I, we asked whether liver fat modifies the detailed metabolic sequelae of acquired obesity. We used a unique twin paradigm to adjust for genetic effects, and employed modern phenotyping tools to quantify even small changes in liver fat content and circulating metabolites.

In Studies II and III, we asked whether common familial hyperlipidemias represent distinct disease entities like FH or merely tails in the population distribution. Specifically, we dissected their:
   a. Genetic architecture
   b. Detailed circulating lipid profiles
   c. Cardiovascular disease risk

3.2  STUDY DESIGN

To assess the metabolic sequelae of acquired obesity, Study I compared NMR metabolomic profiles between monozygotic twins (n = 60 individuals) discordant for BMI. The monozygotic twins were further split into two groups: those discordant for LF and those concordant for LF, quantified by nuclear magnetic spectroscopy. Serum samples were drawn at four timepoints during an oral glucose tolerance test (OGTT) to uncover additional metabolic variation during induced hyperglycemia.

To characterize common familial hyperlipidemias, Studies II and III contrasted hyperlipidemic Finnish families with hyperlipidemic individuals ascertained from the Finnish FINRISK population cohort.

In Study II, 715 individuals from 53 hyperlipidemic families matching the clinical criteria of FCH and 18,715 population samples with similar degrees of hyperlipidemia (high TC and/or TG levels) underwent genotyping and imputation. Known rare variants predisposing to hyperlipidemia were identified and quantified in both cohorts. To assess the contribution of
common variation to the familial and population-ascertained phenotypes, we constructed polygenic risk scores for LDL-C and TGs for each individual.

In Study III, families with an aggregation of high LDL-C or TG traits were contrasted with similarly hyperlipidemic population controls. Family members (755 members from 66 families) and population controls (19,644 individuals) were linked with national hospital discharge and cause-of-death registries to estimate their risk of incident CAD or ASCVD. In addition, to explore potential differences in pathophysiology, we assessed how high LDL-C and TG levels associated with lipidomic profiles composed of 151 circulating lipid species in 550 members from 73 hyperlipidemic families and 897 individuals from the population cohort.
4 METHODS

4.1 STUDY SUBJECTS

The protocols of Studies I-III were approved by the ethics committees of the participating centers. All samples were collected in accordance with the Helsinki declaration. All subjects provided written informed consent.

4.1.1 BMI-DISCORDANT TWINS

Participants for Study I were identified from the FinnTwin12 and FinnTwin16 cohorts (combined $n = 10,384$ individuals) based on their responses to questions concerning weight and height in young adulthood (ages 20.0 to 36.7 years). The mean (± standard deviation) BMI in the FinnTwin12 and FinnTwin16 cohorts was 23.7 (± 3.4) (range: 14.0 to 49.4 kg/m$^2$). Pairs in which one or both cotwins had major somatic or psychiatric diseases, eating disorders, anemia, regular medication (excluding contraceptives), or recent weight cycling (±5 kg change within the past 3 months) were excluded. From the remaining sample, all monozygotic (MZ) pairs (30 pairs, 11 male and 19 female) discordant for BMI (intrapair $\Delta$BMI ≥ 3 kg/m$^2$) were identified. Additional pairs were recruited from FinnTwin16 as controls using a random generator. These included 34 (22 female and 12 male) BMI-concordant (intrapair $\Delta$BMI < 3 kg/m$^2$) MZ pairs, 38 (22 male and 16 female) dizygotic (DZ) BMI-discordant DZ pairs, and 35 (19 male and 16 female) BMI-concordant DZ pairs. In total, 137 twin pairs were included in the study (age range: 22.8 to 36.2, 54% male). The mean BMI of all study subjects was 25.3 (± 4.6) (range: 16.3 to 48.6 kg/m$^2$). 43% of the female study subjects used oral contraceptives. One twin diagnosed with ulcerative colitis and treated with mesalazine and azathioprine was included in the analyses; all other recruited individuals were healthy.

4.1.2 DYSLIPOIDEMIC FAMILY COLLECTION (EUFAM)

The dyslipidemic families for Studies II and III were identified as part of the European Multicenter Study on Familial Dyslipidemias in Patients with Premature Coronary Heart Disease (EUFAM). The probands of the families were identified in participating university hospitals based on premature CAD and simultaneous dyslipidemia (TC, TGs or both ≥ 90th Finnish age- and sex-specific population percentile; or HDL-C (≤ 10th percentile)). If at least one other first-degree relative with a similar dyslipidemia was identified, the family was included in the EUFAM study. The probands were screened for
likely FH with an in-house functional low-density lipoprotein receptor (LDLR) test similar to a test reported by Cuthbert and colleagues. Probands diagnosed with presumptive FH were excluded from subsequent analyses and did not contribute to designations of familial hyperlipidemia among the remaining family members. Individuals with known pre-defined comorbidities (diabetes, hepatic disease, renal disease, thyroid disorders, or malignancy) and pregnant women were similarly excluded from analyses prior to family definitions. Individuals using lipid lowering medication or sex hormones were additionally excluded from samples selected for lipidomic analyses.

Study II included all EUFAM families which met the following definition of FCH: The families had a proband with simultaneous premature CAD and TC or TG ≥ 90th percentile, and at least one first-degree relative with TC or TGS ≥ 90th percentile. Additionally, at least one family member was required to have TGS ≥ 90th percentile to exclude families with isolated hypercholesterolemia. A total of 715 members of 53 FCH families had DNA available and underwent genotyping and imputation.

Study III included all EUFAM families with aggregation of either high (≥ 90th percentile) LDL-C or TG levels in at least two first-degree relatives. Families were designated as “high LDL-C families” or “high TG families” based on the elevated lipid measure. A pedigree could receive both designations if both criteria were simultaneously fulfilled. A total of 755 members from 66 families could be linked with registry data and were free from CAD at baseline. Lipidomics phenotyping of 151 circulating lipid species was also performed for 550 members of 73 families with available serum samples as detailed below.

4.1.3 FINRISK POPULATION SURVEY

Study II and III included participants from the Finnish National FINRISK study as a population-based comparison group. The FINRISK study includes population surveys conducted every 5 years since 1972. Samples and questionnaire data from subjects in the 1992-2012 surveys are stored in the biobank of the National Institute for Health and Welfare (THL).

Participants were excluded from the FINRISK study based on known diabetes, cancer or pregnancy. Individuals using lipid lowering medication were additionally excluded from samples selected for lipidomic analyses. For Study II, a total of 18,715 remaining FINRISK samples were genotyped and imputed and used as controls in genetic analyses. FINRISK individuals meeting the same FCH lipid criteria as were used in the EUFAM cohort were defined as hyperlipidemic. For Study III, all FINRISK individuals who passed exclusion criteria were linked with registry data. A total of 19,644 individuals without
CAD at baseline were included in analyses of incident CAD and ASCVD risk. Additionally, samples from 1141 individuals (897 of whom passed exclusion criteria) from the FINRISK 2012 cohort underwent lipidomics phenotyping.

4.2 CLINICAL MEASUREMENTS AND LABORATORY METHODS

4.2.1 CLINICAL MEASURES AND BLOOD SAMPLING

In Study I, subjects underwent an OGTT after a 12h overnight fast. Venous blood samples were drawn at 0, 30, 60, and 120 minutes during the OGTT and inoculated into fluoride citrate tubes (for glucose) and serum (for insulin and NMR metabolite measurements). Standard methodology was used for measurement of glucose and serum insulin.

Participants in Study I also underwent measurement of anthropometric traits (height and waist circumference), blood pressure (mean of three supine measurements), body composition and abdominal fat volumes. Dual-energy X-ray absorptiometry was used to estimate total fat percentage, android fat mass and percentage, and gynoid fat mass and percentage.

For the BMI-discordant MZ cotwins, subcutaneous, and intra-abdominal fat volumes were additionally quantified from 16 slices of T1-weighted images from a 1.5 Tesla magnetic resonance imaging scanner. Liver fat percentage was estimated using magnetic resonance spectroscopy with the same magnetic resonance imaging scanner. The BMI-discordant MZ twins were further divided into two LF-discordant pairs (n = 24 individuals) and LF-concordant pairs (n = 26 individuals) based on a median difference of > 2% in LF content (Figure 7).

In Study II and III, EUFAM individuals were instructed to fast overnight before drawing of venous blood samples. Serum and EDTA plasma were separated by centrifugation. FINRISK individuals were instructed to fast for four hours prior to blood sampling and to avoid heavy meals earlier during the same day.

Traditional lipid measures were obtained from serum samples from the EUFAM study as follows: Serum TC and TG with an automated Cobas Mira analyzer (Hoffman-La Roche, Basel, Switzerland) by fully enzymatic methods (Hoffman-La Roche kits #0722138 and #0715166, respectively), and serum HDL-C with phosphotungstic acid/magnesium chloride precipitation procedures (Hoffman-La Roche kit #0720674). Measurements were obtained
from the plasma samples of FINRISK individuals using enzymatic assays. The Friedewald formula was used to estimate LDL-C concentration.

Figure 7. Liver fat content for the leaner and heavier cotwin of BMI-discordant pairs in Study I. Twins with >2% difference in LF content were classified as LF-discordant. Reprinted with permission from Rämö et al., Liver Fat and Insulin Sensitivity Define Metabolite Profiles During a Glucose Tolerance Test in Young Adult Twins, The Journal of Clinical Endocrinology & Metabolism, 2017, Volume 102, Issue 1, Pages 220-231, by permission of Oxford University Press.

4.2.2 GENERATION OF NMR METABOLOMICS DATA

In Study I, serum samples underwent centrifugation and freezing to -80°C, and were subsequently analyzed with a high-throughput NMR metabolomics platform. The platform quantified lipoprotein subclass distributions, fats and lipids, tricarboxylic acid cycle intermediates, carbohydrates, amino acids and glycoproteins. Sample preparation was performed automatically using previously established protocols. Data were acquired using separate molecular windows optimized for either lipid molecules or low molecular
weight metabolites. Data were processed and phase corrected following automated protocols. Eighty resulting metabolite measures were selected for Study I prior to analyses, focusing on metabolites previously associated with dyslipidemia, glycemic traits or cardiovascular risk.

### 4.2.3 GENERATION OF LIPIDOMICS DATA

For Study III, lipidomics data was initially generated for 353 individuals from the EUFAM cohort. To assess the quality of lipidomics data, lipid measures were quantified using two separate lipidomics platforms. Platform 1 (an independent research platform based at the Massachusetts General Hospital, Boston, Massachusetts, USA) employed liquid chromatography/mass spectrometry (LC/MS) based lipid profiling to quantify more than 100 lipid analytes from the TG, diglyceride (DG), cholesterol ester (CE), lysophosphatidylcholine (LPC), phosphatidylcholine (PC), lysophosphatidylethanolamine (LPE) and sphingomyelin (SM) classes to the level of acyl chain carbon number and double bond content. Platform 2 (Lipotype GmbH, Dresden, Germany) was a high-throughput shotgun lipidomics platform that employed direct infusion in a mass spectrometer to analyze samples in both positive and negative ion modes in a single acquisition, quantifying over 300 lipid species from diverse classes.

After initial quality control (detailed in the Results section), platform 2 was selected for further analyses and lipidomics phenotyping was extended to a total set of 550 EUFAM family members and 1142 individuals from the FINRISK 2012 cohort. The median coefficient of variation was <10% in each batch. A total of 151 lipid species were detected in at least 80% of samples in both cohorts and were selected for subsequent analyses.

### 4.2.4 GENOTYPING AND IMPUTATION

To confirm the zygosity of the twin pairs included in Study I, genotyping of ten informative genotype markers had been performed previously.

For Studies II and III, EUFAM and FINRISK samples were genotyped with different arrays, either the HumanCoreExome BeadChip, the Human610-Quad BeadChip, the Affymetrix6.0, or the Infinium HumanOmniExpress (Illumina Inc., San Diego, CA, USA). Genotype calls were generated and their quality was controlled at the Institute for Molecular Medicine Finland jointly with other simultaneously available data sets. Phasing of the samples was performed using SHAPEIT (version 2) and imputation using IMPUTE (version 2.3.1). The imputation panel was based on 1000 Genomes Phase 1 integrated haplotypes and an in-house reference panel from 1941 whole
genome sequenced Finnish individuals from population cohorts (FINRISK and Health 2000).

In Study II, 212 SNPs either reported as lead-SNPs associated with LDL-C or TGs in recent genome-wide association studies (171 SNPs) or catalogued in the Online Mendialian Inheritance in Man (OMIM) database within genes potentially involved in primary or secondary hyperlipidemias (44 SNPs) were genotyped or imputed. In Study III, 87 lead variants from published GWA studies of LDL-C and 74 lead variants from GWA studies of TGs were genotyped and imputed for all individuals with lipidomics data available.

4.2.5 HEALTH REGISTRY DATA

In Study III, incident CAD and CVD diagnoses were collected from the National Finnish Hospital Discharge Register and the National Causes-of-Death Register. International Classification of Diseases (ICD)-based diagnoses from these registers have been previously validated.\textsuperscript{437, 438}

CHD was defined as fatal or nonfatal MI or cardiac revascularization (percutaneous angioplasty or coronary artery bypass graft surgery). Codes included for myocardial infarction were as follows: for hospital discharge, ICD-10 codes I20.0 or I21-22, ICD-9 codes 410 or 411.0, and ICD-8 codes 410 or 411.0; for main cause-of-death, ICD-10 codes I21-25, I46, R96, or R98, ICD-9 codes 410-414 or 798 [excluding 7980A], and ICD-8 codes 410-414 or 798.

CVD additionally included stroke, including codes for both hospital discharge and main cause-of death as follows: ICD-10 codes I61 or I63-64 [excluding code I63.6 for subarachnoid hemorrhage]; ICD-9 codes 431, 433.0, 433.1, 433.9, 434.0, 434.1, 434.9, or 436; or ICD-8 codes 431 [excluding codes 431.01 and 431.91 of the Finnish adaptation of ICD-8], 433, 434, or 436.

4.3 STATISTICAL METHODS

4.3.1 TWIN METHODOLOGY

In Study I, to adjust for potential genetic confounding (genetic variation affecting several variables of interest simultaneously), intra-pair differences between cotwins were calculated for both predictor (e.g. liver fat percentage) and outcome variables (e.g. Area Under the Curve values for metabolites as explained below). As MZ twins share 100% of their genetic variation, a previously observed population-based association between two variables would be expected to be absent when analyzing within-pair values if the
association was completely due to genetic confounding. In DZ pairs, the strength of the association would be expected to be diminished with respect to the association observed in the population, but not entirely absent.

4.3.2 CHANGES DURING AN OGTT AND AREA UNDER THE CURVE MEASURES

Given that peak changes in metabolite levels have been shown to occur at different times during and after a 2-hour OGTT, the time point of maximal change was estimated separately for each metabolite. Absolute and percentual changes from baseline levels were then calculated relative to the time point of maximal change. A corrected p-value threshold for metabolite excursions was established by using a principle component analysis-based Bonferroni correction (0.05 divided by 20 – the amount of principal component analysis factors that explained 95% of observed variance in all 80 metabolomics measures).

To capture exposure to a metabolite during the whole OGTT, Area Under the Curve (AUC) values were calculated using data from all timepoints using the trapezoid rule. A nominal significance threshold of 0.05 was used for comparison of metabolite AUCs within and between LF-discordant and LF-concordant pairs.

4.3.3 VARIANT ENRICHMENT ANALYSIS

In Study II, the allele frequencies of 212 lipid-related SNPs were estimated separately in 234 affected FCH family members and 18,715 FINRISK samples, designating the minor allele in FINRISK as the effect allele. The frequencies were divided to yield an enrichment ratio for each SNP. Statistical significance was evaluated under the hypothesis of no enrichment by contrasting each SNP’s enrichment ratio with a 95% credible interval. The credible interval was established in 15 frequency bins by calculating enrichment ratios for all variants across the genome with a MAF >0.001%, excluding variants at known lipid loci.

4.3.4 POLYGENIC SCORES

In Study II, polygenic scores for LDL-C and TGs were calculated for each individual, based on all OMIM SNPs and GWAS lead loci associated with the lipid measure in question. The scores were calculated as the sums of the effect
alleles weighed by the alleles’ effect estimates from a multiple linear model predicting the lipid measure in FINRISK.

4.3.5 PROPORTIONAL HAZARDS MODEL

In Study II, risk of incident CAD or CVD was estimated with the Cox proportional hazards model using age as the time scale. The model was stratified by sex and clustered by family. Individuals with prevalent CAD or CVD were excluded from the analyses of CAD and CVD, respectively. Additional analyses included self-reported smoking status and use of lipid lowering medication at baseline as covariates.

4.3.6 LIPIDOMICS ANALYSES

To estimate the narrow-sense heritability of lipid classes and species, the lipid parameters were first inverse-normalized separately for men and women, and heritability was estimated using linear mixed models with age as a covariate and an empiric genetic correlation matrix as the random effect.

For all subsequent analyses, right-skewed measures were natural logarithm transformed and all species were normalized by standard deviation. Linear mixed models were used to estimate the association between lipid species (or saturation metrics) and predictors (hyperlipidemia status, continuous lipid measurement, or genotype). Other fixed covariates included sex, age, and age². As the lipid species were quantitated in two batches for the EUFAM cohort, we analyzed the cohorts separately and combined the results using fixed effects inverse variance weighted meta-analysis (as implemented in the R package ‘metafor’). P-values were calculated using the Wald test, and statistical significance was evaluated by using the Benjamini-Hochberg method at the 5% false detection rate (FDR) to account for multiple comparisons across the 151 lipid species.

4.3.7 CORRECTION FOR RELATEDNESS

In Study I, when establishing significant metabolite excursions in the whole study sample, survey regression as implemented in the R package ‘survey’ v.3.30 was used to account for the relatedness of cotwins.439

In Study II and III, all linear mixed models included an empirical genetic relationship matrix as the covariance structure of a random effect to account
for relatedness. Additionally, in Study III, Cox proportional hazards models were clustered by family.

4.3.8 STATISTICAL SOFTWARE

Linear mixed models were applied with MMM (version 1.01). R (versions 3.2.1-3.4.3) were used for data transformations and other analyses.
5 RESULTS AND DISCUSSION

5.1 METABOLOMOMIC PROFILES OF ACQUIRED CHANGES IN BMI AND LIVER FAT (STUDY I)

In Study I, we characterized the metabolomic profiles associated with acquired increases in BMI. Further, to identify the contribution of LF to circulating ASCVD risk markers, we separated acquired increases in BMI into two subphenotypes based on whether an increase in LF was also present. To uncover additional variation in circulating metabolites, we drew blood samples at four time points during an OGTT.

In addition to previously established metabolite excursions during an OGTT, we observed considerable changes in lipoprotein parameters, most notably for VLDL- and TG-related measures. We observed widespread metabolite differences between BMI-discordant cotwins. Differences in measures linked with ASCVD risk were more pronounced and widespread between cotwins discordant for LF than between cotwins concordant for LF. The metabolites strongly associated with LF overlapped only partly with measures of insulin sensitivity.

5.1.1 REACTIVITY OF CIRCULATING BIOMARKERS TO ORAL GLUCOSE TOLERANCE TEST

We began by characterizing the reactivity of circulating NMR metabolites to an OGTT in the larger sample of 274 twins. The levels of 54/80 metabolites changed after ingestion of glucose (p < 0.0025, Figure 8).

To assess the biological validity of the NMR platform, we first focused on metabolites known to react during an OGTT. In line with the suppression of release of FAs and lipolysis in subcutaneous adipose tissue, we observed decreases in the levels of most FAs and glycerol. Insulin also inhibits ketone body generation and increases their clearance; we saw decreases of 32-40% in ketone bodies (acetoacetate, acetate, and 3-hydroxybutyrate). Finally, we saw considerable decreases (7-47%) in the levels of most circulating amino acids, as expected after insulin-stimulated intake into muscle cells.
We then focused on metabolites whose reactivity to an OGTT was not previously known, including mostly lipoprotein-related measures (Figure 9). We observed decreases in the levels of total TGs (-10.6%) and VLDL-TGs (-13.7%), potentially reflecting insulin-mediated lipid intake to peripheral tissues and inhibition of VLDL secretion from the liver. The relative TG content in the extremes of HDL particle sizes increased, most evident at 60 minutes for very large (+18.3%) and small (+8.2%) HDL particles.

**Figure 8.** Excursions during the OGTT for key metabolite measures. Median values of percentual excursions from baseline to each time point in the whole study sample (n = 274) for key metabolites. Only measures with significant changes from baseline to peak change (P < 0.0025) are included. For the figure, individual measures were summed for ketone bodies, FAs, and structural amino acid subgroups. Reprinted with permission from Rämö et al., Liver Fat and Insulin Sensitivity Define Metabolite Profiles During a Glucose Tolerance Test in Young Adult Twins, The Journal of Clinical Endocrinology & Metabolism, 2017, Volume 102, Issue 1, Pages 220-231, by permission of Oxford University Press.
Figure 9. OGTT-induced changes in lipoprotein measures. All individuals from the twin cohort (N=274) were used for analysis. Changes from baseline in SD units for each timepoint in the whole study sample (light brown = 30min, middle brown = 60min, dark brown = 120min). Significance of the difference between baseline level and the peak level that most differed from baseline during the OGTT was estimated using survey regression with age and sex as covariates. Asterisks denote significance based on multiple corrected thresholds (*=p<0.0025, **=p<0.0005, ***=p<0.00005). Only measures with significant changes are shown. TG = triglycerides, -C = cholesterol, VLDL = very low-density lipoprotein, IDL = intermediate-density lipoprotein, LDL = low-density lipoprotein, HDL = high-density lipoprotein. Reprinted with permission from Rämö et al., Liver Fat and Insulin Sensitivity Define Metabolite Profiles During a Glucose Tolerance Test in Young Adult Twins, The Journal of Clinical Endocrinology & Metabolism, 2017, Volume 102, Issue 1, Pages 220-231, by permission of Oxford University Press.

The fall in VLDL and TG measures is surprising given that complex meals typically increase serum TGs for several hours. This might reflect a more isolated effect of insulin in inhibiting secretion of TGs from the liver and in directing TGs to peripheral fat tissue (by increasing fat tissue lipoprotein lipase activity and decreasing skeletal muscle lipoprotein lipase activity).
5.1.2 ACQUIRED BMI-DISCORDANCE WITH OR WITHOUT LF DISCORDANCE: TWO DISTINCT SUBGROUPS

Next, we examined the clinical characteristics of an acquired increase in BMI in two subgroups: MZ twins discordant (n = 24 individuals) and concordant (n = 26 individuals) for LF. First, we compared the heavier twins from both subgroups and saw that in addition to increased LF, the heavier twins of the LF-discordant pairs also had increased intra-abdominal fat (p<0.05) compared with their leaner cotwins, but not higher subcutaneous fat in MRI imaging or altered android vs. gynoid fat ratios in DEXA scans. The heavier twins from the LF-discordant group also exhibited higher circulating transaminases and lower indices of insulin sensitivity compared with both their leaner cotwins and the heavier cotwins of LF-concordant pairs.

Analysis of potential confounders showed similar self-reported physical activity (estimated using the Baecke sport index) and alcohol consumption between the heavier twins from both subgroups.

5.1.3 FEW METABOLOMOMIC DIFFERENCES BETWEEN BMI-DISCORDANT MZ COTWINS CONCORDANT FOR LF

Within MZ pairs discordant for BMI but concordant for LF, we observed only small within-pair differences in metabolomic AUCs during the OGTT (Figure 10 and 11). Among lipoprotein parameters, HDL cholesterol was decreased (-13.8%, p = 0.025) in the heavier vs. leaner cotwins. This difference was most marked for larger HDL particle fractions and was also reflected in a smaller HDL diameter (-1.44%, p = 0.015). Additionally, we observed elevations in VLDL cholesterol (+11.5%, p = 0.035) and Apolipoprotein B (+7.9%, p = 0.022). Among small metabolites, glycerol (+23.2%, p = 0.0091), valine (+12.3%, p = 0.015), and tyrosine (+12.5%, p = 0.046) AUCs were increased in the heavier cotwins.

A previous analysis in a smaller sample of these twins (n = 16 individuals) did not uncover significant differences in traditional lipid measurements. Here, we report differences in HDL-C, VLDL-C and Apolipoprotein B, suggesting that acquired BMI, even when not accompanied by LF, is not free of metabolic sequelae associated with ASCVD risk.
Figure 10. Dynamic metabolite levels in subtypes of BMI discordance. Mean ± SE values of selected NMR measures are presented for 4 groups at each time point during the OGTT: cotwins with higher BMI within LF-discordant MZ pairs (solid line with black squares, n = 12), cotwins with lower BMI within LF-discordant MZ pairs (solid line with white squares, n = 12), cotwins with higher BMI within LF-concordant MZ pairs (dashed line with black triangles, n = 13), and cotwins with lower BMI within LF-concordant MZ pairs (dashed line with white triangles, n = 13). Student’s t test was used to estimate significance between the AUC values of the matched cotwins with higher and lower BMI (paired testing) and between the AUC values of the two groups of twins with higher BMI (nonpaired testing). P values are marked by asterisks next to the symbols of the corresponding groups. *P < 0.05; **P < 0.01. Reprinted with permission from Rämö et al., Liver Fat and Insulin Sensitivity Define Metabolite Profiles During a Glucose Tolerance Test in Young Adult Twins, The Journal of Clinical Endocrinology & Metabolism, 2017, Volume 102, Issue 1, Pages 220-231, by permission of Oxford University Press.
We repeated the analysis in MZ pairs discordant for both BMI and LF, and saw more widespread and marked differences in metabolite AUCs (Figure 10 and 11). Among lipoprotein parameters, the heavier cotwins had markedly elevated AUCs for total TGs (+87.5%, small VLDL particles (+30.5%, \(p = 0.037\)), and small LDL particles (+24.7%, \(p =0.033\)), as well as reduced HDL2 cholesterol AUC (-17.9%, \(p = 0.045\)). Among small metabolites, AUCs for monounsaturated FAs (+37.5%, \(p =0.027\)), isoleucine (+33.2%, \(p = 0.028\)) and glycoproteins (+12.0%, \(p =0.0064\)) were also increased in the heavier twins of LF-discordant pairs. Although isoleucine and monounsaturated FAs decreased during the OGTT, their levels remained higher in the heavier vs. leaner cotwins even at the end of the test at 120 minutes (\(p < 0.05\)) (Figure 10).

Compared with the heavier twins of LF-concordant pairs, the heavier twins of LF-discordant pairs also exhibited higher AUCs for Apolipoprotein B (+15.9%,

**Figure 11.** Dynamic metabolite levels in subtypes of BMI discordance (continued). Reprinted with permission from Rämö et al., Liver Fat and Insulin Sensitivity Define Metabolite Profiles During a Glucose Tolerance Test in Young Adult Twins, The Journal of Clinical Endocrinology & Metabolism, 2017, Volume 102, Issue 1, Pages 220-231, by permission of Oxford University Press.

5.1.4 WIDESPREAD METABOLOM IC DIFFERENCES BETWEEN BMI-DISCORDANT MZ COTWINS DISCORDANT FOR LF
Our findings are in line with cross-sectional associations published nearly concurrently from the The Cardiovascular Risk in Young Finns Study.\textsuperscript{445} In the cross-sectional study of 1,939–2,002 individuals, employing the same metabolite panel, after adjusting for BMI, ultrasound-quantified LF was most strongly associated with VLDL particles in all except very small size fractions, serum and VLDL TGs, MUFAs and SFAs, leucine, and isoleucine. The study also reported numerous additional metabolite associations suggestive of a widespread atherogenic profile linked with LF even after adjusting for BMI.

Previous studies have also reported a connection between LF and detailed FA parameters. TGs with low carbon number and double bond content appear to be enriched in liver biopsies of NAFLD patients and in the VLDL particles of individuals with insulin resistance.\textsuperscript{23, 446} Increased liver content of saturated and monounsaturated TGs and free fatty acids, ceramides and dihydroceramides appear to characterize NAFLD with concomitant insulin resistance compared with NAFLD associated with the \textit{PNPLA3} I148M variant ("metabolic" vs. genetic NAFLD subtypes).\textsuperscript{447}

Studies of circulating lipid profiles in individuals with NAFLD have also noted increases in TGs with low carbon number and lipid species containing SFAs and MUFAs, as well as decreases in species containing PUFAs and ether lipids.\textsuperscript{20, 21} In a validation cohort of one of these studies, a combination of three lipid species (PC(O-24:1/20:4), PC(18:1/22:6) and TG(48:0)) had a sensitivity of 69.1% and a specificity of 73.8% for diagnosing NAFLD.\textsuperscript{20} In the other study, TG 50:1, diSM 18:0, CE 20:3, SM 18:0, and TG 52:1 emerged as the most important lipid biomarkers of liver fat.\textsuperscript{21} Although precise species vary between studies, it has been proposed that these overall findings may reflect increased secretion of VLDL particles carrying saturated TGs due to accelerated de novo lipogenesis in the liver. On the other hand, NAFLD is also associated with a reduction in HDL particles, which appear enriched for many of the lipid species negatively associated with NAFLD.\textsuperscript{382-384}

Hepatic overproduction of VLDL-TG and ApoB in NAFLD has previously been reported.\textsuperscript{448} Altered VLDL and TG parameters in our study are in line with this hypothesis. The subsequent generation of smaller VLDL remnants and small LDL particles, which penetrate arterial walls easier, may predispose to accelerated atherosclerosis.\textsuperscript{123}

Our findings also suggest that obesity with concomitant LF may confer metabolic sequealae and risk of morbidity even beyond traditional lipoprotein measures. The heavier twins with increased LF had elevated levels of circulating glycoproteins, a marker of all-cause mortality in a previous
We saw differences in saturated and monounsaturated FAs, which have previously been associated with diabetes and ASCVD risk over traditional risk factors, and in isoleucine, which has been associated with incident diabetes risk. These differences persisted between the heavier and leaner twins throughout the OGTT despite considerable negative excursions in the metabolites, suggesting that insulin resistance in individuals with LF may contribute to the observed metabolite differences.

Our analyses have several limitations. The sample size was small, and although we could contrast cotwins with each other, the study design was cross-sectional. Inferences with regards to any individual lipid measure should be drawn with care, as our results are liable to both over- and underestimation of true effect sizes. Although we adjusted for genetic differences through analyses of cotwins, we could not analyze genetic variation potentially contributing to differences in the risk of LF accumulation between twin pairs.

### ASSOCIATIONS OF METABOLOMIC MEASURES WITH INDICES OF INSULIN SENSITIVITY

Many metabolite changes in NAFLD have also been associated with dysglycemia and diabetes, and – although the direction of causality remains unclear – hepatic insulin resistance has been proposed as a key feature of NAFLD. To address this connectedness, we examined the extent to which various OGTT-derived indices of insulin sensitivity and secretion correlated with the metabolic alterations seen with LF discordance.

In addition to fasting insulin, two insulin sensitivity indices were strongly associated with metabolite AUCs: the Homeostasis model assessment index of insulin resistance (HOMA-IR), and adipocyte IR (Figure 12). The three parameters were associated ($p < 0.0025$) with total and VLDL-TG ($r = 0.53–0.67$), cholesterol in large and medium VLDL particles ($r = 0.13–0.22$), monounsaturated FAs ($r = 0.48–0.64$), and saturated FAs ($r = 0.46–0.60$). Among LDL particles, only smaller particle sizes were associated with insulin sensitivity indices ($r_{\text{HOMA-IR}} = -0.41$; $r_{\text{Adipocyte IR}} = 0.44$). We observed only weak associations between HDL particles and insulin sensitivity. Adipocyte IR was additionally associated with isoleucine ($r = 0.47$). The Matsuda insulin sensitivity index and the Liver IR index were most notably associated with VLDL and TG measures ($r_{\text{Matsuda}} = -0.47$; $r_{\text{Liver IR}} = 0.43–0.44$ for total and VLDL TGs). Two indices of beta cell function, the insulin secretion index and the oral disposition index were not significantly associated with any AUC measures.
Figure 12. Associations between intrapair differences in insulin-based indices and the AUC values of metabolites in all 64 monozygotic twin pairs. (Figure legend is continued on next page)
Intrapair differences were calculated by subtracting the measurement of the cotwin with lower BMI from the measurement of the cotwin with higher BMI. Indices and AUC values were rank-normalized prior to analysis while adjusting for sex and age, and Pearson correlation coefficients were estimated. Color denotes the strength of correlations (red = positive, blue = negative) and asterisks mark their significance (*P < 0.0025; **P < 0.0005; ***P < 0.00005). IDL, intermediate-density lipoprotein. Reprinted with permission from Rämö et al., Liver Fat and Insulin Sensitivity Define Metabolite Profiles During a Glucose Tolerance Test in Young Adult Twins, The Journal of Clinical Endocrinology & Metabolism, 2017, Volume 102, Issue 1, Pages 220-231, by permission of Oxford University Press.

In summary, indices of insulin sensitivity were associated with many, but not all, of the metabolites that differed most strongly between LF-discordant cotwins. Surprisingly, indices based on fasting values were most strongly associated with several metabolite exposures, highlighting the predictive value of fasting insulin compared with more intricate OGTT-derived parameters.

5.2 THE GENETIC ARCHITECTURE OF FCH (STUDY II)

The aim of Study II was to characterize the contribution of known rare and common lipid-associated variants to FCH. Genotyping and imputation was performed for 715 FCH family members (including 234 individuals affected by FCH) and 18,715 individuals from the FINRISK population study. We found that both common and uncommon lipid-associated variants are enriched in affected FCH members. Only a small proportion (3%) of affected FCH individuals carry known rare high-impact variants, whereas up to a third have high polygenic risk scores, similar to the polygenic background of hyperlipidemic individuals in the population.

5.2.1 LIPID-ELEVATING VARIANTS ARE ENRICHED IN INDIVIDUALS WITH FCH ACROSS THE ALLELE FREQUENCY SPECTRUM

We began by analyzing the enrichment of lipid-associated variants (including GWAS lead-SNPs and variants catalogued in the OMIM database) in affected FCH family members compared with population controls. Only variants with at least one carrier among the affected FCH individuals were included in the analysis (194 out of a total of 212 catalogued variants). We found an overall enrichment of lipid-elevating variants and depletion of lipid-lowering variants in affected FCH family members (sign test \( p = 0.0025 \) for LDL-C-associated variants and \( p = 0.0016 \) for TG-associated variants). In total, 60/95 LDL-C-elevating variants were more prevalent and 57/99 variants were less prevalent
in affected FCH individuals compared with population controls. Similarly, 61/96 TG-elevating variants were more prevalent and 57/98 TG-lowering variants were less prevalent in affected FCH individuals compared with population controls (Figure 13).

Figure 13. Enrichment of LDL-C or TG associated SNPs in FCH affected individuals by their frequency and effect on LDL-C and TG levels. Enrichment ratio is the ratio of the allele frequencies in the affected individuals (n = 234) to the allele frequencies in the Finnish FINRISK population cohort (n = 18,715). Only individuals without diabetes or other relevant confounders were included. Under the null hypothesis of no enrichment, a 95% credible interval (shaded area) was estimated by calculating the enrichment statistic (enrichment ratio) for all variants with MAF > 0.001% across the genome excluding the loci of the 212 SNPs. Variants are designated as either lipid level elevating (red) or lowering (blue) for (a) LDL-C and (b) TG based on $\beta$ estimates from linear regression in the FINRISK samples. Point size and color intensity reflect the magnitude of the effect. Only SNPs with at least one heterozygous carrier are shown (n = 194). Reproduced with permission from: Ripatti P, and Rämö JT, et al. (2016). The Contribution of GWAS Loci in Familial Dyslipidemias. PLOS Genetics 12(5): e1006078. https://doi.org/10.1371/journal.pgen.1006078

Notably, enrichment of lipid-elevating variation and depletion of lipid-lowering variation was observed throughout the allele frequency spectrum. Most of the enriched variants were originally identified in population studies and include genes whose function is not yet well established, such as $MTHFD2L$ and the $PIGV-NR0B2$ region. Our results are concordant with studies observing enrichment of common variants in familial dyslipidemia.
hypercholesterolemia and highlight the utility of genetic association analyses in the general population for understanding familial disorders.36

5.2.2 HIGH-IMPACT VARIANTS IN A MINORITY OF SUBJECTS WITH FCH OR POPULATION-ASCERTAINED HYPERLIPIDEMIA

Next, we examined the prevalence of putative high-impact lipid-elevating variants in the affected FCH family members. Only two such variants were observed in the sample: APOA rs3135506, which has been proposed to predispose to hypertriglyceridemia in homozygotes; and the APOE ε2ε2 haplotype (a combination of the rs7412 and rs429358 variants), which has been reported to predispose to type III hyperlipoproteinemia (Tables 2 and 3).451, 452

Among the FCH family members, six individuals (five of whom were affected) were homozygous carriers of rs3135506. In addition, we observed 105 heterozygotes (43 of whom were affected) within the FCH families. Overall, the rs3135506 variant was 1.8-fold enriched in the FCH families compared with population controls.

The APOE ε2ε2 haplotype was observed in three FCH family members (two of whom were affected) and was 3.7-fold enriched in affected FCH family members compared with population controls. Consistent with previous reports, the three carriers of the haplotype appeared to have elevated cholesterol and TGs in the VLDL and IDL lipoprotein fractions, and lower LDL-C compared with other APOE isoforms (Table 2).

Additionally, the rare OMIM-catalogued LIPC variant rs28933094 was observed to be 2.6-fold enriched in affected FCH family members (MAF = 0.041) compared with population controls (MAF = 0.016), in addition to being 4.8-fold enriched in FINRISK population controls compared with Non-Finnish Europeans (based on data from the Exome Aggregation Consortium). While rs28933094 predisposes to hepatic lipase deficiency, the lipid profiles associated with rs28933094 have been reported to vary, potentially depending on the underlying lipid phenotype.453-454

As FH might potentially explain part of the FCH phenotype in some families, in addition to the functional LDL receptor phenotyping conducted during family recruitment, we imputed one out of five variants previously reported to underlie FH in Finland (FH-Pogosta), but observed no carriers. As a
limitation, we were unable to successfully genotype other Finnish FH variants, including the large FH-Helsinki deletion. Similarly, the scarcity of other catalogued dyslipidemia-predisposing variants may partly reflect the limits of current genotyping and imputation methods.


Finally, to estimate the contribution of common variation to hyperlipidemia in the FCH families and population controls, we calculated polygenic risk scores for LDL-C and TGs in each subject. Among the 234 affected FCH individuals, 83 (35%) had high polygenic burden (defined as polygenic score over the 90th population percentile) for either LDL-C or TGs (Table 3). A similar proportion (32%) of hyperlipidemic individuals in the population cohort had a high polygenic score for LDL-C or TGs. This suggests a similar contribution of the catalogued variants to FCH and population-ascertained hyperlipidemias of comparable magnitude.

We observed considerable heterogeneity in polygenic scores between and within the FCH families (Figure 14). In a quarter (14 out of 53) of the FCH families, more than half of affected family members had high polygenic scores.
In six out of 53 families (11%), all of the affected individuals either carried putative high-risk variants or had high polygenic scores.

**Table 3.** Number of affected individuals with high polygenic lipid score or Mendelian SNPs. Reprinted with permission from: Ripatti P, and Rämö JT, et al. (2016). The Contribution of GWAS Loci in Familial Dyslipidemias. PLOS Genetics 12(5): e1006078. https://doi.org/10.1371/journal.pgen.1006078

<table>
<thead>
<tr>
<th></th>
<th>High* LDL-C score</th>
<th>High* TG score</th>
<th>Either score high</th>
<th>Both scores high</th>
<th>OMIM SNP†</th>
<th>Either score high or OMIM SNP†</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCH affected (n = 234)</td>
<td>40 (17%)</td>
<td>59 (25%)</td>
<td>83 (35%)</td>
<td>16 (6.8%)</td>
<td>7 (3.0%)</td>
<td>86 (37%)</td>
</tr>
<tr>
<td>FCH probands (n = 48)</td>
<td>2 (4.2%)</td>
<td>13 (27%)</td>
<td>13 (27%)</td>
<td>2 (4.2%)</td>
<td>2 (4.2%)</td>
<td>14 (29%)</td>
</tr>
<tr>
<td>FINRISK hyperlipidemic‡ (n = 2653)</td>
<td>479 (18%)</td>
<td>479 (18%)</td>
<td>839 (32%)</td>
<td>119 (4.5%)</td>
<td>29 (1.1%)</td>
<td>850 (32%)</td>
</tr>
<tr>
<td>FINRISK statin users (n = 6300)</td>
<td>934 (15%)</td>
<td>777 (12%)</td>
<td>1504 (24%)</td>
<td>207 (3.3%)</td>
<td>39 (0.62%)</td>
<td>1516 (24%)</td>
</tr>
</tbody>
</table>

OMIM, Online Mendelian Inheritance in Man; FCH, familial combined hyperlipidemia; FINRISK, The National FINRISK Study.

*Polygenic lipid score > 90th Finnish population percentile, calculated in the Finnish FINRISK population cohort.

†Carriers of SNPs that cause Mendelian dyslipidemias and are observed in the FCH families (APOE ε2ε2 or APOA5 rs3135506 homozygotes).

‡Hyperlipidemia in the FINRISK population samples is defined as TC or TG (non-fasting) > 90th age- and sex-specific population percentile, analogously with the FCH diagnostic criteria.

do:10.1371/journal.pgen.1006078.003

Lastly, we examined the positive predictive value of high polygenic scores in both cohorts. A high polygenic score for either LDL-C or TGs had a positive predictive value for FCH of 0.45 and a negative predictive value of 0.71 in the FCH families. Positive and negative predictive values were 0.23 and 0.89 in the population cohort, respectively, using similar lipid criteria.

As a limitation, the use of dichotomized cutoffs may lead to misleading classification of polygenic burden in individuals close to the 90th PRS percentiles. Some individuals who have lipid PRSs over the 90th population percentile have normal lipid values, reflecting the relative importance of other risk (and protective) factors for dyslipidemia. Conversely, a PRS > 90th percentile does not conclusively establish polygenic burden as the sole or even main contributor to an individual family member’s dyslipidemia. The precise lipid criteria used to define FCH vary between studies; ours represented common criteria at the time of sample collection. It is uncertain how our results would generalize to criteria based on ApoB measurements or more stringent lipid thresholds. The larger PPV and NPV in the FCH families may at least partly reflect the proportion of affected individuals in each cohort. Despite these limitations, the considerable similarity of increased lipid PRSs in dyslipidemic individuals in both the FCH families and in the population suggests a significant role for polygenic burden in FCH.
Figure 14. Number of affected and unaffected individuals with high polygenic lipid scores or carriers of high-impact Mendelian variants. The black shading presents the number of (a) affected, and (b) unaffected individuals with high polygenic lipid scores (LDL-C, TG, or both polygenic scores over the 90th percentile in the population) or carriers of a high-impact Mendelian variant (APOE ε2ε2, n = 2; or homozygosity for APOA5 rs3135506). The families are sorted by the number of affected individuals with high polygenic lipid scores or a high-impact Mendelian variant in (a). The grey shading presents the number of other (a) affected, and (b) unaffected individuals in the family. Reprinted with permission from: Ripatti P, and Rämö JT, et al. (2016). The Contribution of GWAS Loci in Familial Dyslipidemias. PLOS Genetics 12(5): e1006078. https://doi.org/10.1371/journal.pgen.1006078

5.3 CORONARY ARTERY DISEASE RISK AND LIPIDOMIC PROFILES IN FAMILIAL AND POPULATION-ASCERTAINED HYPERLIPIDEMIAS (STUDY III)

The first aim of Study III was to compare the risk of CAD and CVD between familial and population-ascertained hyperlipidemias with high LDL-C or high TGs. Additionally, we aimed to characterize the lipidomic profiles associated with high LDL-C and TG levels, and to compare them between familial and population-ascertained hyperlipidemias. We expected major differences in
pathophysiology underlying familial and population-ascertained hyperlipidemias to be reflected as different lipidomic profiles.

We found that although high LDL-C and TGs were associated with increased CAD and CVD risk, the familiality of the hyperlipidemia conferred no additional risk increase. We observed distinct lipidomic profiles for high LDL-C and TG levels, with TG being uniquely associated with a larger set of lipid species. The lipidomic profiles were highly similar, however, for familial and population-ascertained forms of these hyperlipidemias, suggesting similar and overlapping pathophysiology.

5.3.1 CORONARY ARTERY DISEASE RISK IN FAMILIAL AND POPULATION-ASCERTAINED HYPERLIPIDEMIAS

We began by estimating the risk of incident CAD in familial and population-ascertained hyperlipidemic individuals. We performed separate risk analyses in the families with high aggregation of either LDL-C or TGs (n = 755 individuals), and the FINRISK population cohort (n = 19,644 individuals). The mean (range) follow-up time to first CAD endpoint, death, or end of registry follow-up was 16.1 (0.1–20.1) years in the hyperlipidemic families and 12.6 (0.02–19.0) years in the population cohort.

In the population cohort, individuals with high LDL-C (over 90th population percentile) had an increased risk of incident CAD compared with other individuals (hazard ratio [HR] 1.74; 95% confidence interval [CI] 1.48–2.05) (Figure 15). We observed a similar risk estimate for hyperlipidemia in the 47 “high LDL-C” families (n = 633 individuals) (HR 1.71; 95% CI 0.94–3.10). There was no significant difference in risk between the cohorts (p = 0.84).

The risk of incident CAD was also elevated in individuals with high TGs in the population cohort (HR 1.38; 95% CI 1.09–1.75) and in the “high TG” families (n = 375 individuals) (HR 1.35; 95% CI 0.52–3.51), with no significant difference between the cohorts (p = 0.82).

We analysed the risk of incident CVD similarly, and found no significant differences in the risk associated with hyperlipidemia between the families and population samples. Having high LDL-C conferred hazard ratios for incident CVD of 1.54 (1.33–1.78) and 1.23 (0.69–2.19) in the population cohort and “high LDL-C” families, respectively. Having high TG conferred hazard ratios for
incident CVD of 1.37 (1.12-1.67) and 1.63 (0.81-3.3) in the population cohort and “high TG” families, respectively.

### Figure 15. Risk of incident coronary artery disease (CAD) in hyperlipidemias with family history and population-ascertained hyperlipidemias. To assess the risk of incident CAD associated with the hyperlipidemia types, we used Cox proportional hazards models using age as the time scale, stratified by sex and clustered by family, to estimate hazard ratios (HRs) for incident CAD events, excluding individuals with prevalent CAD. LDL-C indicates low-density lipoprotein cholesterol; TG, triacylglycerides. Reprinted with permission from: Rämö JT et al. Coronary Artery Disease Risk and Lipidomic Profiles Are Similar in Hyperlipidemias With Family History and Population-Ascertainment Hyperlipidemias. Journal of the American Heart Association. 2019;8:e012415. doi:10.1161/JAHA.119.012415.

Finally, we repeated the analyses of incident CAD and CVD risk by additionally adjusting for BMI, self-reported lipid lowering medication usage, and smoking. No between-cohort differences in CAD or CVD risk were found for individuals with high LDL-C (p = 0.92 and 0.59, respectively) or high TGs (p = 0.75 and 0.81, respectively). For this and all previous analyses, meta-analyses of the family and population cohorts closely approximated estimates derived in the population cohort.

Our results contrast with earlier reports of considerably elevated CAD risk in FCH, including up to 5.1-fold increased odds ratios (ORs) for CAD in relatives of Dutch FCHL probands compared with spouses.230-232 However, these analyses have not adjusted for lipid levels, and the definition of the FCH phenotype itself classically includes early coronary heart disease in the

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>Total subjects (hyperlipidemic/ others)</th>
<th>Incident CAD (hyperlipidemic/ others)</th>
<th>HR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDL-C &gt; 90th percentile</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In 'high LDL-C' families</td>
<td>625 (136/489)</td>
<td>45 (16/29)</td>
<td>1.71 (0.94-3.1)</td>
</tr>
<tr>
<td>In the population</td>
<td>19644 (2175/17469)</td>
<td>904 (176/728)</td>
<td>1.74 (1.48-2.05)</td>
</tr>
<tr>
<td>Meta-analysis</td>
<td>20209 (2311/17958)</td>
<td>949 (192/757)</td>
<td>1.74 (1.48-2.04)</td>
</tr>
<tr>
<td>TGs &gt; 90th percentile</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In 'high TG' families</td>
<td>371 (72/299)</td>
<td>21 (3/18)</td>
<td>1.35 (0.52-3.51)</td>
</tr>
<tr>
<td>In the population</td>
<td>19644 (1405/18239)</td>
<td>904 (74/930)</td>
<td>1.36 (1.09-1.75)</td>
</tr>
<tr>
<td>Meta-analysis</td>
<td>20015 (1477/18538)</td>
<td>925 (77/848)</td>
<td>1.36 (1.09-1.74)</td>
</tr>
</tbody>
</table>
proband. The increased OR in the FCH relatives might thus reflect higher lipid levels and also non-lipid-mediated hereditary CAD risk factors. Here, to isolate the effect of familiality, we contrast individuals with familial hyperlipidemia and population samples with similar lipid levels. Further, we differentiate between LDL-C and TGs to isolate their contributions.

Increased CAD risk, even after adjusting for lipid levels, has been reported for individuals with monogenic FH. Here, we excluded FH using functional phenotyping and genotyping. For the remaining individuals with familial high LDL-C, we observed lower additive CAD risk than previously reported for monogenic FH, in line with the smaller LDL-C elevation (+1.9 mmol/l). After exclusion of FH, familial hyperlipidemias appear to be very polygenic. This multiplicity of pathways and pleiotropic influences, in contrast with a single pathway in monogenic FH, may translate to the reduced influence of familiality observed in this study.

Our analyses have several limitations. We were not able to conclusively exclude all Finnish FH variants using genotyping and imputation, although preliminary and unpublished analyses of whole-genome sequencing data in the EUFAM families support their scarcity. We also did not have family history information for FINRISK participants, some of whom may in fact come from families with a history of hyperlipidemia. Reflecting the moderate sample size, the confidence intervals from CAD and CVD risk estimation are large; future studies and meta-analyses are needed to conclusively assess ASCVD risk in familial hyperlipidemias. Nevertheless, excepting analyses of FH, our study represents the largest and most thorough study isolating the effect of familiality in hyperlipidemic individuals to date.

5.3.2 ASSESSING THE BIOLOGICAL VALIDITY OF TWO LIPIDOMICS PLATFORMS

Next, we wanted to characterize the lipidomic profiles associated with familial and population-ascertained hyperlipidemias. To quantify circulating lipid species, preliminary lipidomics phenotyping was first performed for 353 EUFAM family members using two independent platforms as detailed in the Methods section. A total of 179 lipid species were quantified using platform 1 and 325 lipid species were quantified using platform 2.
To assess the biological validity of the platforms, we began by comparing lipidomics measurements of TG species with standard TG measures in the EUFAM cohort. Platform 1 and 2 included 57 and 39 TG species, respectively, and 30 TG species were common to both platform. The sums of individual TG species were compared with standard TG measures separately for each platform. Pearson’s correlation coefficients were 0.68 and 0.82 for platform 1 and 2, respectively (Figure 16).

![Figure 16. Correlation between traditionally measured serum TG and the total TG concentration measured with each lipidomics platform. Total TGs were calculated from the lipidomics data as the sum of the concentration of individual TG species. Platform 1 included 57 TG species, and platform 2 included 39 TG species.](image)

Next, we performed a genome-wide association analysis for the lipid species quantified by Platform 1 (Figure 17) and Platform 2 (Figure 18).

Overall, stronger associations were noted for lipid species quantified by Platform 2 (Figure 18). In addition, likely reflecting biological validity and accuracy of quantification, analysis of lipid species quantified by Platform 2 highlighted known lipid loci including variants in or near FADS2, LIPC, and HAS1-LILRA3 even in the small initial family cohort. In subsequent analyses employing larger cohorts, a total of 35 lipid species associated loci have been identified, 20 of which are in or near known lipid loci, and 15 represent potential novel loci.\textsuperscript{459}
Figure 17. Genome-wide association analysis results for lipid species quantified by Platform 1. P-values (natural logarithm transformed) are presented on the y axis, and individual lipid species grouped within structural classes are presented on the x axis. Lipid values were natural logarithm transformed. Association between genetic variants and lipid species was estimated using linear mixed models with age, sex, and FCH status as covariates and an empiric genetic correlation matrix as the random effect. Modified with permission from Dr. Rubina Tabassum (verbal communication).

Figure 18. Genome-wide association analysis results for lipid species quantified by Platform 2 (annotated with respect to known lipid loci). Modified with permission from Dr. Rubina Tabassum (verbal communication). Note: The y-axis is truncated to highlight SNPs near the significance threshold.
Next, we estimated the heritability of lipid species from each structural class separately for both platforms using linear mixed models (Figure 19). In line with the higher correlation between traditional TG measures and sums of individual lipid species in platform 1 vs. platform 2, the heritability of the sum of TG species was higher for platform 2 vs 1 (mean ± SE: 0.41 ± 0.14 vs 0.18 ± 0.14). Heritability estimates were also higher for PI, PE and LPC species in platform 1 vs 2. However, both platforms yielded high heritabilities for DG and SM species.

**Figure 19.** Heritability estimate (± SE) for the total concentration of lipid species within each lipid class. Lipid values were summed within each class and inverse-normalized separately for men and women. Narrow-sense heritability was estimated using linear mixed models with age as a covariate and an empiric genetic correlation matrix as the random effect. Platform 1 is presented with red bars and platform 2 with blue bars. DG = diacylglyceride, LPC = lysophosphatidylcholine, LPE = lysophosphatidylethanolamine, PC = phosphatidylcholine, PE = phosphatidylethanolamine, PI = phosphatidylinositol, SM = sphingomyelin, TG = triacylglyceride.

Finally, we estimated heritability and between-platform correlation separately for all individual TG species. We observed modest variation between individual TG species, with most heritability estimates ranging between 0.25 and 0.45 (Figure 20). The between-platform correlation differed widely...
between individual TG species, with a range of 0.03–0.9. Between-platform correlations were highest for TG species with the shortest carbon chains. Surprisingly, for some intermediate and long-chain TG species for which both platforms reported high heritability, the between-platform correlation for the measures was very low. For example, TG(52:2) and TG(56:5), which had heritabilities over 0.4 for both platforms, correlated poorly between platforms ($r = 0.03$ and $r = 0.19$, respectively). This suggests that both panels capture biologically relevant variation, but they may be measuring different molecular species.

**Figure 20.** Heritability and between-platform correlation of individual TG species. Panel A) Heritability estimates of each of the 30 TG species that were quantified by both platforms. Panel B) Correlation of the concentration of each TG species between the two platforms.

The heritability of traditional plasma lipids, including TGs, is typically reported in the 40–60% range.\textsuperscript{460-462} The heritabilities of many lipid classes have not been conclusively established, with prior estimates from platform 1 ranging from 0–65% with considerable within-class variation in one study.\textsuperscript{463} In most cases, we were thus unable to contrast the heritability values estimated
in our study with golden standards. Nevertheless, as the method for estimating heritability was the same for both platforms, major differences between platforms are likelier to reflect imperfect quantitation rather than artificially inflated heritability for one platform.

In summary, correlations with standard measures, genome-wide association analyses, and heritability analyses suggested superiority of platform 2 for several lipid classes, but also revealed potential species-specific variation in biological validity. An initial attempt was made to reconcile and further study these differences, but this was not feasible for platform 1.

Concurrently with the preliminary analyses in EUFAM, platform 2 was selected as the lipidomics platform for the FINRISK population cohort. The decision was thus made to extend lipidomics phenotyping using platform 2 to all other EUFAM cohort members with available lipid samples. All subsequent analyses included data produced using platform 2 only.

### 5.3.3 LIPIDOMIC PROFILES OF HIGH LDL-C

Next, we characterized the lipidomic profiles of familial and population-ascertained hyperlipidemias. For both high LDL-C and high TGs, we first characterized the association of hyperlipidemia with individual lipid species. We then additionally estimated the association of hyperlipidemia with saturation of fatty acid chains within each lipid class.

Familial high LDL-C was associated with elevated levels of 99 lipid species including most of the analyzed lipid classes (Figure 21.A.). We also observed reduced levels for six species (three LPC, two LPE and one PCO species). In the population samples, high LDL-C exhibited similar overall lipidomic profiles, with significant associations for 51 elevated lipid species (Figure 21.B.). The effect estimates derived from the two cohorts correlated highly ($r = 0.8$, Figure 23.A.), and we observed no significant between-cohort differences in effect estimates for any lipid species at 5% FDR.

Familial high LDL-C was also associated with altered saturation of fatty acids in several lipid classes, including increased saturation of LPCs and ceramides and reduced saturation of LPEs, PCS, PCOS and PIs ($p$-value range = 0.019–0.0014). Increased saturation of LPCs was also seen in the population cohort ($p=7.2\times10^{-4}$), and the trends for other classes appeared similar to those
observed in the families, with no significant between-cohort differences at 5% FDR.

**Figure 21.** Associations between high low-density lipoprotein cholesterol (LDL-C) status and the levels of 151 lipid species. A, Individuals affected by high LDL-C levels (n=105) were compared with their unaffected relatives (n=358) in the 53 “high LDL-C” families. B, Individuals affected by high LDL-C (n=56) were compared with other individuals (n=841) in the FINRISK study population cohort. The association of high LDL-C status with the lipid species was estimated using linear mixed models with age, age², and sex as the other fixed effect covariates. Statistical significance was evaluated using the Benjamini-Hochberg method at a 5% false discovery rate (FDR). Cer indicates ceramide; DG, diacylglyceride; FDR, false discovery rate; LDL-C, low-density lipoprotein cholesterol; LPA, lysophosphatic acid; LPC, lysophosphatidylcholine; LPE, lysophosphatidyl-ethanolamine; PC, phosphatidylcholine; PCO, phosphatidylcholine-ether; PE, phosphatidylethanolamine; PEO, phosphatidylethanolamine-ether; PI, phosphatidylinositol; CE, cholesteryl ester; SM, sphingomyelin; ST, sterol; TG, triacylglyceride. Reprinted with permission from: Rämö JT et al. Coronary Artery Disease Risk and Lipidomic Profiles Are Similar in Hyperlipidemias With Family History and Population-Ascertained Hyperlipidemias. Journal of the American Heart Association. 2019;8:e012415. doi:10.1161/JAHA.119.012415.

**5.3.4 LIPIDOMIC PROFILES OF HIGH TGS**

We analysed the lipidomic profiles of high TGs identically to high LDL-C, and found widespread differences in lipid species related to high TGs in both the families (elevated levels of 107 lipid species and reduced levels of seven PCO,
two LPC and one PI species; Figure 22.A.) and in the population (elevated levels of 108 species and reduced levels of ten PCO and one LPC species; Figure 22.B.). The effect estimates correlated very highly between the cohorts \( (r = 0.96; \text{Figure 23.B.}) \), with no significant between-cohort differences for any lipid species at 5% FDR.

**Figure 22.** Associations between high triacylglyceride status and the levels of 151 lipid species. A, Individuals affected by high triacylglycerides (n=64) were compared with their unaffected relatives (n=223) in 39 “high TG” families. B, Individuals affected by high triacylglycerides (n=65) were compared with other individuals (n=832) in the FINRISK study population cohort. Abbreviations are displayed in Figure 21 legend. Reprinted with permission from: Rämö JT et al. Coronary Artery Disease Risk and Lipidomic Profiles Are Similar in Hyperlipidemias With Family History and Population Ascertained Hyperlipidemias. Journal of the American Heart Association. 2019;8:e012415. doi:10.1161/JAHA.119.012415.

Compared with the lipidomic profiles of high LDL-C, high TGs appeared more strongly associated with several classes, including TGs, DGs, Pes, PCs, PCOs and PIs. However, high LDL-C appeared more uniquely associated with elevated SM species.

Finally, high TGs were associated with increased saturation of fatty acids in the TG, DG, LPC and CE classes in both the hyperlipidemic families and the general population \( (p\text{-value range} = 0.0012–5.9\times10^{-11}) \). There were no between-cohort differences in effect estimates at 5% FDR. Differences in fatty acid saturation may partly reflect different dietary intake between

93
hypertriglyceridemic and normotriglyceridemic individuals, but they are also affected by endogenous metabolism.\textsuperscript{464}

\textbf{Figure 23.} Correlation of effect estimates for hyperlipidemia status between the hyperlipidemic families and the population samples. The correlation between the effect estimates observed in the family and population cohorts is presented in A) for high LDL-C (effect estimates presented in Figure 21) and B) for high TGs (effect estimates presented in Figure 22). Abbreviations are displayed in Figure 21 legend. Reprinted with permission from: Rämö JT et al. Coronary Artery Disease Risk and Lipidomic Profiles Are Similar in Hyperlipidemias With Family History and Population-Ascertained Hyperlipidemias. Journal of the American Heart Association. 2019;8:e012415. doi:10.1161/JAHA.119.012415.

Finally, despite distinct lipid profiles for LDL-C and TGs, the lipidomic profiles were strikingly similar for familial and population-ascertained forms of these hyperlipidemias. Major differences in metabolic pathways would be expected to translate to differences in the studied lipid classes, especially considering the distinct profiles observed for high LDL-C and TGs. Our results suggest that familial and population-ascertained hyperlipidemias share overlapping and heterogeneous pathophysiology.

\textbf{5.3.5 ADJUSTMENT FOR COMBINED HYPERLIPIDEMIA}

As the lipidomic profiles of high LDL-C and TGs were largely overlapping, we asked to what extent this was due to combined hyperlipidemia in some hyperlipidemic individuals. We observed simultaneously high LDL-C and TGs in 31 individuals in the family cohort and 9 individuals in the population cohort. To further separate the role of LDL-C and TGs, we estimated their independent association with the lipid species in co-adjusted models (Figure 24).
Figure 24. Independent (coadjusted) associations of low-density lipoprotein cholesterol (LDL-C) and triacylglycerides with 151 lipid species. Effect estimates for LDL-C and triacylglycerides were derived from linear mixed models with the lipid species as outcomes and LDL-C, log(triacylglycerides), age, age², and sex as fixed-effect covariates. The effect estimates were derived separately in the hyperlipidemic families (n=550 individuals; A) and the FINRISK study population cohort (n=897 individuals; B). Effect estimates are presented for LDL-C in orange and triacylglycerides in purple. Statistical significance was evaluated using the Benjamini-Hochberg method at a 5% false discovery rate (FDR). Abbreviations are displayed in Figure 21 legend. Reprinted with permission from: Rämö JT et al. Coronary Artery Disease Risk and Lipidomic Profiles Are Similar in Hyperlipidemias With Family History and Population-Ascertained Hyperlipidemias. Journal of the American Heart Association. 2019;8:e012415. doi:10.1161/JAHA.119.012415.

LDL-C was associated with 83 and 91 lipid species in the hyperlipidemic families and the population, respectively, after adjusting for TGs. The associations of many TG, DG, and PC species with LDL-C were reduced in magnitude. Nevertheless, LDL-C remained strongly associated with CE, ceramide, PC and PCO species.

TGs, in contrast, remained strongly associated with 125 and 124 lipid species from various lipid classes in the hyperlipidemic families and the population, respectively, after adjusting for LDL-C. These included all TG, DG, PE, ceramide and PI species. Overall, 42 lipid species were found to be uniquely associated with TGs in either cohort, while only 13 species were uniquely associated with LDL-C (Figure 25).
The overlap in the lipidomic profiles of LDL-C and TGs is expected not only from their intercorrelation, but also from lipoprotein particle metabolism. LDL particles arise from TG-rich lipoproteins in large part through the action of lipoprotein lipase and hepatic lipase, conserving many particle surface phospholipids and core lipids. The lipoprotein particles, however, are further modified by the enzymes such as the phospholipid transfer protein and cholesteryl ester transfer protein.

Earlier studies have reported large variation in the percentual lipid compositions between different lipoprotein classes. Here, we characterized the association of lipid species with more clinically employed lipid parameters. For example, we find that PCs, which have been reported to form 12% of LDL particle lipids and only 3–9% of TG-rich lipoprotein lipids, are more strongly associated with TGs than with LDL-C.

As limitations of our analyses of lipidomic profiles, we excluded many lipid species due to their incomplete capture between the three batches. Future
studies may better capture such species, and these could theoretically highlight species that differ between familial and population-ascertained hyperlipidemias. However, if this were the case, given the considerable intercorrelation of lipids within and between structural classes, we would have expected to have seen some differences even in the current analyses. The fasting time differed between EUFAM and FINRISK samples; this would have been a greater concern had we observed differences between the cohorts. Further, current recommendations support the use of non-fasting samples for the assessment of traditional lipoprotein parameters.467

The definite roles of most circulating lipid species are yet to be established. Several lipid species have been proposed to predict ASCVD risk over traditional risk factors.272-274 Importantly, many of these lipid species appeared to be associated with both LDL-C and TGs in co-adjusted models, including the ceramides Cer(42:1;2) (presumably Cer(d18:1/24:0)) and Cer(42:2;2) (presumably Cer(d18:1/24:1)), the sterol esters CE(16:1) and CE(18:0), and the sphingomyelin SM(34:1;2). All four ceramides in our sample were independently and positively associated with both LDL-C and TGs, and LDL-C was also associated with increased saturation of ceramides.

In contrast with most CE species which were associated with LDL-C, CE(16:1) appeared more strongly linked with TGs than LDL-C. Among SMs, only SM(34:1;2) was negatively correlated with TGs, and only after controlling for LDL-C. Despite positive associations with LDL-C and TGs, some species such as Cer(42:1;2) and TG(56:6) have been reported to be associated with reduced ASCVD risk.272, 274 Such discordances and co-associations may explain why these species have improved ASCVD prediction in recent studies, and highlight the need for better pathophysiological understanding.
6 CONCLUSIONS

6.1 ACQUIRED BMI AND METABOLOMICS – THE ROLE OF LIVER FAT

In Study I, we explored the extent of alterations in circulating metabolites due to acquired differences in BMI between monozygotic twins. In accordance with our hypothesis, we saw that increased BMI was associated with a more atherogenic circulating metabolic phenotype when accompanied by increases in LF.

Importantly, we did not study NAFLD, as has typically been done in studies employing ultrasound or computed tomography. Based on more sensitive magnetic resonance spectroscopy, we were able to quantify even small differences in liver fat in many individuals who would not qualify for a diagnosis of NAFLD. Given that NAFLD predisposes to a spectrum of diseases ranging from simple steatosis to steatohepatosis and fibrosis, analyses of NAFLD cases may be biased by changes seen in more advanced disease of long duration. Here, we were able to more sensitively assess changes due to early differences in hepatocyte lipid content. Using the most comprehensively employed NMR metabolomics platform in the field allowed for an overview of the metabolic changes associated with LF, as well as linking our findings to extensive literature of metabolite-disease associations.

Our results point to the conclusion that acquired obesity with liver fat may represent a more atherogenic phenotype predisposing to ASCVD. The results add to existing literature implicating dysregulated hepatic lipid metabolism as the bridge between obesity, LF and dyslipidemia. Considerable overlap between metabolite profiles associated with LF and measures of insulin sensitivity, as well as the imperfect resolution of differences in many metabolites during the hyperglycemic phase of the OGTT, additionally support a role for dysglycemia and insulin resistance in the pathogenesis of LF.

Interestingly, and importantly, BMI-discordant cotwins without LF discordance also exhibited putatively atherogenic changes, albeit to a lesser extent and excluding many metabolites such as circulating fatty acids. Our results conflict with previous findings from a smaller sample of the same twins, and emphasize caution in interpreting an obese phenotype without LF as “metabolically healthy”.

The most important limitations of our study are its small sample size and the use of a nominal significance threshold to establish differences in metabolite
AUCs related to BMI discordance. Inferences relating to any particular metabolite should thus be drawn with care. Our results are, however, in line with later observations in a larger cross-sectional cohort of young Finns.445

To summarize, accumulating evidence supports LF as a key delineating factor for atherogenic metabolic sequelae in obesity. Although the routine measurement of LF is not feasible in clinical practice, these findings provide motive and avenues for further dissection of the pathophysiology of LF, with potential for eventual targeted prevention of ASCVD in obese individuals.

6.2 FAMILIAL AND POPULATION-ASCERTAINED HYPERLIPIDEMIAS – SIMILAR GENOMIC, LIPIDOMIC AND CLINICAL CHARACTERISTICS

Motivation for Studies II and III arose from uncertainty regarding the etiology and clinical significance of FCH, a relatively common familial hyperlipidemia often present in individuals with premature CAD.

In Study II, we show that, unlike FH, FCH appears to be due to rare variants in only few cases; further, the identified genes (APOE and APOA5) point to distinct metabolic pathways and rather represent genetic reclassification than a monogenic basis for classical FCH. In contrast, up to a third of FCH families exhibited high polygenic burden for either TCs or TGs. High polygenic burden can be inherited separately for either lipid, suggesting that in many FCH families, the combined hyperlipidemic phenotype does not reflect a single cause, but rather a combination of genetic (and likely environmental) causes.

We followed our findings in Study III by analyzing familial traits of high LDL-C and TGs separately. We began with an analysis of detailed circulating lipid phenotypes in what is to our knowledge the most comprehensive lipidomic analysis in common hyperlipidemias to date. Although LDL-C and TGs associated with distinct lipid species, familial and population-ascertained hyperlipidemias were associated with strikingly similar lipid phenotypes, suggesting similarities in pathophysiology. Further, familial and population-associated hyperlipidemias conferred similar risk if incident CAD and CVD.

In addition to casting doubt on the phenotype of FCH as a distinct pathophysiological entity, our results have a reassuring implication. Beyond knowledge of a given family member’s lipid levels, FCH or its constituent familial traits may not be useful as stratifying ASCVD risk factors. This is in stark contrast with monogenic FH, in which CAD risk is elevated even when adjusted for LDL-C levels. As an important limitation, the number of cases with incident CAD during follow-up in Study III was moderate, and we were
not able to contrast treatment responses in individuals with FCH and population-ascertained hyperlipidemias. However, our results do not support the current Finnish recommendations for the management of hyperlipidemias, which emphasize FCH as a stratifying factor.\textsuperscript{229} As part of an evolving evidence base, our findings on the genetics of FCH have been incorporated into the most recent Finnish and European guidelines for the management of dyslipidemias.\textsuperscript{33, 229}
CONCLUDING REMARKS AND FUTURE PROSPECTS

The role of LF in mediating cardiometabolic outcomes associated with obesity remains imperfectly understood. In this thesis, we showed that acquired BMI increases are accompanied by more widespread differences in circulating ASCVD risk markers when LF is also increased. The extent to which these markers represent causative pathways or “bystander” associations is not certain. Mechanistic studies are needed to pinpoint causative targets. Recent genetic data suggest that different ways to lower LF may lead to very different metabolite profiles and ASCVD risk. At the individual level, it is not certain whether measuring LF or novel intermediate cardiometabolic parameters could significantly improve the management of obesity over traditional anthropometric and biochemical markers. Clinical studies with both imaging and metabolomic data are needed to answer these questions.

Many familial hyperlipidemias have been considered phenotypically distinct from common hyperlipidemias in the population. We found that FCH, unlike FH, is a genetically heterogenous and, to a large extent, polygenic disorder. It has been of interest whether genetic tests for FCH should be developed and employed in clinical practice similar to genetic screens for FH. Our data suggest limited utility for such screens. However, as we only studied previously catalogued variation, we were not able to assess potential novel high-impact variants predisposing to FCH. Future studies may uncover such variants in selected patients. Replication of our analyses in other FCH cohorts is also needed to assess the utility of genetic screens in other countries.

Parallelling our genetic analyses, we found that, excepting monogenic FH, familial hyperlipidemias were not associated with higher ASCVD risk compared with population-ascertained hyperlipidemias of similar magnitude. Together, our results do not support the inclusion of FCH as “high ASCVD risk” conditions in Finnish national dyslipidemia guidelines. Given the large confidence intervals in our analyses, replication in other cohorts is needed. Clinical studies are also needed to assess potential differences in treatment responses in familial compared with population-ascertained hyperlipidemias. Finally, our sample was not enriched for certain rare familial hyperlipidemias, such as familial chylomicronemia or type III hyperlipoproteinemia. Future studies could interrogate their ASCVD risk in more detail.

This thesis has focused on potential novel risk stratifying factors in cardiometabolic diseases. While medicine is becoming ever more personalized, it must be noted that measures to address more common cardiometabolic risk factors should not be undervalued. Given the gap
between smoking, diet, and exercise recommendations and behavioural patterns in the population, there remains considerable potential for benefits from even simple public health initiatives.\textsuperscript{168, 469} To have a large impact, targeted screening and therapeutics must also prove cost-effective. PCSK9 inhibitors, for example, still remain considerably less cost effective than statins.\textsuperscript{470} Ultimately, it should not matter most whether an intervention arises from “simple” or precision medicine – but rather, that it enables the greatest possible benefit to human health.
ACKNOWLEDGEMENTS

The work for this thesis was carried out in the laboratories of Professor Aarno Palotie, Professor Samuli Ripatti and Professor Kirsi Pietiläinen at the Institute for Molecular Medicine Finland (FIMM), the Obesity Research Unit at the Research Programs Unit of the University of Helsinki, the Stanley Center at the Broad Institute of MIT and Harvard, and the Analytical and Translational Genetics Unit (ATGU) at Massachusetts General Hospital (MGH). I wish to acknowledge the directors of these research institutions, including Professor Mark Daly, Professor Olli Kallioniemi, Professor Jaakko Kaprio, Professor Olli Silvennoinen, Professor Eric Lander, Professor Steven Hyman, and Professor Henna Tyynismaa.

Financial support for this work has been granted by the MD/PhD Programme and the Doctoral Programme in Population Health of the University of Helsinki, the Finnish Foundation for Cardiovascular Research, the Academy of Finland, the National Institutes of Health, the Finnish Medical Foundation, the Emil Aaltonen Foundation, the Biomedicum Helsinki Foundation, the Paulo Foundation, the Idman Foundation, and the Veritas Foundation. I thank all of the foundations and institutes who have supported our research, as well as the research participants whose contributions have made these studies possible.

I want to express my deepest gratitude to my supervisors Aarno Palotie, Samuli Ripatti and Kirsi Pietiläinen. I have had the unique privilege of learning from three supervisors with unique fields of interest ranging from clinical management to statistical methodology. You have created close-knit communities to which one is time and time again glad to return. A special thanks to Aarno for your graciousness, kindness and wisdom in matters both scientific and personal. I am particularly thankful for the opportunity to spend a research year abroad in Boston during my medical studies. This exchange not only allowed me to focus on research, but also exposed me to a culture of international collaboration and ambition that will forever affect the way I approach science. I thank Samuli for your astuteness, openness and never-ending humour. You have impressed on me the idea that science is never done. I only hope that I am able to incorporate your “get things done” attitude in any future endeavours. Kirsi’s kindness and perseverance continue to guide me, as does your focus on research questions deeply rooted in the patients that you care for. You have my particular thanks for your unyielding support during the lengthy review process of our first manuscript. Together, I could not have hoped for a better group of supervisors.
I want to thank Professor Katriina Aalto-Setälä for taking time from her busy schedule to accept the role of Opponent at my thesis defence. I am similarly thankful to Professor Johanna Kuusisto and Professor Emma Ahlqvist for reviewing and providing valuable comments on this dissertation.

For guidance and support during my doctoral studies, I thank my thesis committee members Professor Tiinamaija Tuomi and Professor Markus Perola. Your bird’s-eye view and input have been essential in the midst of daily analysis plans. I suspect that the speedy format of thesis committee meetings, in which years of research are condensed into a set number of slides, have, in retrospect, been crucial for maintaining focus.

I am grateful to the senior investigators who have co-led our studies. Marja-Riitta Taskinen has given invaluable advice and supervision throughout my doctoral studies. Your candor and expertise on the diagnosis and clinical management of dyslipidemias have ensured that our research has always remained topical. I thank Aila Rissanen for giving freely and generously of your time, advising me on both research strategy and personal aims. You continue to mentor and nurture established and beginning researchers alike with wisdom and grace. Jaakko Kaprio belongs to a rare group of medical researchers equally talented in medicine and statistical methodology. Your advice on analysis strategies and scientific writing have been invaluable. Veikko Salomaa not only is an expert of Finnish health registries, but is also an astute reader of manuscripts, often improving that which we had expressed ambiguously or incompletely. I was privileged to receive remote guidance from Nelson Freimer during regular phone calls early in my studies. You taught me how it can be more valuable to understand precisely a narrow subset of patients, rather than reach less certain conclusions about a larger group. I thank Kai Simons for kindly inviting me to your laboratory in Dresden, providing an invaluable immersion to the world of lipidomics very early in our collaboration. Your curiosity, diligence and desire to improve the whole field of lipidomics continue to inspire me. Nathan Stitziel gave time from his busy clinical and research schedule when we first launched into analyses of lipidomics data based on two very different platforms. Your eye for the relevant helped us choose among many forking paths.

I thank all the other group leaders who have contributed to the research communities I have been privileged to work in during my doctoral studies and rotations, including but not limited to Ben Neale, Daniel MacArthur, Elise Robinson, Elisabeth Widén, Matti Pirinen, Taru Tukiainen, Henna Tyynismaa, Anu Wartiovaara, Pentti Tienari and Pertti Panula.

The analyses would not have been performed nor the manuscripts written if it were not for my coauthors. Above all, I thank Pietari Ripatti, with whom I’ve shared a journey from medical student to Boston expat to physician-
researcher. I was privileged to work together with you from day one at FIMM. The ease with which you crush through to-do lists is incomprehensible. Your wit and humour have ensured that research has never been boring, even in those moments when it has felt laborious. Your skill in writing with brevity continues to inspire me even as I fail to meet your standards, as evidenced again by the length of this dissertation.

A thousand thanks to my other coauthors: Sanni Söderlund for your professionalism, sincerity and good humour; Sakari Jukarainen for your critical outlook infused with a delightful wit and the joy of discovery; Rubina Tabassum for your inexhaustible skills in the most diverse of analyses; Sanna Kaye for teaching me prioritization in matters both scientific and personal; Leonie Bogl for your kindness and thoughtful approach to study design; Niina Matikainen for your clinical expertise and mentorship; Matti Pirinen and Susan Service for your crucial advice on statistics, confounders and biases; Ida Surakka for your genetic expertise and help with all things rare; Aki Havulinna for your deep expertise on health registries and analyses of outcome data; Antti Sarin for your bioinformatics skills and fundamental work on genotyping and imputation; Mathias Gerl and Christian Klose for your meticulous work, passion and guidance in the field of lipidomics; Pirkka-Pekka Laurila, Christian Ehnholm and Päivi Pajukanta for your work in expanding the analytic scope of EUFAM; and Jesper Lundbom, Nina Lundbom and Antti Hakkarainen for your imaging expertise.

I thank Eija Hämäläinen for laboratory and sample management. My thanks also to the coordinators and administrators without whom none of us would have been able to collaborate, including Sari Kivikko, Ulla Tuomainen, Shen Huei-Yi, Nina Blom, Katri Vihma and Laura Viertola.

Research has introduced me to a host of wonderful colleagues. A particular thanks to Sini Kerminen, Lea Urpa, Nina Mars, and Taru Tukiainen for your leadership in organizing events, meetings and get-togethers for doctoral students. I thank Elmo Saarentaus for Boston Eagles, a shared fridge and long sunny walks on the streets of Cambridge; Tuomo Kiiskinen for your longstanding work as a masterful commentator of everyday life and your endpoint wizardry; and Pyry Helkkula for your constant curiosity, good cheer and genuineness. Thank you also to everyone else who has contributed to the communities at FIMM and Biomedicum.

I spent a transformative year in Boston during my studies, which would not have been the same if it were not for a special group of individuals. I luckily met Olli Pietiläinen early on during my introduction to FIMM before he moved to Harvard, and deduced by inference that FIMM must be formed of only the best people. I may never have met as delightful a person as you. You quite literally ensured that I was not homeless during my first days across the
Atlantic. Verneri Anttila convinced me to join FIMM in an early Skype call. Your sheer force of curiosity and ambition in tackling the biggest questions are in a league of their own. Taru Tukiainen has been a wonderful colleague for life and career discussions, as well as for reminding me of the importance of having “Onni” in one’s life. I thank the “Palotie Boys”: Padhraig Gormley, Mitja Kurki, Dennis Lal and Kumar Veerapen. I could not have hoped for a more wonderful crew with which to share a work space and countless free hours. Thank you to Konrad and Alicia for your inspiring science and acumen, and that soft spot in your heart for this chilly city of Helsinki. Finally, thanks to Tim Heinemann for never forgetting, nor letting others forget, what is truly important in life.

Many a morning, day and evening – mostly day and evening – have been spent on the couches of Mocoma, the recliners of Terkko, or someone’s sofa, partly working on research projects and partly chatting. Sakari Jukarainen, Juho Pirhonen and Pekka Vartiainen have been members of this guild of sunlight avoidance, for which I give my heartfelt thanks – as well as for your genuine and lasting desire to pursue intrinsically worthy things in life.

My family is my bedrock. I thank my mother Ilona and my father Lauri for your unwavering love, wisdom and patience. Your guidance has been just that, and never direction. I thank my brother Tuomas for summers at the archipelago, and for always being reliably older than me.

Finally, my doctoral studies quite literally introduced me to Kaitlin Samocha. You, to me, are the most significant of findings.

Joel Rämö
September 28, 2019
in Helsinki
8 REFERENCES


27. Goldstein JL, Schrott HG, Hazzard WR, Bierman EL, Motulsky AG. Hyperlipidemia in coronary heart disease. II. Genetic analysis of lipid


76. Popkin BM, Slining MM. New dynamics in global obesity facing low-and middle-income countries. *Obes. Rev.* 2013;14 Suppl 2:11-20


122. Kersten S. Physiological regulation of lipoprotein lipase. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids.* 2014;1841:919-933


166. Popkin BM, Adair LS, Ng SW. Global nutrition transition and the pandemic of obesity in developing countries. *Nutr. Rev.* 2012;70:3-21


169. Ioannidis JPA. The challenge of reforming nutritional epidemiologic research the challenge of reforming nutritional epidemiologic research the challenge of reforming nutritional epidemiologic research. *JAMA*. 2018;320:969-970


182. Mills KT, Bundy JD, Kelly TN, Reed JE, Kearney PM, Reynolds K, Chen J, He J. Global disparities of hypertension prevalence and control: A systematic analysis of population-based studies from 90 countries. *Circulation.* 2016;134:441-450


206. Welfare NiIHHa. Smoking in finland. 2018;2019


244. Mora S, Glynn RJ, Ridker PM. High-density lipoprotein cholesterol, size, particle number, and residual vascular risk after potent statin therapy. *Circulation.* 2013;128:1189-1197

245. Mora S. Advanced lipoprotein testing and subfractionation are not (yet) ready for routine clinical use. *Circulation.* 2009;119:2396-2404


301. Lihn AS, Richelsen B, Pedersen SB, Bruun JM. Monocyte chemoattractant protein-1 release is higher in visceral than subcutaneous human adipose tissue (at): Implication of macrophages
resident in the at. *The Journal of Clinical Endocrinology & Metabolism*. 2005;90:2282-2289


alcoholic fatty liver disease (NAFLD): A population based study. *J. Hepatol.* 2007;47:711-717


325. Tulloch-Reid MK, Williams DE, Looker HC, Hanson RL, Knowler WC. Do measures of body fat distribution provide information on the risk of type 2 diabetes in addition to measures of general obesity? *Comparison of anthropometric predictors of type 2 diabetes in Pima Indians.* 2003;26:2556-2561


330. Expert Panel on Detection E, Adults ToHBCi. Executive summary of the third report of the national cholesterol education program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (adult treatment panel iii). *JAMA.* 2001;285:2486-2497


344. Lee JJ, Pedley A, Hoffmann U, Massaro JM, Fox CS. Association of changes in abdominal fat quantity and quality with incident


349. Korenblat KM, Fabbrini E, Mohammed BS, Klein S. Liver, muscle, and adipose tissue insulin action is directly related to intrahepatic triglyceride content in obese subjects. *Gastroenterology.* 2008;134:1369-1375


wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* 2018;50:1219-1224


441. R Core Team. R: A language and environment for statistical computing, 2014


