Heritability and family-based GWAS analyses to discover novel lipidomic biomarkers of cardiovascular disease

A thesis submitted to the University of Manchester
for the degree of Doctor of Philosophy
in the Faculty of Biology, Medicine and Health

2020

Kathryn A. McGurk

School of Medical Sciences
Division of Cardiovascular Sciences

Table of contents

Word count: 52, 203

List of tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Appendix

List of figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Appendix

List of equations

Chapter 2

List of abbreviations

Abbreviation	Definition
2-AG	2-arachidonyl glycerol
2H or D	deuterium
2SMR	two-sample Mendelian randomisation
AA	arachidonic acid
ACE	angiotensin converting enzyme
AEA	anandamide/arachidonoyl ethanolamide
AFR	African
ALA	alpha-linolenic acid
AMI	acute myocardial infarction
AMR	American
ASN	Asian
BMI	body mass index
C	carbon
C18DS	dihydrosphingosine
C18S	sphingosine
C18S1P	sphingosine-1 phosphate
C1P	ceramide 1-phosphate
CAD	coronary artery disease
CB1	cannabinoid receptor 1
CD83	inflammatory protein CD83
CER	the sphingolipid species studied in the ceramide array
CER[ADS]	alpha-hydroxy fatty acid dihydroceramide species
CER[AS]	alpha-hydroxy fatty acid ceramide species
CER[NDS]	non-hydroxy fatty acid dihydroceramide species
CER[NS]	non-hydroxy fatty acid ceramide species
CERS	ceramide synthase
CHF	chronic heart failure
CI	confidence interval
CID	collision-induced dissociation
COX	cyclooxygenase
CRM	certified reference material
CVD	cardiovascular disease
CYP450	cytochrome P450
DEGS1	delta 4-desaturase, sphingolipid 1
DGLA	dihomo-gamma-linolenic acid
DHA	docosahexaenoic acid
DHEA	docosahexaenoyl ethanolamide
DHET	dihydroxyeicosatrienoic acid
DiHDPA	dihydroxydocosapentaenoic acid
DiHOME	dihydroxyoctadecenoic acid
DPEA	docosapentaenoyl ethanolamine

eCB	endocannabinoid
EDTA	ethylenediaminetetraacetic acid
Eico	the species studied in the eicosanoid and related mediator array
ELOVL	elongation of very long chain fatty acids protein
EPA	eicosapentaenoic acid
EpOME	epoxyoctadecenoic acid
eQTL	expression quantitative trait locus
ESI	electrospray ionisation
EUR	European
eV	electron volt
FA	fatty acid
FAAH	fatty acid amide hydrolase
FADS	fatty acid desaturase
FBXO28	F-box only protein 28
FLMM	FaST-LMM
FS	full siblings
GCTA	genome-wide complex trait analysis software
GIF	genomic inflation factor
GPR55	G protein receptor 55
GREML	genome-based restricted maximum likelihood
GRM	genetic relationship matrix
GTEx	genotype-tissue expression project
GWAS	genome-wide association study
H2	broad-sense heritability
h2	narrow-sense heritability
HDHA	hydroxydocosahexaenoic acid
HDL	high-density lipoprotein
HEA	heptadecanoyl ethanolamide
HED	high-energy dynode
HETE	hydroxyeicosatetraenoic acid
Hg	mercury
HODE	hydroxyoctadecdienoic acid
HOM	observed homozygosity
HOTrE	hydroxyoctadecatrienoic acid
HR	hazard ratio
HRC	Human Reference Consortium
HS	half siblings
HTO	Hypertension Oxford Cohort
HWE	Hardy-Weinberg Equilibrium
IBD	identical by descent
IH	intracerebral haemorrhage
IS	internal standard
LA	linoleic acid
LC	liquid chromatography

LDL	low-density lipoprotein
LEA	linoleoyl ethanolamide
LLE	liquid-liquid extraction
LMM	linear mixed model
LOX	lipoxygenase
Lp(a)	lipoprotein(a)
m/z	mass-to-charge ratio
MAF	minor allele frequency
min	minutes
MRM	multiple reaction monitoring
MS	mass spectrometry
MS/MS	tandem mass spectrometry
n-3	omega-3
n-6	omega-6
NAE	the N-acyl ethanolamine species studied
NM	not mentioned or total observations
NS	not significant
OEA	oleoyl ethanolamide
OR	odds ratio
OT	other
P	P-value
PDEA	pentadecanoyl ethanolamide
PEA	palmitoyl ethanolamide
pg/ml	picogram per millilitre
PGs	prostaglandins
PheWAS	phenome-wide association study
PLINK	whole-genome association analysis toolset
PO	parent-offspring
POEA	palmitoleoyl ethanolamide
PUFA	polyunsaturated fatty acid
QC	quality control
QQ	Quantile-Quantile
QTDT	quantitative trait-disequilibrium test software
R	R programming software
S/N	signal/noise ratio
SD	standard deviation
SE	standard error
SGPP1	sphingosine-1-phosphate, phosphatase 1
SNP	single nucleotide polymorphism
SPE	solid phase extraction
SPTLC3	serine palmitoyltransferase subunit 3
SRM	selected reaction monitoring
STEA	stearoyl ethanolamide
T2D	type-2 diabetes

TDT	transmission-disequilibrium test
TG	triglyceride
TXB2	thromboxane
UKB	UK Biobank
V	voltage
Ve	environmental variance
VEA	vaccinoyl ethanolamide
Vg	genetic variance
WHR	waist hip ratio

Abstract

Background: Genetic studies of lipids have shown that, while not DNA-encoded, their activities and metabolism are strongly influenced by DNA variants. Signalling lipids are emerging as novel biomarkers of cardiovascular disease (CVD). Those with a strong genetic influence (high heritability) are the strongest candidate species of which common DNA variants influencing lipid concentrations may be identified. Such DNA variants could then be used in Mendelian randomisation analyses to explore causality of each lipid species in CVD and other phenotypes.

Methods: In this study, lipidomics, genome-wide association studies (GWAS), and two-sample Mendelian randomisation (2SMR) using the UK Biobank study, were carried out to identify heritable species of signalling plasma lipids and the common DNA variants that influence their concentration, to allow for assessment of their causality in CVD. A family cohort (196 families, 999 individuals) was collected through a proband with hypertension in each family to assess for common DNA associations with lipids in an exemplar British Caucasian population. Arrays of 83 eicosanoids and related species (Eico), 54 ceramides and related sphingolipids (CER), and 28 *N*-acyl ethanolamines including endocannabinoid anandamide (NAE), were analysed in a range finding study of 204 plasma samples (31 families) by targeted mass spectrometry-based lipidomics (LC-ESI-MS/MS) to identify the most heritable lipid classes for full-cohort analysis (999 samples). Heritability was estimated by pedigree-based QTDT and SNP/GREML-based GCTA software, and family-based GWAS were completed using GCTA software on Human Reference Consortium imputed genotyping data.

Results and Interpretation: The range finding study of 31 families showed that NAE and CER lipid classes were more heritable than the Eico species; 9 NAE, 10 CER, and 4 Eico were significantly heritable. Variants in the gene encoding the rate-limiting step of CER biosynthesis (*SPTLC3*) were identified for 3 CER traits at this stage. The full cohort analysis of 999 plasma samples (196 families) of the NAE and CER classes showed the lipids were significantly heritable over a wide range (h^2 = 18%-87%). A missense variant (rs324420) in the gene encoding the enzyme fatty acid amide hydrolase (*FAAH*), which degrades NAEs, associated at GWAS significance ($P<5x10^{-8}$) with four NAEs (DHEA, PEA, LEA, VEA). Additionally, a previously described GWAS association between a SNP in the gene of the enzyme serine palmitoyltransferase (*SPTLC3*), was extended to a wider range of plasma CER species (7 CER[NS], 2 CER[NDS]). Novel SNP associations (*CD83, SGPP1, DEGS1*) influencing plasma CER lipid species were identified, two of which (*SGPP1* and *DEGS1*) implicate CER species in haematological phenotypes. This genetic analysis of a wide range of plasma NAE and CER species highlights that these bioactive lipids are substantially heritable and are influenced by SNPs in key metabolic enzymes, however, their causality in CVD remains unconfirmed.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright statement

Acknowledgements

A huge thank you goes to my coaches over the last four years, Anna and Bernard. Thank you for giving me the chance to experience a PhD and study a topic that I have remained excited about. I have learnt so much under your supervision, about both research and life as an academic. Thank you both for indulging my research ideas when I asked you your thoughts on them; this has given me the confidence to question what I read and think up new research ideas. In particular, Anna, thank you for your honest hallway conversations over the past four years. Bernard, thank you for all the support, opportunities, and numerous letters of recommendation for all the activities I was involved in during my PhD. I will be extremely lucky in the future to have as great supervisors as you have been.

Thank you to my parents, Shirley and Brian, who have enabled me to study at top-level universities. You taught, funded, and supported me through all levels of education and I would never have reached this level without you. In particular, Mum, thanks for the practical help over the last few years and your steadiness when I first moved country. Dad, thanks for all your help with all of my CVs and applications, and for all of the calls. Thank you to my partner, Ben, for your unwavering support during my PhD, your patience, and all of your very rational advice. You have never once doubted me and always stabilise me. Thank you to my siblings, Sarah, Andrew and Niall, for always reminding me there is more to life than a PhD, and for keeping me cool.

To all the friends and colleagues that I have been so lucky to make in Manchester, their support and fun, those who have shaped my last four years: Caroline, Becca, Ross, Fi, Joe, Katie, Joanna, Megan, Norah, Alya, Anggit, Luke, Yousef, Josh, James, Emma, Steph, Shona, Lisa, the Bedian's Wags, Sara, Matt, Bev, Mark, Shona, Rob, Toby, Alise, Mia, Jade, Lynne, Yingjuan, Leslie, Gennadiy, Thiara, Caitlinn, Alex, Nikhil, Lorna, Jayesha, Halah, Amal, Gio, Naz, Elisa, Harry, Sakis, Amira, Maryam, Alanna. To my long standing friends from Dublin who are always there for me and supported me leaving, in particular; Georgia, Dara, Clodagh, Holly, Connie, Honor, Larissa, Clare, Katie, Ciara, Rob, Naomi, Mitch, Mark, Graeme, Alice.

Thank you to the staff of UoM who gave me a chance: David, Doug, Roz, Jennifer, Adam, Claire. Thank you to Tony and Nophar for believing in the research I did at Stoller. Thank you to the friends from the University of Cape Town, in particular Mark for all his help while I was over there. Thank you for the computing and stats help Simon, Heather, Martin, Hui. Thank you to Laura for the serum analysis and for the lab training. Thank you to Alex for the mass spec training. Thank you to Loli for the quality control advice. Thank you to Neil for the help when the mass spec was playing up and for the mass spec discussions. Thank you to Jonny and Marta for the lipid comparisons. Thank you for all the admin help James, Sarah, Jen, Joy, Caroline. Thank you to the Newton Fund, Manchester Doctoral College, President's Doctoral Scholar Award, FBMH Doctoral Society, and the travel grants I received. Finally, I'd like to thank the Medical Research Council and the University of Manchester for funding my stipend, lab costs, and tuition fees over the last four years.

Publications

The final results of the full cohort analysis of this PhD project have been written up as a manuscript, submitted to the preprint server bioRxiv, while undergoing the publication submission process: McGurk et al. *Heritability and family-based GWAS analyses of the N-acyl ethanolamine and ceramide lipidome reveal genetic influence over circulating lipids.* bioRxiv 815654; doi: https://doi.org/10.1101/815654.

Keavney, McGurk. *The Open Science of Atrial Fibrillation*. Circulation Research 2020;126:210-211.

McGurk et al. *The use of missing values in proteomic data-independent acquisition mass spectrometry to enable disease activity discrimination.* Bioinformatics 2019:btz898.

Nethononda, McGurk, et al. *Marked variation in the heritability estimates of LVM depending on mode of measurement*. Scientific Reports 2019;9:13556.

Conference contributions

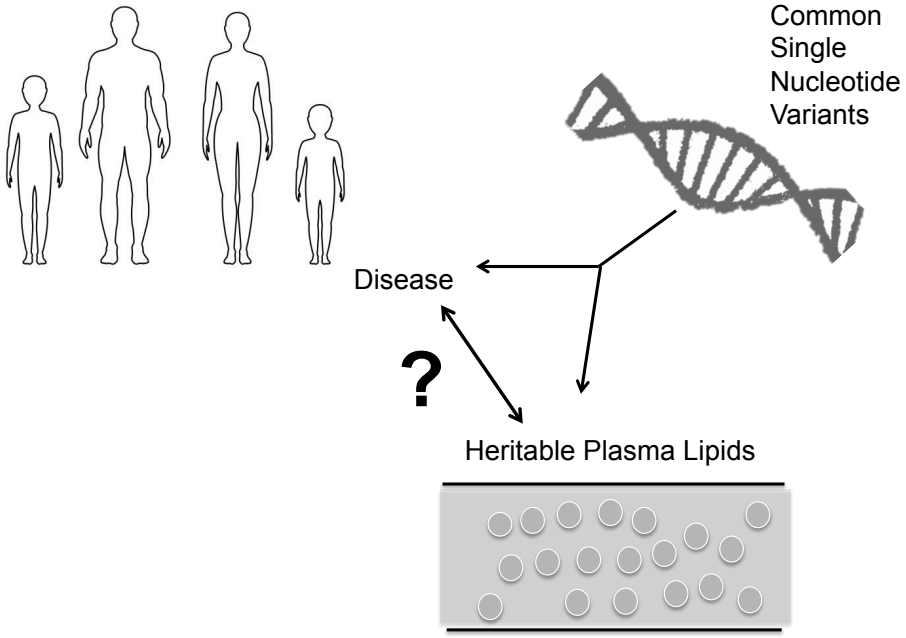
McGurk et al. *104 Heritability and family-based GWAS analyses of the circulating ceramide, endocannabinoid, and N-acyl ethanolamide lipidome*. Heart 2019;**105**:A86. British Atherosclerosis Society/British Society for Cardiovascular Research Spring Meeting, UK, 2019

McGurk et al. *Genetic determinants of bioactive lipid species in a hypertension cohort*. European Heart Journal 2017;38:ehx504.P4249.
European Society of Cardiology Congress, Spain, 2017

McGurk et al. *141 Heritability and family-based GWAS analyses to discover novel lipidomic biomarkers of cardiovascular disease.* Heart 2017;103:A106.
British Atherosclerosis Society/British Society for Cardiovascular Research Spring Meeting, UK, 2017

Chapter 1

Introduction



1.1. Cardiovascular disease

Cardiovascular disease (CVD) is a group of diseases of the heart and vasculature. More than 30% of all deaths worldwide are due to CVD, the leading cause of death and disability in western countries (Lawes *et al.*, 2008). The cost on health systems due to CVD is substantial; it has been estimated that $350 billion was spent on CVD between the year 2014-2015 in the United States (Benjamin *et al.*, 2019). Risk factors of CVD include smoking, obesity, raised blood low-density lipoprotein (LDL; >130 mg/dl), hypertension (>130/80 mmHg (Whelton *et al.*, 2018)), diabetes, chronic kidney disease, and lack of recommended exercise (>150 minutes per week of moderate intensity, or >75 minutes a week of vigorous intensity aerobic physical activity, or equivalent combination (Piercy *et al.*, 2018)) (Benjamin *et al.*, 2019).

Atherosclerotic CVD is the development of atheromatous plaques in the inner lining of arteries, and is affected by diet, smoking, hypertension, dyslipidemia, diabetes and physical activity. It is the main cause of heart attacks and strokes, which kill 1 million patients per year in the US (Topol *et al.*, 2006). Pathology includes cholesterol, inflammation, plaques, calcification, and clot formation in blood vessels (Libby *et al.*, 2011). It is a chronic inflammatory disease of the arterial wall explained by several hypotheses; The Response to Injury Hypothesis – aberrant immune reactions (Ross *et al.*, 1977), The Oxidation Hypothesis – oxidised LDL triggers arterial wall injury and foam cell formation (Steinbrecher *et al.*, 1990), and The Thrombogenic Hypothesis – the role of thrombosis in the formation of plaques (Rokitansky, 1849).

Clinical lipid panels of LDL cholesterol, high-density lipoproteins (HDL), triglyceride (TG), and total cholesterol (LDL, HDL, and TG summed) are heritable and modifiable risk factors for atherosclerotic CVD (Libby *et al.*, 2011). LDL and HDL blood levels are proportional to CVD risk, and plasma LDL is one of the most used and few causal, clinical biomarkers for CVD risk (Wilson *et al.*, 1980). However, about 50% of patients who present with coronary artery disease (CAD), the build up of atherosclerosis in the coronary artery, are classified as low or intermediate risk with current risk algorithms (Hoefer *et al.*, 2015). Thus, there is a need for new biomarkers of atherosclerotic cardiovascular disease for diagnostic purposes and risk

stratification, and the identification of novel drug targets may aid in decreasing the mortality rate associated with CVD.

Genetic analyses have shown that the predicted genetic variance for CVD is largely determined by common SNPs of small effect size (Nikpay *et al.*, 2015; Khera *et al.*, 2018) and most of the identified DNA variants are found near to genes with roles in lipoprotein variation and hypercholesterolemia (O'Donnell *et al.*, 2011). Lipids therefore have a key role in CVD risk (Libby *et al.*, 2011) and recent advances in lipidomic bioanalytics have enabled quantitative analyses of a greater proportion of the lipidome in blood (Quehenberger *et al.*, 2010), with targeted analyses supporting attempts to potentially identify further disease biomarkers, in studies of low-concentration lipid species (Fahy *et al.*, 2009; Stephenson *et al.*, 2017; Kendall *et al.*, 2019).

1.2. Lipids and the lipidome

Lipids are a diverse group of compounds that are insoluble in water but soluble in organic solvents. They are defined as "hydrophobic or amphipathic small molecules that may originate entirely or in part by carbanion-based condensations of thioesters (fatty acyls, glycerolipids, glycerophospholipids, sphingolipids, saccharolipids, and polyketides) and/or by carbocation-based condensations of isoprene units (prenol lipids and sterol lipids)" (Fahy *et al.*, 2009). The lipidome is defined as the entire collection of lipid species found in any given system (eukaryotes and prokaryotes) (Fahy *et al.*, 2009). Lipids have more distinct molecular species than the other biomolecules comprising the human body (e.g. nucleic acids, amino acids, and carbohydrates) (Quehenberger *et al.*, 2010). They are naturally occurring molecules with central roles in membrane structure, energy production, cell signalling, gene expression regulation, and the immune response. Eight lipid categories are defined by the LIPID MAPS consortium (Fahy *et al.*, 2005, 2009): fatty acyls, sphingolipids, glycerolipids, glycerophospholipids, sterol lipids, prenol lipids, saccharolipids, and polyketides. Lipoproteins, with causal roles in CVD, are not lipids but protein aggregates carrying many lipid classes (Rolim *et al.*, 2015). Certain "bioactive" lipids are now measurable; unique species that are potent, complex biomolecules, that can

exert diverse signalling functions, and range in concentration and structure (Quehenberger *et al.*, 2010).

This project studies three groups of bioactive lipid mediators found in plasma: eicosanoids and related species, *N*-acyl ethanolamines including endocannabinoid anandamide, and ceramides and related sphingolipid mediators. For clarity, all species analysed in the assay for eicosanoids and related species are abbreviated as "Eico", the *N*-acyl ethanolamines are abbreviated as "NAE", and the ceramides and related sphingolipids are abbreviated as "CER". The number of species studied here in each of the three classes of lipids cannot be identified by current untargeted, shotgun lipidomics due to their low concentration in blood, so they are less well studied and require targeted lipidomics techniques for measurement.

1.2.1. Endocannabinoid anandamide and *N*-acyl ethanolamines

N-acyl ethanolamines (NAEs) are fatty acid derivatives, derived from membrane phospholipid precursors, and degraded by the enzyme fatty acid amide hydrolase (*FAAH*; Figure 1-1). This class of bioactive lipids includes the endocannabinoid (eCB) anandamide (AEA), which has strong affinity for the cannabinoid receptors CB1 and CB2 (Devane *et al.*, 1992). The rest of the NAE species are structural analogues of AEA, termed "eCB congeners", and have varying affinities for the cannabinoid receptors, but have been suggested to enhance the binding of AEA to the receptors, termed "the entourage effect" (Ho *et al.*, 2008). Many NAEs signal through other receptors of which AEA too binds (e.g. G protein receptor 55 (*GPR55*)) in neurotransmission) (Godlewski *et al.*, 2009), highlighting the overlap of NAEs with eCB signalling. Such NAEs include the nuclear factor agonist palmitoyl ethanolamide (PEA), the anorexic mediator oleoyl ethanolamide (OEA), and a number of other species with roles in neuronal signalling, pain, and obesity (Devane *et al.*, 1992; Calignano *et al.*, 2001; Rodríguez de Fonseca *et al.*, 2001; Wilson *et al.*, 2001; Engeli *et al.*, 2005; Hohmann *et al.*, 2005). Direct cannabinoid receptor 1 (CB1) antagonist drugs, for anorectic, anti-obesity treatments, have caused severe adverse psychiatric effects (Mach *et al.*, 2009). FAAH inhibitors are being evaluated as an alternative approach to modulating eCB signalling for treatment of multiple diseases (Mallet *et al.*, 2016).

The contribution of genetic factors to the variation in circulating NAEs has not been studied at a genome-wide level. At the time of commencement of the project (October 2015), no GWAS had been completed for this class of lipids. Since then, the NAE species oleoyl ethanolamide (OEA) was identified in a GWAS study using untargeted, shotgun mass spectrometry, which identified a variant in the *FAAH* gene to associate with this lipid (Long *et al.*, 2017).

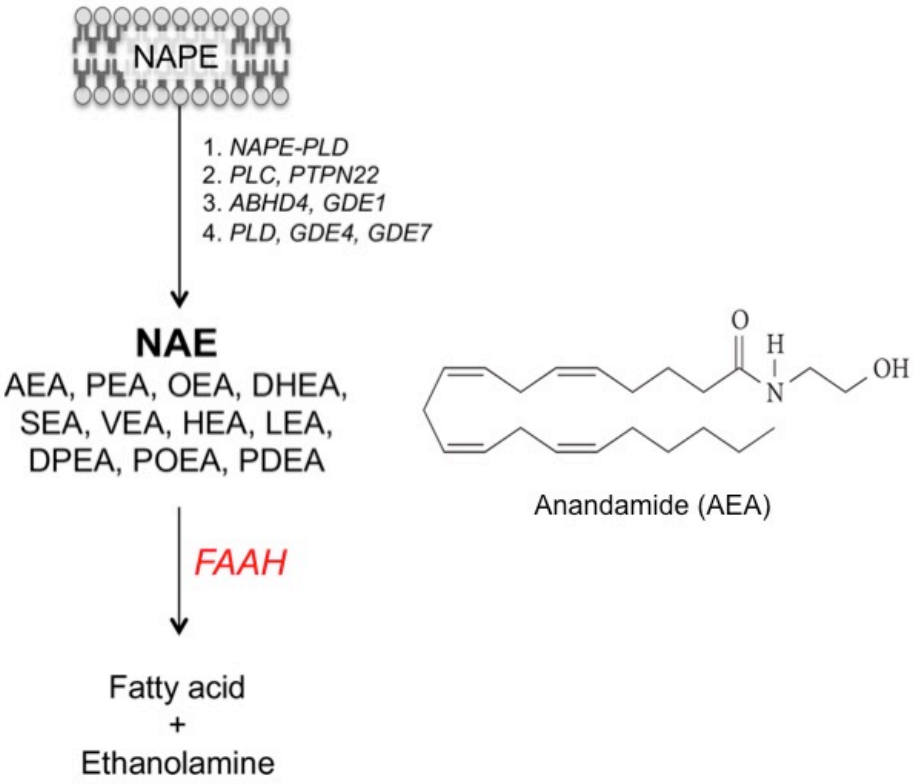


Figure 1-1: Schematic overview of the biosynthetic pathway for *N*-acyl ethanolamines

N-acyl ethanolamines (NAE) are produced through four independent enzymatic pathways from membrane phospholipid precursors (*N*-acyl phosphatidylethanolamine; NAPE), and differ by fatty acid substrate. Fatty acid amide hydrolase (*FAAH*) degrades the relevant NAEs to free fatty acids (such as arachidonic acid in the case of anandamide (AEA)) and ethanolamine. Genes encoding enzymes are in italics. *NAPE-PLD, N*-acyl phosphatidylethanolamine phospholipase D; *PLC*, heparin sulfate proteoglycan 2; *PTPN22,* protein tyrosine phosphatase non-receptor type 22; *ABHD4*, abhydrolase domain containing 4; *GDE1,* glycerophosphodiester phosphodiesterase 1; *GDE4*, glycerophosphodiester phosphodiesterase 4; *GDE7,* glycerophosphodiester phosphodiesterase 7; *PLD*, glycosylphosphatidylinositol specific phospholipase D1; *FAAH*, fatty acid amide hydrolase.

1.2.2. Ceramides and other sphingolipids

Ceramides are derivatives of sphingoid bases (e.g. sphingosine and dihydrosphingosine) and fatty acids (Figure 1-2). The first step of their *de novo* biosynthesis is catalysed by the enzyme serine palmitoyltransferase enzyme, a heterodimeric protein whose monomers are encoded by the *SPTLC1-3* genes, the rate limiting step of sphingolipid biosynthesis (Perry *et al.*, 2000). The CER species measured in this project include: non-hydroxy fatty acid-based species (CER[NS]); the CER[NS] precursors, dihydroceramides (CER[NDS]), which contain a dihydrosphingosine backbone; CER[AS] species with an alpha-hydroxy fatty acid; and the cell survival (Cuvillier *et al.*, 1996) mediators, sphingosine (C18S) and sphingosine-1 phosphate (C18S1P) (Figure 1-2).

The CER[NS] have been well studied for their important roles in apoptosis (Perry *et al.*, 1998). They can initiate apoptosis directly at the mitochondrial membrane or can stimulate multiple apoptosis-inducing signalling pathways. For example, CER[N(16)S(18)] has been shown to initiate apoptosis by creating a mitochondrial membrane channel or pore, increasing the permeabilisation of the membrane (Siskind *et al.*, 2000, 2002). Signalling at the cell membrane by stress response molecules, such as TNFα, causes increases in cellular CER levels that in turn activate multiple signalling cascades to induce apoptosis (reviewed in Pettus *et al.* (2002)).

The nomenclature of the CER has evolved as the field expands with the current used notation in "NS" format. The CER[NS] notation denotes a CER that contains a non-hydroxy fatty acid attached to a sphingosine base, for example, a 16-carbon non-hydroxy fatty acid joined to a 18-carbon sphingoid base is denoted as CER[N(16)S(18)], where N(16) represents a 16-carbon non-hydroxy fatty acid, and S(18) represents a 18-carbon sphingosine base attached (Figure 1-3).

A small subset of CER species found at high concentration can be measured by shotgun mass spectrometry, and this has allowed for the study of CER species as biomarkers of disease and for inclusion in lipidomic genetic analyses. Recently, some circulating CER[NS] derivatives of 18-carbon sphingosine (e.g. CER[N(16)S(18)]) have been identified as novel biomarkers of cardiovascular fatal outcome (Laaksonen *et al.*, 2016), type-2 diabetes, and insulin resistance (Haus *et al.*, 2009). However, the

association between CER species and CVD has not been confirmed in all studies, reviewed in Table 1.1.

Genetic analyses have been completed for seven such CER species. Heritability has been estimated in two studies. CER[N(16)S(18)], CER[N(20)S(18)], CER[N(22)S(18)], CER[N(23)S(18)], CER[N(24)S(18)], CER[N(24:1)S(18)], and corresponding dihydroceramide species (CER[NDS]), had estimated heritability of 37%-51% ($P<0.01$) for CER[NS] and 9%-34% ($P<0.01$) for CER[NDS] in 42 Mexican American families (Bellis *et al.*, 2014). A recent study of CER[N(22)S(18)], CER[N(22:1)S(18)], CER[N(24)S(18)], and CER[N(24:1)S(18)], estimated heritability of 35%-40% for the four species (Tabassum *et al.*, 2019). The same seven CER[NS] species have been assessed by GWAS (Hicks *et al.*, 2009; Demirkan *et al.*, 2012; Tabassum *et al.*, 2019) and the association results are depicted in Table 1.2.

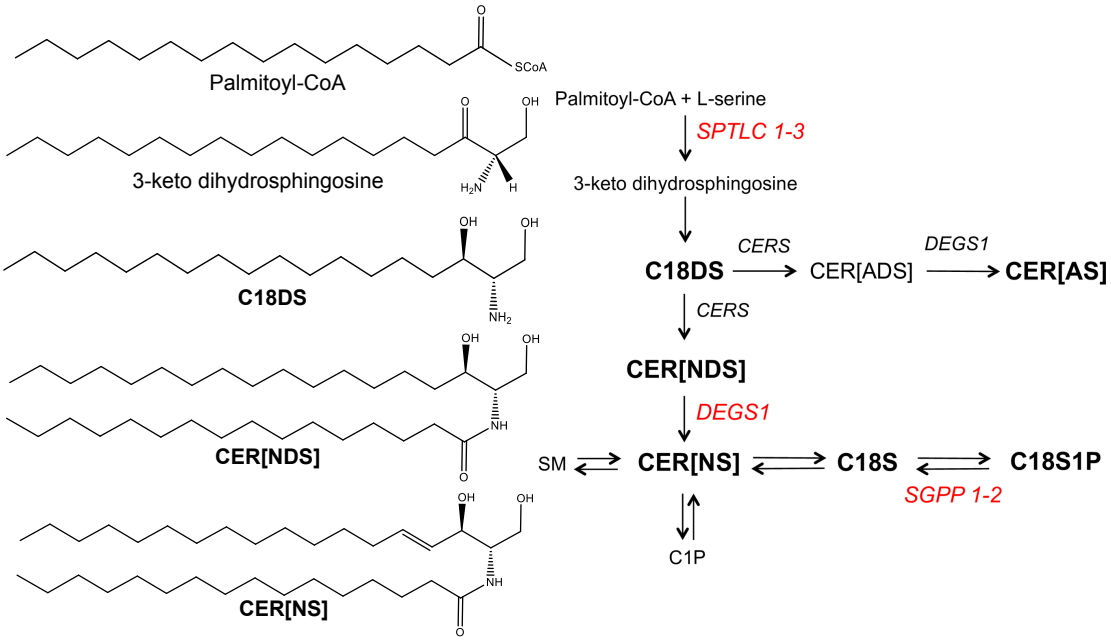


Figure 1-2: Structures and schematic overview of the biosynthetic pathway for ceramides and related sphingolipid species

In the *de novo* pathway, Ceramide (CER) species are biosynthesised via the enzyme serine palmitoyltransferase (*SPTLC1-3*) that converts palmitoyl-CoA and L-serine to 3-keto dihydrosphingosine in the rate-limiting step of the sphingolipid de novo pathway. The resulting dihydrosphingosine (C18DS) is coupled to various fatty acids via ceramide synthases (*CERS*) to generate dihydroceramides CER[NDS] that are further converted to CER[NS] via the enzyme delta 4-desaturase, sphingolipid 1 (*DEGS1*). Conversion of CER[NS] species to sphingosine (C18S) and sphingosine 1-phosphate (C18S1P), is through reversible reactions via the action of sphingosine-1 phosphate, phosphatase (*SGPP1-2*). CER[NS] are also reversibly converted to sphingomyelin (SM) and ceramide 1-phosphate [C1P]). In a similar way to the *de novo* production of CER[NS], addition of alpha-hydroxy fatty acids to C18DS results in CER[ADS] species. The structures of the precursor species in the creation of apoptotic CER[NS] are depicted to the left of the image. Measured lipid species are depicted in bold.

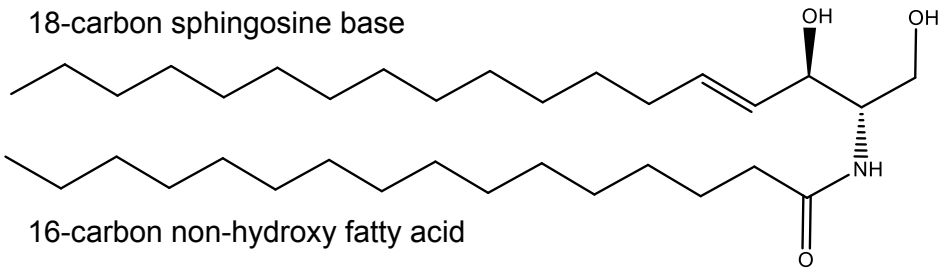


Figure 1-3: Structure of ceramide species CER[N(16)S(18)]

The ceramide CER[N(16)S(18)] is structured with a 18-carbon sphingoid base (sphingosine) attached by amide bond to a 16-carbon non-hydroxy fatty acid. Ceramide species vary in carbon length on both the sphingoid base and fatty acid chains.

Table 1.1: Summary of studies associating CER and calculated traits with fatal outcome by cardiovascular disease

Results of hazard (HR) or odds ratio are shown together for the assessment of circulatory CER, total sum, and further ratios of the species studied. Presented significant results includes a confidence interval (CI) of greater than 1. Positive: HR > 1.00; Negative: HR < 1.00; MACE, major adverse cardiac events; CAD, coronary artery disease; AMI, acute myocardial infarction; CHF, chronic heart failure; NS, not significant; NM, not mentioned even though the species was measured (likely the result was NS); a, summation of the specific ceramides studied in a particular study, which varies; b, depending on adjustment used.

Reference	(Tarasov *et al.*, 2014)	(Laaksonen *et al.*, 2016)	(Laaksonen *et al.*, 2016)	(Havulinna *et al.*, 2016)	(Havulinna *et al.*, 2016)	(Yu *et al.*, 2015)	(de Carvalho *et al.*, 2018)
Mortalities (n)	258	81	51	Fatal Incident MACE 116	Fatal Recurrent MACE 70	200	26
Controls (n)	Stable CAD 187	Stable CAD 1499	Stable CAD 1586	Incident MACE 7589	Recurrent MACE 326	CHF 223	AMI 288
N(14)S(18)							**NS**
N(16)S(18)		**Positive**	**Positive**	**Positive**[b]	**Positive**		**Positive**
N(18)S(18)		**Positive**[b]	**Positive**	**Positive**[b]	**Positive**[b]	**NM**	**NS**
N(18:1)S(18)							**P = 0.047**
N(20)S(18)						**NM**	**NS**
N(20:4)S(18)							**NS**
N(22)S(18)						**NM**	**NS**
N(22:1)S(18)							**NS**
N(23)S(18)							**NS**
N(24)S(18)		**NS**	**Negative**[b]	**NS**	**NS**	**NM**	**NS**
N(24:1)S(18)		**Positive**	**Positive**[b]	**NS**	**Positive**		**Positive**
N(25)S(18)							**Positive**
N(26)S(18)							**NS**
N(26:1)S(18)							**Positive**
N(16)S(18)/ N(24)S(18)	**Positive**	**Positive**	**Positive**	**NS**	**NS**		
N(18)S(18)/ N(24)S(18)	**Positive**	**Positive**	**Positive**	**Positive**	**NS**		
N(20)S(18)/ N(24)S(18)							
N(24)S(18)/ N(24:1)S(18)	**Negative**						
N(24:1)S(18)/ N(24)S(18)		**Positive**	**Positive**	**NS**	**NS**		
Total CER[a]						**Positive**	

Table 1.2: Summary of results from previous GWAS of CER species

The table depicts the nearest gene of variants identified from GWAS that associated with CER species to date. *SPTLC3,* serine palmitoyl transferase subunit 3; *CERS4*, ceramide synthase 4; *ZNF385D*, Zinc Finger Protein 385D.

CER	Gene	Ref
CER[N(16)S(18)]	*SPTLC3*	Hicks *et al.*, 2009
CER[N(20)S(18)]	*CERS4*	Hicks *et al.*, 2009
		Demirkan *et al.*, 2012
		Tabassum *et al.*, 2019
CER[N(22)S(18)]	*SPTLC3*	Hicks *et al.*, 2009
		Demirkan *et al.*, 2012
		Tabassum *et al.*, 2019
CER[N(22)S(18)]	ZNF385D	Tabassum et al., 2019
CER[N(22:1)S(18)]	*SPTLC3*	Tabassum *et al.*, 2019
CER[N(23)S(18)]	*SPTLC3*	Hicks *et al.*, 2009
		Demirkan *et al.*, 2012
CER[N(24)S(18)]	*SPTLC3*	Hicks *et al.*, 2009
		Demirkan *et al.*, 2012
CER[N(24)S(18)]	*ZNF385D*	Tabassum *et al.*, 2019
CER[N(24:1)S(18)]	*SPTLC3*	Hicks *et al.*, 2009
		Demirkan *et al.*, 2012
		Tabassum *et al.*, 2019

1.2.3. Eicosanoids, octadecanoids, and docosanoids

Oxygenation of polyunsaturated fatty acids (PUFAs) creates many lipid mediators, including; (i) eicosanoids, derivatives of the 20-carbon (C) PUFA arachidonic acid (AA) (ii) docosanoids, derivatives of the 22-C PUFA docosahexaenoic acid (DHA), and (iii) octadecanoids, derivatives of the 18-C PUFAs linoleic acid (LA) and α-linolenic acid (ALA). AA and LA are omega-6 (n-6) PUFAs, and ALA and DHA are omega-3 (n-3), where the last double bond is located at the 6^{th} and 3^{rd} C from the methyl terminus, respectively. PUFAs have been implicated in altering inflammation, with n-3 PUFAs having perceived anti-inflammatory roles from multiple epidemiology studies in the 1980s identifying an inverse relationship between dietary

intake of n-3 PUFAs with cardiovascular morbidity in the Greenland Inuit population (Dyerberg, 1989). CVD risk guidelines continue to recommend the replacement of less healthy food for seafood 1-2 times a week to reduce CVD risk (Rimm *et al.*, 2018), however, current studies of PUFAs in inflammation show that the relationship is unclear (Innes *et al.*, 2018), as is the evidence of dietary intake of n-3 PUFAs providing CVD benefit (Bowen *et al.*, 2016). The bioactive, signalling, mediator lipids produced from PUFAs have known roles in inflammation and immunity (Wall *et al.*, 2010).

PUFAs esterified in cellular membranes are released by cystolic phospholipase A2 and metabolised by several enzymes to produce such mediators; (i) the metabolism of AA by cyclooxygenase enzymes (COX; *PTGS* genes) produces eicosanoids and prostanoids: prostaglandins (PGs), prostacyclin (PGI_2), and platelet aggregating thromboxane (TXA_2); (ii) lipoxygenase enzymes (LOX; *ALOX* genes) produce hydroxyeicosatetraenoic acids (HETE), hydroxyoctadecdienoic acids (HODE), hydroxyoctadecatrienoic acids (HOTrE), and hydroxydocosahexaenoic acids (HDHA); and (iii) the CYP450 monoxygenases (*CYP1A, -2B, -2C, -2D, -2G, -2J, -2N, -4A*; including epoxygenases, mid-chain and terminal monooxygenase) produce dihydroxydocosapentaenoic acids (DiHDPA), epoxyoctadecenoic acids (EpOME), dihydroxyoctadecenoic acids (DiHOME), and dihydroxyeicosatrienoic acids (DHET) (Figure 1-4) (Hamberg *et al.*, 1975; Quehenberger *et al.*, 2010; Massey *et al.*, 2013; Astarita *et al.*, 2015).

Numerous small-scale studies of these unique lipids species have implicated them in CVD inflammation as they are found at increased concentration have in the blood of patients compared to healthy controls (Lötzer *et al.*, 2005; Theken *et al.*, 2012; Schuck *et al.*, 2013; Yang *et al.*, 2013; Oni-Orisan *et al.*, 2016). The roles of each species are not all fully understood, however in CVD, the hydrolysis of EET species to DHET species has been implicated in the regulation of vasoconstriction caused by EET species (Yu *et al.*, 2000), HODE species have been identified as major components of oxidised LDL (Nagy *et al.*, 1998), and HETE species are thought to remodel the vasculature during hypertension (Gainer *et al.*, 2005; Ding *et al.*, 2013).

Neither heritability nor GWAS analyses have been undertaken for the eicosanoid, octadecanoid, and docosanoid species measured in this project. However their precursor PUFA species have been estimated to be substantially heritable in Caucasian populations; 12%-59% of the variance in their blood levels is due to genetic factors (Bray *et al.*, 2007; Shah *et al.*, 2009; Tanaka *et al.*, 2009). Other plasma bioactive mediator species (leukotriene B4, thromboxane B2, prostaglandin E2, and thromboxane A2) have also shown substantial heritability; 28%-75% in Caucasian populations (Bray *et al.*, 2007; Vila *et al.*, 2010; Camacho *et al.*, 2012). Genomic loci have also been found with variants that significantly associate at GWAS with PUFAs, for example, genes of the enzyme fatty acid desaturase (*FADS*), which unsaturates fatty acids to PUFAs and the gene encoding the PUFA elongase enzyme (*ELOVL2*) (Tanaka *et al.*, 2009; Lemaitre *et al.*, 2011).

The identification of heritable bioactive lipid species, and the common DNA variants influencing their circulating concentrations, would allow for assessment of a causal role for these signalling lipid species in CVD risk.

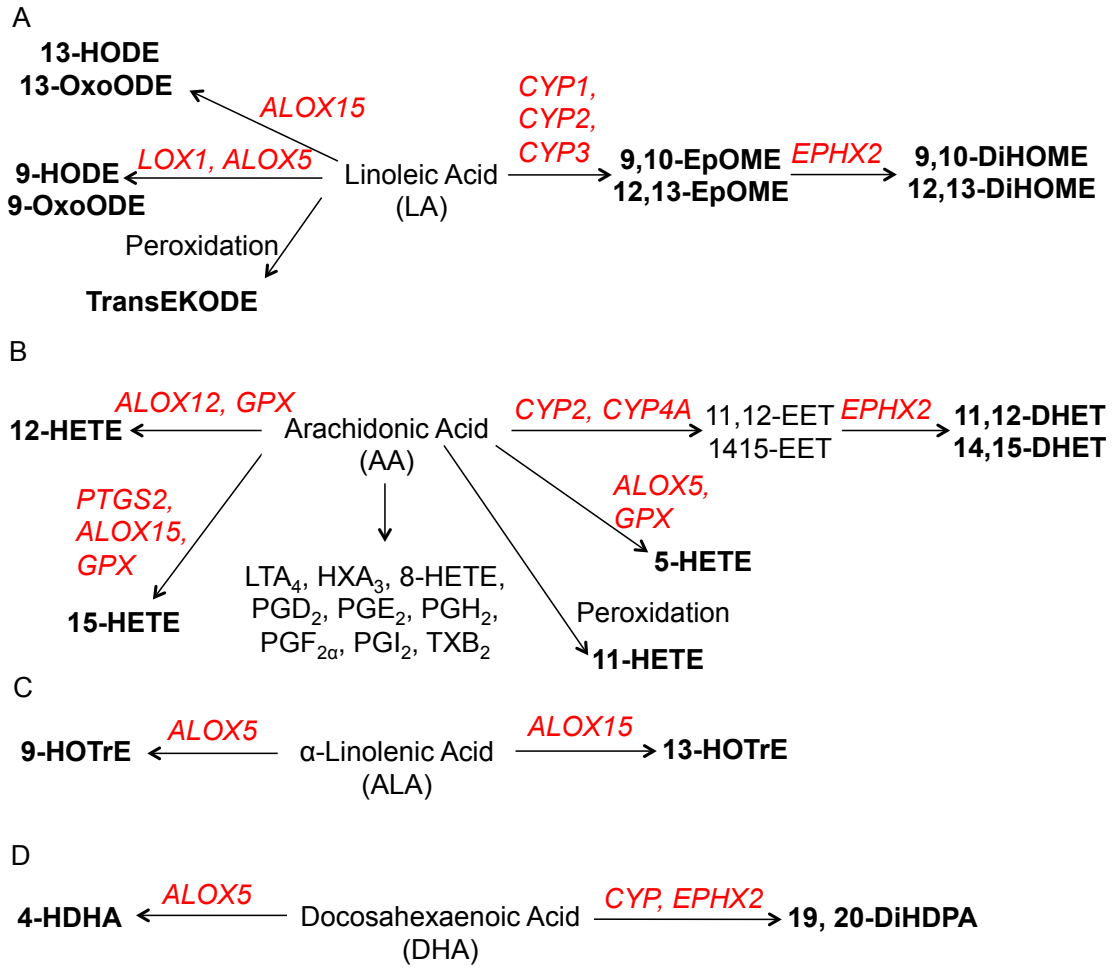


Figure 1-4: Schematic biochemical pathways of eicosanoids, octadecanoid, and docosanoid species

A) Linoleic acid (LA) metabolism via lipoxygenases (*LOX1*, *ALOX-5, -15*), CYP450 enzymes (*CYP*), and epoxide hydrolase (*EPHX2*). B) Arachidonic acid (AA) metabolism via cyclooxygenases (*PTGS1-2*), lipoxygenases (*ALOX5, -12, -15*), CYP450 enzymes (*CYP*), epoxide hydrolase (*EPHX2*), and glutathione peroxidase (*GPX*). C) Alpha-linoleic acid (ALA) metabolism via lipoxygenases (*ALOX5, -15*). D) Docosahexaenoic acid (DHA) metabolism via lipoxygenase (*ALOX5*), CYP450 enzymes (*CYP*), epoxide hydrolase (*EPHX2*). Measured lipid species are in bold.

1.3. Mass spectrometry in bioanalysis and lipidomics

Lipidomics is the comprehensive analysis and characterization of lipid molecules (Fahy *et al.*, 2005). The biochemistry of lipids results in the production of many closely related molecules and thus, analysis of the unique species requires high sensitivity (Murphy *et al.*, 2011a). For example, the sphingolipid class of CER studied here contain the most numerous lipids of any lipid class identified in plasma samples, and can differ from each other by the addition of a carbon to their structure (Quehenberger *et al.*, 2010). Furthermore, potent, signalling lipids, including those of the CER, NAE, and Eico groups, are found at low abundance in blood samples and require an analysis of high sensitivity.

Mass spectrometry analyses allow for high sensitivity and high mass accuracy of lipids of interest in a sample (Murphy *et al.*, 2011a). Shotgun mass spectrometry, an untargeted, unbiased technique, allows for the analysis of a broad coverage of many lipid classes identifiable in a biological sample. However, this approach is affected by ion suppression by high concentration lipids, limiting its ability to detect minor species. Thus, it mostly identifies the most abundant lipids in a sample, for example, phospholipids, cholesterol, and triacylglycerol in blood samples (Quehenberger *et al.*, 2011).

In comparison, targeted mass spectrometry undertaken in this study, uses chromatographic separation to isolate selected lipid classes of interest prior to mass spectrometry analysis. This allows for the analysis of many lipid species from a selected lipid class, with information about their mass spectrometry fragmentation and hydrophobicity known *a priori* (Murphy *et al.*, 2011b). This requirement of a targeted protocol has inhibited the analyses of low concentration plasma lipid mediators in substantial sample sizes (Hinterwirth *et al.*, 2014), but allows for the assessment of an extended array of bioactive lipid mediators of the Eico, NAE, and CER classes in this study, where most species have not been assessed via genetic analyses which require the analysis of hundreds of samples.

Lipid extraction from a biological sample is undertaken using organic solvents via liquid-liquid extraction (such as for the analysis of CER and NAE lipids) or methods targeting hydrophobicity via solid phase extraction (SPE) (as for the analysis of Eico

species). Liquid chromatography (LC) with tandem mass spectrometry (MS/MS) is the analytical platform of choice for low concentration lipids as LC allows for chromatography under particularly high pressure resulting in reduced retention times compared to other chromatography methods, no derivatisation step is required, and the MS has rapid scanning, allowing for a rapid analysis of tens of lipids at high sensitivity and specificity (Murphy *et al.*, 2011b). Lipid species are not typically separated by normal phase analysis (separation by polarity), therefore reverse-phase LC is required to separate lipids according to their known hydrophobic properties, deemed "lipophilicity".

The triple quadrupole mass spectrometer was found to be favourable for the analysis of lipids due to its collision-induced dissociation (CID) of molecular ions and the modes of tandem mass spectrometry (MS/MS) operation, including product ion scanning, precursor ion scanning, and multiple reaction monitoring (MRM), compared to ion trap and time-of-flight type mass spectrometers. Coupled to this, the ionization methods of electrospray ionization (ESI) is capable of readily generating molecular ion species from all known lipids. ESI is a soft, low energy desorption ionisation technique that allows for minimal insource fragmentation. ESI is a generally applicable ionization method that yields rich fragmentation patterns of lipids (Murphy *et al.*, 2011a).

The mass spectrometry pipeline used in this study is as follows (depicted in Figure 1-5); the lipid extract from a tissue of interest is injected via automated robot to the LC for separation of the lipids by reversed phase chromatography. The lipids mix with aqueous and organic mobile phases and reach the column; a silica-bonded C8 or C18 stationary phase. The method uses a specific flow of altering gradients to separate out the lipid species of interest depending on their affinity with the column, the aqueous mobile phase, or organic solvent, passing them to the mass spectrometer.

The mass spectrometer can only measure the mass of a molecule by its conversion to gas-phase ions. Through ESI, the separated lipids in liquid form receive a high voltage from an electric field in a high-pressured capillary needle, charging and oxidising the droplets to form a mist. The mist moves into the desolvation capillary by an applied voltage where dehydrated, heated nitrogen drying gas causes the charged

droplets in the mist to evaporate (desolvation). They decrease in size and split based on charge (coulombic explosion) and exit through the Taylor cone into the mass analyser.

The mass analysers are in a vacuum environment to prevent the ions from interacting with atmospheric compounds. The triple quadrupole mass spectrometer has three quadrupoles; two mass analysers and a collision cell. The first quadrupole contains four opposing electric field rods (two positive and two negative) that separate ion trajectories by mass-to-charge ratio (m/z). The masses of selected compounds are calculated based on the structure of the compound. The selected precursor ions move through the first quadrupole without touching the charged rods, while unwanted compounds meet the charged rods and are deemed neutral and undetectable. The precursor ions are fragmented in the second quadrupole, or collision chamber/cell, by inactive argon gas (via CID) to create fragmented product ions. Selected product ions are filtered in the third quadrupole, based on the known mass of the fragments desired, to achieve precise identification of the lipid species of interest.

The use of commercial standards and an increase in the assessment of multiple fragmentation patterns (transitions) for each species allows for a low probability of identifying false positive molecules. However, assessment of multiple transitions decreases the ability of the mass spectrometer to identify fragments; there is a limit to the number of transitions available for analysis from each injection. The addition of many extra transitions increases the number of injections required for analysis, leading to increases in mass spectrometry time and processing time (Arnott, 2001; Matuszewski *et al.*, 2003; Lu *et al.*, 2017). Challenges still remain with mass spectrometry-based lipidomics due to lack of high throughput analyses and the insufficient number of internal standards and reference standards available for each lipid species now measureable (Murphy *et al.*, 2011b).

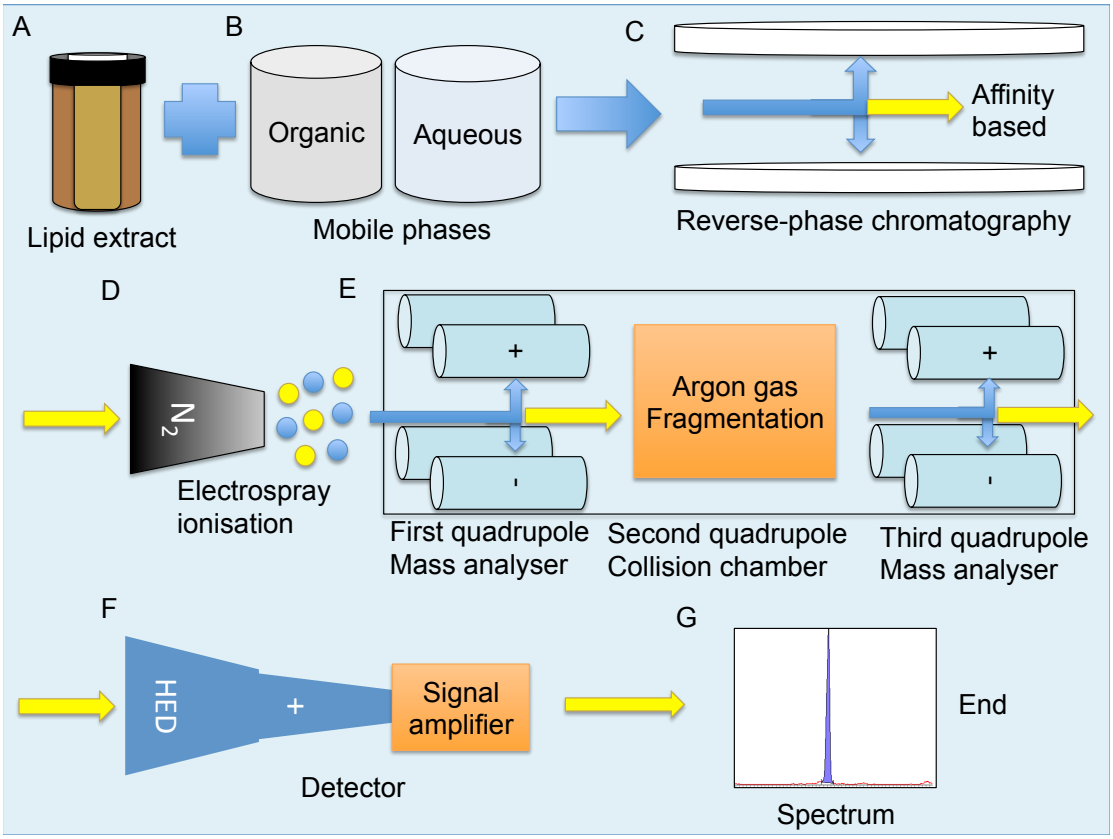


Figure 1-5: Schematic overview of the LC-ESI-MS/MS analysis pipeline

The lipid extract (A) is separated on a reverse-phase column (C) by affinity with the mobile phases (B). The lipids selected with the correct affinity/hydrophobicity for species of interest, move into the cone for electrospray ionisation (ESI; D) to gas ions. Precursor ions are selected based on their mass-to-charge ratio in the first mass spectrometry quadrupole, they are fragmented in the second quadrupole, and expected product ions are selected in the third quadrupole of the mass spectrometer (E). The product ions are funnelled through a high-energy dynode (HED; F) and the analog signal is recorded and converted to a digital signal as spectra (G).

1.4. Genetic analyses

1.4.1. Heritability

Heritability is the portion of phenotypic variance due to genetic factors, as opposed to non-genetic, environmental factors like diet and measurement error. Heritability allows for the comparison of different traits within a population, determines the efficiency of prediction of genetic risk of disease, and can lead to insights for intervention strategies. Heritability is dependent upon the population it is estimated in because both the genetic and environmental factors are specific to a population, e.g. allele frequencies and diet (Visscher *et al.*, 2008).

Narrow-sense heritability (h^2) estimates the variance in a phenotype due to additive genetic variance, the simple additive effects of alleles, allowing for the estimation of the correlation between relatives depending on additive effects only. Broad-sense heritability (H^2) estimates the variance that is due to total genetic values such as dominance, epistatic and parental effects, which are assumed to be negligible when estimating h^2. The difference between dominance and additive effects is depicted in Figure 1-6. Most relatives (excluding monozygotic twins and full siblings) share less than one copy of alleles that are identical by descent (IBD), so effects based on sharing two copies of alleles (dominance, interactions, etc.) do not contribute to their phenotypic resemblance. Thus, h^2 (as opposed to H^2) has been more studied in cohort analyses as the correlation of most relatives depends on h^2 only (Visscher *et al.*, 2008).

Observed variation (e.g. plasma lipid concentration) is partitioned into unobserved genetic and environmental factors. The estimate of heritability is a measure of the amount of variation in the phenotype between individuals in a population, in a given environment, that is due to their genotypes, and it determines the resemblance between relatives (Visscher *et al.*, 2008). Such genetic variance is estimated on a scale of 0.00-1.00 (or as a percentage, i.e. 0% - 100%), with the remaining variance due to environmental factors. Height is an example of a trait that is highly heritable, the variation in measured height is estimated to be ~80% due to genetic factors (Silventoinen *et al.*, 2003; Yang *et al.*, 2010), and therefore ~20% is estimated to be

due to non-genetic, environmental factors. Nutrition and lack of dietary protein is the most important environmental factor influencing height (Bozzoli *et al.*, 2009).

While a large estimate of heritability implies a strong correlation between phenotype and genotype, and the genomic loci influencing a trait can be more easily detected, it does not give information on the number of loci affecting a trait. A trait with a low estimate of heritability can be influenced by a single, large-effect genetic locus, and the converse is also true; a trait estimated as highly heritable can have hundreds of small effect loci, for example, blood pressure (Evangelou *et al.*, 2018) and height (Yang *et al.*, 2010). A mixture of large and small effect loci can also influence moderately heritable traits (Lorenz *et al.*, 2012). However, a highly heritable phenotype is more likely to have a major-gene effect. Examples of this are the gene of angiotensin converting enzyme (*ACE*) on plasma ACE, and the gene of lipoprotein(a) (*LPA*) on plasma lipoprotein(a) (Lp(a)). Among the large number of measureable lipids, identification of heritable species will allow for the identification of a subset of lipids in which a major gene effect can be detected.

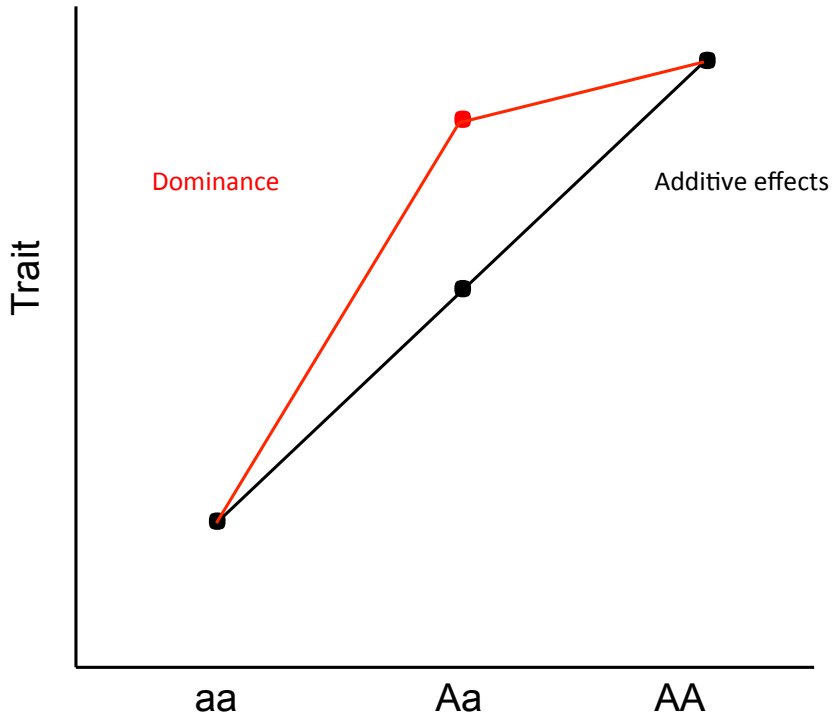


Figure 1-6: The difference between dominance and additive genetic effects

The figure depicts a comparison of dominance (red) and additive (black) acting genetic effects in the variation of an example trait (Y-axis) by an example genotype (X-axis). Additive effects (black) are assumed to increase with genotype in a linear manner. Dominance effects (red) show a near equal increase in a trait based on the presence of one or two dominant alleles. The most common heritability and GWAS protocols assume genotypes have additive effects.

1.4.2. Family-based genetic analyses

Family-based genetic studies have dominated genetic analyses as far back to Mendel's concepts of inheritance in plants. Linkage analyses involving families with affected individuals have led to the discovery of many Mendelian diseases and traits. Heritability traditionally was estimated from the analyses of regressions of offspring phenotype values on mean parental values, by methods such as path analysis (Wright, 1921), or correlations between siblings and twin studies. These preliminary analyses provided a backbone for the estimation of variance by assessing further correlations between extended relatives, using the expected resemblance through the fraction of alleles that are IBD. Using ANOVA, this allowed for the assessment of offspring within an immediate family and between families, and commenced the analyses of transmission-disequilibrium test (TDT) for parent-offspring trios (Spielman *et al.*, 1998) and quantitative trait-disequilibrium test (QTDT) of extended family-based cohorts (Abecasis *et al.*, 2000), around the time the family-based cohort analysed in this study was collected.

An example of this methodology is if a single gene influences a trait, the covariance between the trait values for two individuals depends on the genetic variances and kinship coefficient. The kinship coefficient is half the expected proportion of alleles shared IBD, inherited from the most recent common ancestor, representing the degree of relatedness between pairs of individuals (e.g. unrelated = 0, identical twins = 1, parent-offspring/full-siblings = ½, half-siblings = ¼). Those who are more distantly related have lower kinship values and less genetic correlation is expected in their trait values.

Linear mixed model approaches (LMM) were used as the backbone of variance components linkage analyses. While traditional family-based tests are robust, they have lower power due to the lower number of effective sample size as controls. LMM approaches adjust for relatedness and allow for the analysis of all data provided, from both related and unrelated individuals. Furthermore, the use of pedigree data in family-based analyses assumes that the reported pedigrees are correctly specified and all family founders are completely unrelated, sharing no alleles identical by descent. This is usually violated in real family-based studies.

Since 2010, SNP-based estimations of heritability using LMM approaches have become preferential in modelling phenotypic correlations between individuals (Yang *et al.*, 2010). LMMs are used to account for relationships and population stratification, and can estimate the heritability partitioned by measured SNPs to investigate genetic correlations between traits. LMM is a model in which the dependent variable is a linear function of both fixed and random independent variables. Fixed effects are fixed at their measured values (e.g. single nucleotide polymorphisms (SNPs)) while random effects are sampled from a distribution. This adjusts for relatedness by taking into account kinship coefficients without random effects (family correlations) and only the fixed effects (SNPs) are tested.

The use of LMMs with the addition of estimating a genetic relationship matrix (GRM) allows for the adjustment for relatedness in family-based or seemingly unrelated cohorts for both heritability and GWAS analyses. Furthermore, Yang *et al.*, showed that modelling the effects of all genotyped SNPs simultaneously, including those in linkage disequilibrium and not at GWAS significance, explained the full reported heritability for height (~80%). Such an estimate is called SNP-based heritability and is modelled via LMM. This SNP-based heritability, and a complimentary estimate of pedigree-based heritability using variance components analysis (QTDT), are undertaken in this project. Variance components analysis uses a more general model than that of LMM, calculating the kinship matrix based on the reported pedigrees.

1.4.3. Genome-wide association studies

GWAS are unbiased scans of genetic markers across the genome to identify genetic variants that statistically associate with variations in a phenotype, the frequency of the alleles vary as a function of phenotypic trait values (Marees *et al.*, 2018). GWAS are used to identify associations between common DNA variants (minor allele frequency >1%) in a population and the variance in a phenotype. For DNA variants of small effect sizes on a phenotype, association analyses have greater power over traditional linkage analyses. Association testing measures the correlation between alleles near a marker locus with disease status across individuals. If significant, either the marker is causal to the disease/phenotype or the marker is in linkage disequilibrium with the

locus having the effect; alleles found close together on the genome tend to correlate with each other due to lack of recombination over many generations.

Since 2002, GWAS have been used to map DNA loci to traits (Ozaki *et al.*, 2002), however in 2007, the approach was first validated through GWAS that coupled genome-wide coverage with a substantial sample number and strong replication, identifying susceptibility variants for several diseases (Burton *et al.*, 2007). Candidate gene studies, the genotyping of only known risk genes or variants, have not been as successful as GWAS that genotype polymorphic markers dispersed at intervals across the genome. The variants included on genotyping arrays are based on surveys of human genetic variation, such as the HapMap Project (Belmont *et al.*, 2003) and 1,000 Genomes Project (Auton *et al.*, 2015), ignoring genomic loci that are identical across humans and loci that are in high linkage disequilibrium in European populations. Genotyping chips allow for the analysis of variants outside protein-coding regions (as opposed to exome-only analyses), are unbiased to the discovery of variants, and are exceedingly successful with the ability of imputation techniques, such as the Haplotype Reference Consortium (McCarthy *et al.*, 2016) which can impute the non-genotyped regions of DNA in a genotyped cohort, based on reference genomes. In it's first release, the reference consisted of 64,976 haplotypes at 39 million SNPs, all with an estimated minor allele count of greater than 5, from 34,000 individuals of multiple cohorts.

While GWAS do not require families, association analyses based on parent-offspring trios were at the forefront of genetic studies in the 1980s-1990s (Rubenstein *et al.*, 1981), the time at which this cohort was collected. GWAS studies in large numbers of individuals have identified loci where common genetic variation influences the prevalence of major plasma lipid species, such as HDL- and LDL-cholesterol, triacylglycerides, and PUFAs (Tanaka *et al.*, 2009; Teslovich *et al.*, 2010; Willer *et al.*, 2013). These study provide evidence that although lipids are not DNA-encoded, certain species are heritable and have been found to significantly associate with DNA variants at GWAS significance, mainly in the genes of enzymes involved in their metabolism.

1.4.4. Two-sample Mendelian randomisation

Mendelian randomisation (MR) allows for testing the causality of a trait on disease risk, using genetics. That is, the assessment of whether a trait, such as a lipid, has a causal role in disease. Genetic analyses are useful in this setting as the DNA code is set at birth and mostly invariable. Therefore, it is not influenced or altered over time or by external/environmental facts, as an assessment of lipid levels and disease risk is, by potential factors, for example diet. Therefore, measuring lipid levels in a case-control analysis to understand their role in disease could be influenced by such factors. However, if a DNA variant that is shown to increase disease risk through a disease GWAS, and independently shown to alter blood lipid levels via a lipid GWAS, the DNA variant confirms the causal role of the lipids in disease risk, as an individual with the variant will have both an increased risk of disease and altered lipid levels. The DNA variant in this setting is called an "instrument" as it is used to assess the relationship between lipids and disease without the influence of cofounding factors.

MR is a framework for the causal inference of risk factors on disease (depicted in Figure 1-7). While inference is regularly assessed, causal inference is difficult to confirm in observational epidemiology due to unobserved or unmeasured confounding (e.g. existing disease and environmental factors), reverse causation, and selection bias. MR uses SNPs as instrumental variables, intermediaries used to estimate causal effects, as genetic variants are fixed at conception and are mostly invariable. MR estimates the causal effect of an exposure (e.g. risk factor) on an outcome (e.g. disease) using a SNP as a genetic instrument; the exposure is estimated from the instrument (e.g. by the number of alleles) and the outcome is regressed on the exposure to obtain a causal effect estimate. MR is analogous to randomized control trials with an unconfounded exposure-outcome relationship. The approach answers whether the exposure defines the status of the outcome, and causality can therefore be addressed. Two-sample MR (2SMR) is the use of two different study samples to estimate the instrument-risk factor and instrument-outcome associations, which can be from published data (Davies *et al.*, 2018; Teumer, 2018)

Particularly relevant to this study are the discoveries regarding the genetic control of lipoprotein(a) (Lp(a)), a highly heritable lipid particle with roles in cardiovascular disease (CVD). Lp(a) was discovered in 1963 by Kare Berg using antigens for serum lipoproteins. While its role in CVD was debated for some time, genetic studies helped resolve the epidemiological controversies. In 1991, a family study showed that Lp(a) was highly heritable in 12 families (Lackner *et al.*, 1991). Although Lp(a) was found to associate with risk within a population, between populations some of the risk were inverse, for example, women have higher Lp(a) then men (Frohlich *et al.*, 2004) although women are at lower risk of CVD morality (Bots *et al.*, 2017). However, Clarke et al. (2009) undertook a case/control analysis to show that two DNA variants in the gene *LPA* correlated with both levels of Lp(a) and coronary artery disease; Lp(a)'s role was confirmed as causal in coronary artery disease. The lipoprotein is highly heritable (~75%) and mostly unaltered by environmental factors (Berg, 1994; McCormick, 2004; Clarke *et al.*, 2009). The history of Lp(a) is an example of the importance of genetic studies for the discovery of lipidomic biomarkers of disease risk. Another example includes the causal role of triacylglycerides on coronary heart disease risk (Holmes *et al.*, 2015).

The identification of novel, causal, intervenable risk factors is a priority for cardiovascular epidemiology. Observational epidemiology has led to a number of "blind alleys" where strongly associated risk factors have, when modified in clinical trials, proven not to affect disease risk; examples include plasma HDL-cholesterol (Wright, 2013). The massive resource costs of drug development and clinical trials aiming to modify putative risk factors which turn out to be non-causal could perhaps be avoided in future through the use of MR, on which there is now extensive literature (Keavney *et al.*, 2006; Timpson *et al.*, 2012; Voight *et al.*, 2012; Smith *et al.*, 2014; Holmes *et al.*, 2015; Mokry *et al.*, 2015; Burgess *et al.*, 2016).

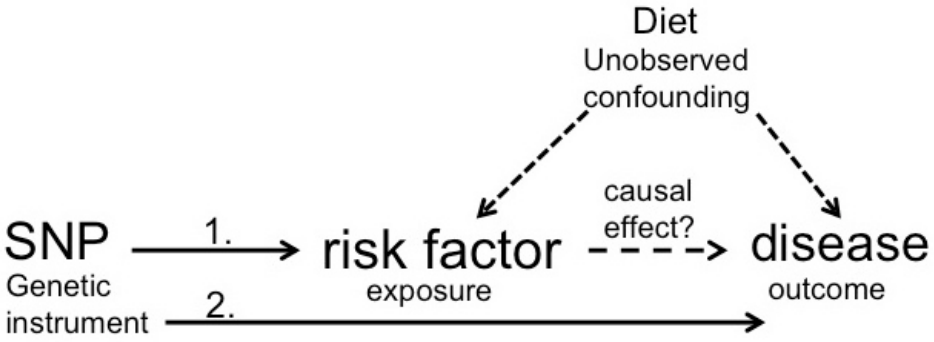


Figure 1-7: Schematic of Mendelian randomisation

Causal inference can be assessed by observational epidemiology, but this is affected by unobserved, unmeasured confounding, such as diet influencing both a risk factor and the presence of disease. Mendelian randomisation (MR) overcomes this by using genetic factors as instruments to estimate the causal effect of a risk factor on disease. For example, if a GWAS association identifies a SNP to associate with risk of disease (instrument-outcome assessment - depicted as point 2), and independently associates with a risk factor (instrument-exposure assessment - depicted as point 1), then the causal effect of the exposure on the outcome can be estimated using the genetic factor as an instrument, as it is not affected by confounding and is mostly invariable. The two assessed steps can be completed using two separate cohorts, denoted as "two-sample Mendelian randomisation" (2SMR). This figure is adapted from Teumer, 2019.

1.5. Study hypothesis, aims, and objectives

As described, CVD is the leading cause of death in western countries, therefore there is a need for novel diagnostic biomarkers and targets of treatment (Lawes *et al.*, 2008; Hoefer *et al.*, 2015). Dyslipidaemia has a key role in atherosclerotic CVD risk (Libby *et al.*, 2011) and there is a growing body of evidence implicating blood lipid mediators of the CER, NAE, and Eico classes as potential biomarkers of CVD risk; circulating apoptotic CER[NS] species with an 18-C sphingosine backbone have been shown in epidemiological studies to predict risk of CVD (Siskind *et al.*, 2010; Havulinna *et al.*, 2016; Laaksonen *et al.*, 2016); NAE and eCB species have roles in satiety (among other roles), and have been targeted by pharmacology as potential treatment of CVD (Rodríguez de Fonseca *et al.*, 2001; Mach *et al.*, 2009; Pacher *et al.*, 2013; Mallet *et al.*, 2016), and vasoactive, inflammatory Eico species have been found increased in the blood of CVD patients compared to healthy controls (Theken *et al.*, 2012; Schuck *et al.*, 2013; Oni-Orisan *et al.*, 2016).

The question of causality and mechanism of involvement of these three lipid groups in CVD is undetermined, and there are likely confounding factors influencing the current evidence, such as diet and underlying inflammation. However, although such signalling lipids are not DNA-encoded, enzymes and other proteins strictly regulate their metabolism and activities, and the few genetic studies completed have shown that some blood lipid species from these classes, or their respective precursor PUFAs, are substantially heritable and in the few cases studied by GWAS, have associated to GWAS significance with common DNA variants that influence their levels in circulation (Bray *et al.*, 2007; Tanaka *et al.*, 2009; Bellis *et al.*, 2014; Long *et al.*, 2017; Tabassum *et al.*, 2019).

The ability of MR techniques to estimate the causality of traits in disease without the influence of confounding factors that inhibit epidemiology studies, allows for the question of causality to be addressed for these lipid species and confirm or deny their roles as causal in CVD (Davies *et al.*, 2018). While low concentration circulating lipid mediators are now measurable by targeted mass spectrometry techniques (Quehenberger *et al.*, 2010; Stephenson *et al.*, 2017; Kendall *et al.*, 2019), genetic analyses have not been completed for the numerous uniquely structured, circulating,

bioactive lipid species of each of the three lipid classes studied here, to date. A cohort of substantial size for genetic analyses, with genotyping data and blood samples available for mass spectrometry-based lipidomics, and an exemplar of a subset of a general population, was required to enable the assessment of causality of these low concentration lipids on CVD.

This study's hypothesis is that a subset of low concentration, plasma CER, NAE, and Eico species are substantially heritable, and the major DNA variants influencing their plasma levels can be found using GWAS in a moderately-sized cohort. The aim of the project was to measure the low concentration plasma lipid species from the NAE, CER, and Eico classes of lipids in the blood of a cohort of genotyped participants to estimate the heritability of the lipids, identify the DNA variants with major effect over the levels of the circulating lipid species, and confirm if the lipid species are causal in attenuating CVD, or other diseases and phenotypes, using two-sample MR (Figure 1-8).

The following specific questions in the field will be addressed;

1. A subset of CER species have been explored as biomarkers of cardiovascular disease (Laaksonen *et al.*, 2016) and type-2 diabetes (Havulinna *et al.*, 2016; Jensen *et al.*, 2019). Three genetic studies have identified DNA variants associating with high concentration blood CER species (Hicks *et al.*, 2009; Demirkan *et al.*, 2012; Tabassum *et al.*, 2019). The causal involvement of CERs in CVD remains unclear. Identification of the common variants influencing an extended array of CER species in plasma will allow for the assessment of causality of the lipid species in CVD via two-sample MR (Davies *et al.*, 2018).

2. The *N*-acyl ethanolamine species (NAE), particularly eCB anandamide (AEA), have been investigated as targets of drug therapeutics for CVD (Sipe *et al.*, 2005; Schaich *et al.*, 2014). To date only one GWAS has been completed measuring oleoyl ethanolamide (OEA) via shotgun lipidomics (Long *et al.*, 2017). The identification of further DNA variants associating with this class of lipids would aid current therapeutics in progress that target

the NAE pathway in the aim of modulating eCB signalling without side effects (Mallet *et al.*, 2016), by providing further information on the underlying genetic influences on this class of lipids.

3. The expression of the enzymes involved in the respective lipid pathways is known and is mostly systemic (Uhlen *et al.*, 2015), however it is unclear which tissues contribute most to the plasma pool of lipids. Identification of the main DNA variants influencing these plasma lipid species may reveal expression quantitative trait loci (eQTLs); variants that have been confirmed to alter the expression of genes in specific tissues, and thus highlight the tissues having a major influence on plasma lipid species concentrations.

4. The identification of highly heritable lipid species from each class of lipids studies will highlight a subset of genetically-influenced lipids for follow up in future large cohort studies to further explore associations with disease.

The study objectives are as follows:

1. Assess the quality of the lipidomics results
2. Assess the bioinformatics approaches used to undertake genetic analyses
3. Undertake a range finding study of all three classes of lipids (CER, NAE, and Eico) in a subset of the cohort plasma samples to identify the most heritable lipids
4. Analyse the more genetically-influenced lipid classes in the full cohort
5. Complete GWAS to identify the common DNA variants influencing plasma concentrations of the lipids measured
6. Undertake further exploration of the GWAS results to discover whether the DNA variants identified;
    a. are found near the genes of enzymes involved in the respective lipid metabolic pathways
    b. are confirmed to alter the expression of genes in specific tissues (i.e. are eQTLs)

c. have been identified in previously published GWAS of disease and other phenotypes
d. are causal in disease or in influencing traits, through two-sample MR analysis

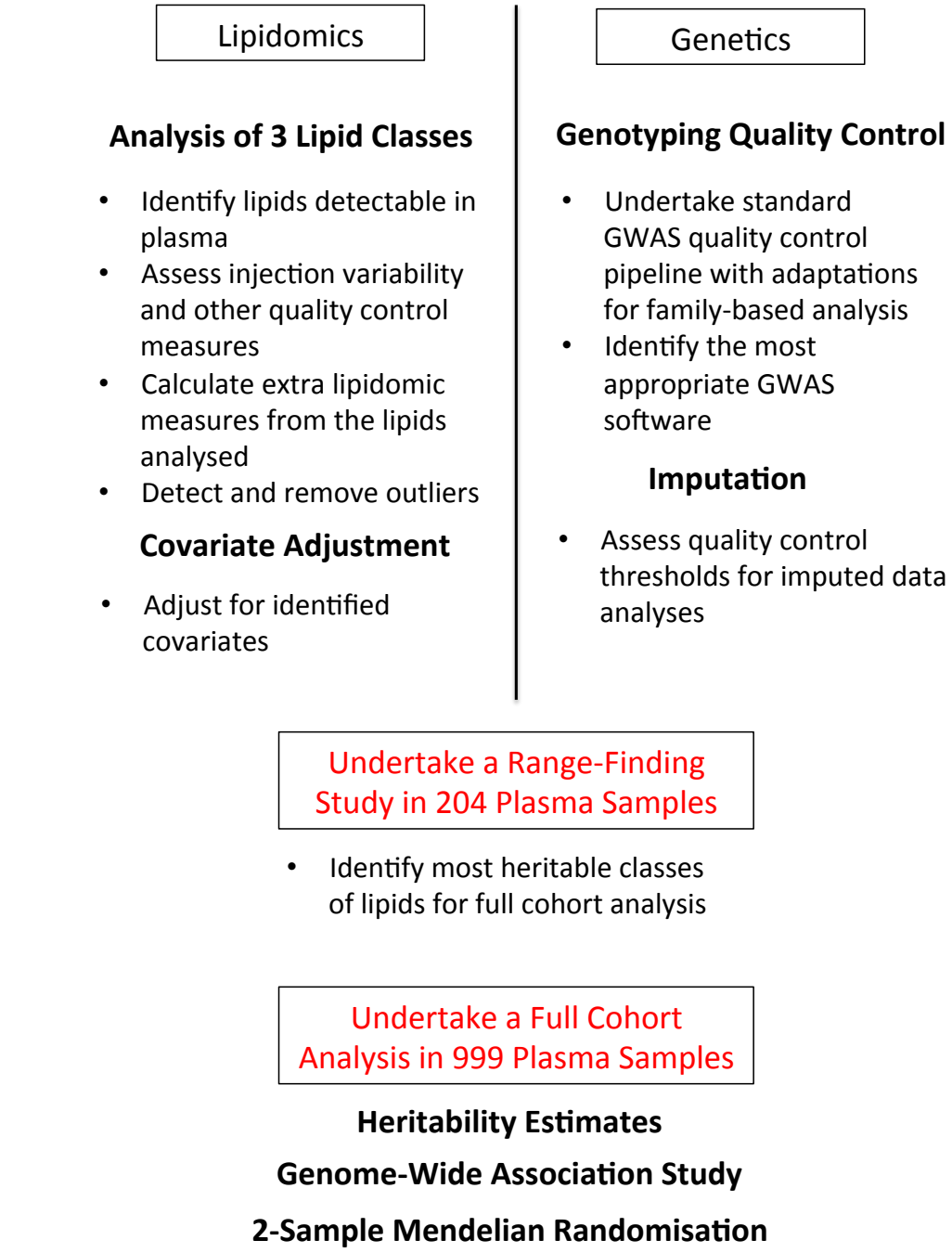


Figure 1-8: Diagrammatic overview of thesis plan

At the commencement of this project, raw genotyping data and unopened plasma samples were available for the cohort. To understand the genetic influence over three classes of bioactive plasma lipids, a range-finding study was undertaken to identify those lipids at particular genetic influence in a subset of the cohort (204 plasma samples). The information gained from this study allowed for a focused lipidomics

assessment of the most genetically-influenced lipid species using the full cohort samples for the final genetic analyses (999 plasma samples).

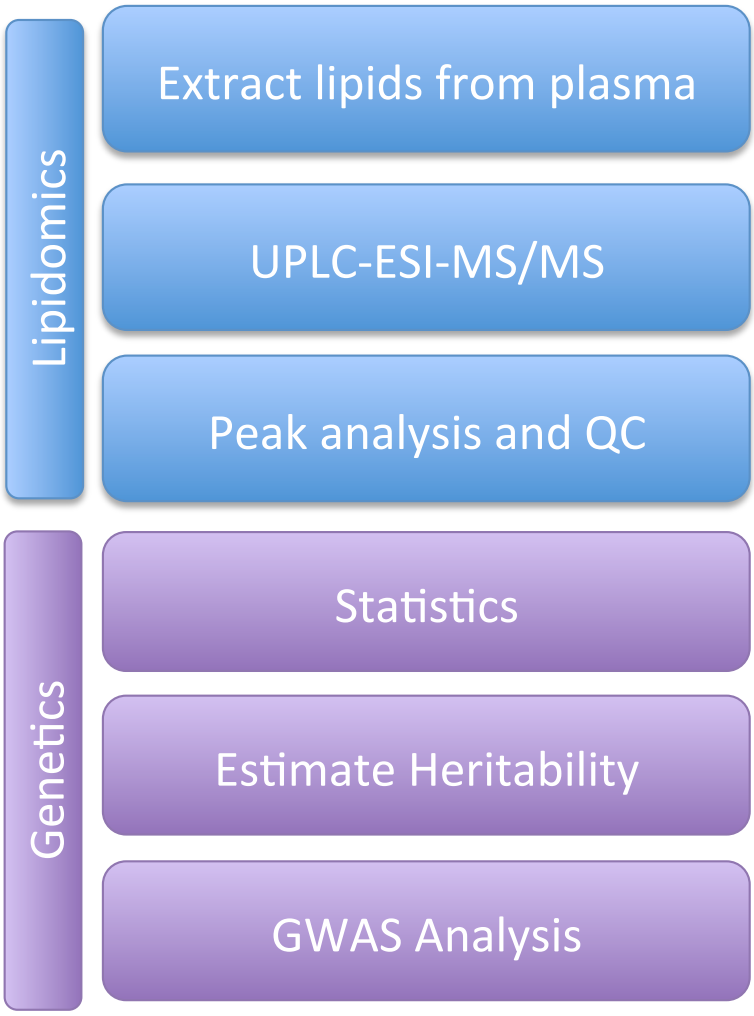
Having undertaken a lipidomics analysis, the objectives were to identify lipids from the three classes that were quantifiable in the plasma samples, assess the quality of the lipidomics data, calculate ratios and summations for the lipid classes analysed, and to identify and remove outliers. From this data, covariates were assessed to identify those of particular influence over the lipid values, and adjust for them accordingly.

With the genotyping data, standard quality control was required before completion of the genome—wide association study (GWAS) and SNP-based heritability assessment. Standard genotyping quality control for common SNPs was undertaken in this family-based cohort to subset SNPs of high quality for SNP-based genetic analyses. Family-based GWAS software were assessed to identify the most appropriate for the analysis. HRC imputation was to be undertaken and more specialised quality control for the imputed data analyses was assessed to identify quality thresholds that retained only the most confidently imputed SNPs.

Once the initial lipidomics and genetics procedures and results were identified from the range finding study, the full cohort integrated them. The results address the three main questions; 1) how influenced are the lipids by genetics? (heritability estimates); 2) where on the DNA are common variants found that are involved in such genetic influence (GWAS); 3) are the variants causal in cardiovascular disease or other phenotypes, potentially highlighting them as novel biomarkers? (2-sample Mendelian Randomisation).

Chapter 2

Materials and methods



2.1 Study participants and phenotyping

The Hypertension Oxford (HTO) Family Cohort has been shown to have adequate power to detect moderate-sized genetic influences on quantitative traits, as exampled by previous publications (Table 2.1). The cohort allowed for plasma lipidomic analysis of 999 individuals (196 families) in the time frame of the project, as the genotyping data was readily available. The cohort is an exemplar of the British Caucasian Population. This number of samples is modest for genetic studies, but significant genotype associations with quantitative traits have been identified through analysis of this cohort previously. Of note is the study of angiotensin converting enzyme (ACE), where plasma ACE activity was measured by HPLC with the use of a synthetic substrate. PCR genotyping was completed for ten ACE locus polymorphisms, and through haplotype analysis the cohort identified DNA variants influencing variation in ACE activity (Keavney *et al.*, 1998).

2.1.1 Cohort recruitment

Participants from 248 families (1,425 individuals) were recruited for a quantitative genetic study of hypertension and other cardiovascular risk factors, selected via a proband with essential hypertension between 1993 and 1996 (secondary hypertension was excluded using standard clinical criteria) (Keavney *et al.*, 2005; Mayosi *et al.*, 2008). Probands were recruited from outpatients attending the John Radcliffe Hospital Oxford hypertension clinic or via their family doctors. Hypertensive Proband eligibility included an ambulatory blood pressure monitor reading of mean daytime systolic blood pressure >140 mm Hg and mean daytime diastolic blood pressure >90 mm Hg, repeated clinic readings above >160 mm Hg and >95 mm Hg, or taking two or more antihypertensive drugs and still found to be hypertensive (>140/90 mmHg clinical pressure).

Included families were U.K. residents of self-reported white ethnicity and were required to consist of 3 or more siblings quantitatively assessable for blood pressure if one parent of the sibship was available for blood sampling, or 4 or more siblings if no parent was available. First, second and third degree relatives were then recruited to assemble a series of extended British families. Approximately a third of all

participants were found to have hypertension, representing that of the British Caucasian population (Joffres et al., 2013; Public Health England, 2018).

The collection protocol obtained ethical clearance from the Central Oxford Research Ethics Committee (ethics application number: 06/Q1605/113) and it corresponds with the principals of the Declaration of Helsinki. Written informed consent was obtained from all participants.

2.1.2 Phenotyping

The participants were fully phenotyped for cardiovascular risk factors, blood biochemical measures, anthropometric traits, and plasma was stored. The time of day at which blood was drawn was not standardised, although the majority of visits took place in the early evening, and individuals were not fasting. Trasylol and O-phenanthroline protease inhibitor was added to the EDTA anticoagulant. Blood samples were kept on ice and plasma separated and frozen at −80°C within four hours of the blood draw.

DNA was extracted from whole blood by standard methods and quantified by a proprietary Pico Green assay (Singer *et al.*, 1997). Genotyping was performed using the Illumina 660W-Quad bead array chip that includes 557,124 SNPs at the Centre National de Genotypage (Evry Cedex), France.

Upon commencement of this project, 999 plasma samples (2ml, EDTA) were available for lipidomics analyses, with the corresponding genotyping data available in PLINK (Purcell *et al.*, 2007) binary format containing information on 557,124 SNPs for 1,234 individuals. For the range finding study (Chapter 4), families were only included if their DNA was genotyped and two aliquots of EDTA samples were available to avoid freeze/thaw-cycles (one aliquot for the Eico assay, and one for the NAE and CER joint assay), this resulted in 31 families of 196 participants.

Table 2.1: Significant genetic analyses using the Hypertension Oxford Family Cohort in literature

The table depicts the significant heritability and genetic associations identified previously using the British Caucasian family-based cohort.

Trait	Reference	Comment
Angiotensin-1 converting enzyme gene	(Keavney *et al.*, 1998)	Cladistic/measured haplotype analysis of polymorphisms in the gene allowed for localisation of variants with influence on ACE variability
ECG measures of left ventricular hypertrophy more heritable than Echo	(Mayosi *et al.*, 2002)	Significant heritability estimated for Sokolow-Lyon voltage, echocardiographic left ventricular mass, electrocardiographic left ventricular mass, Cornell voltage, and Cornell product
Interleukin-6 gene associated with plasma CRP	(Vickers *et al.*, 2002)	Significant estimate of heritability identified for CRP and association between CRP and Interleukin-6 gene
Haploype analysis of aldosteron synthase gene and heart size	(Mayosi *et al.*, 2003)	Significant association between CYP11B2 and cardiac wall thickness and left ventricular cavity size
Plasma 11-deoxycortisol and cortisol, urinary tetrahydrodeoxycortisol	(Keavney *et al.*, 2005)	Significant heritability estimated for 11-deoxycortisol and cortisol, estimated and candidate gene study associations with CYP11B1 and CYP11B2
Leptin gene in blood pressure and carotid intima-medial thickness	(Gaukrodger *et al.*, 2005)	Significant association for pulse pressure and carotid intima-medial thickness and leptin gene
Proopiomelanocortin gene association with body fat distribution	(Baker *et al.*, 2005)	Significant association between proopiomelanocortin gene and waist-to-hip ratio
Interleukin-6 gene associated with carotid artery intimal-media thickness	(Mayosi *et al.*, 2005)	Significant heritability estimated for carotid artery intimal-media thickness and association of Interleukin-6 gene and maximal carotid IMT
Aldosterone association with variants in CYP11B1	(Imrie *et al.*, 2006)	Significant heritability of urinary tetrahydroaldosterone excretion estimate and significant association with CYP11B1 found
INSIG-2 gene associated with obesity	(Hall *et al.*, 2006)	No association found between a SNP in upstream to INSIG-2 and obesity-related phenotypes
Angiotensinogen gene associated with pulse pressure	(Baker *et al.*, 2007)	Significant association between angiotensinogen gene and pulse pressure
ECG and Echo	(Mayosi *et al.*, 2008)	Genome-wide linkage analysis identified associations with Sokolow-Lyon voltage, electrocardiographic left ventricular mass, and echocardiographic left ventricular mass and chromosomal loci
Ambulatory blood pressure association with P2X receptor genes	(Palomino-Doza *et al.*, 2008)	SNP association in P2X7 gene with blood pressure
Plasma potassium level associated with beta-subunit of epithelial sodium channel gene	(Gaukrodger *et al.*, 2008)	Significant association between SCNN1B gene and plasma potassium

STK39 association with blood pressure	(Cunnington *et al.*, 2009)	No association was found between SNPs tested in STK39 and blood pressure measurements, with blood cell allelic expression differences identified
Familial and phenotypic associations of the aldosterone renin ratio	(Alvarez-Madrazo *et al.*, 2009)	Significant heritability estimates for aldosterone/renin ratio and association between CYP11B2 and plasma aldosterone
CD36 association with left ventricular mass	(Hall *et al.*, 2011)	Significant association between CD36 gene and left ventricular mass
Hexose-6 phosphate dehydrogenase associated with carotid intima-medial thickness	(Rahman *et al.*, 2011b)	Significant association between H6PD gene and mean carotid intima-medial thickness measurement
Left ventricular mass associated with 11-beta hydroxysteroid dehydrogenase type 1	(Rahman *et al.*, 2011a)	Significant association between HSD11B1 gene and left ventricular mass
MiRNA 22 associated with left ventricular mass	(Harper *et al.*, 2013)	Significant association between miR-22 locus and left ventricular mass determind by Sokolow-Lyon voltage
Variation in plasma angiotensin-I converting enzyme shows allelic heterogeneity in the ABO blood group locus	(Terao *et al.*, 2013)	Significant association between alleles and intermediate plasma ACE activity, with heterogeneity among A alleles
Left ventricular phenotypes	(Nethononda *et al.*, 2019)	Heritability estimates of left ventricular phenotypes are greater when measured by ECG than CMR

The table depicts the significant heritability and genetic associations identified previously using the British Caucasian family-based cohort.

2.2 Plasma Lipidomics

2.2.1 Materials

Methyl formate HPLC grade ≥99.0%, methanol HPLC grade ≥99.9%, hydrochloric acid ACS grade; 36.5-38%, acetonitrile LC-MS grade ≥99.9%, acetic acid HPLC grade ≥99.7%, formic acid for mass spectrometry ~98% (Sigma Aldrich, UK). Ethanol HPLC grade >99.8%, chloroform HPLC grade >99.8%, hexane HPLC grade ≥97.0%, isopropanol HPLC grade >99.8% (Fisher Scientific, UK). Four deuterated internal standards all 10ng/µl in ethanol: PGB_2-*d*4, 8,9-EET-*d*11, 8,9-DHET-*d*11, 12-HETE-*d*8 (Cayman Chemical, US). 1ml of 25 µM Ceramide/Sphingoid Internal Standard Mixture I (Avanti Polar Lipids, USA) containing 25 µM of 10 compounds in ethanol: C17S, C17S1P, C17DS1P, lactosyl(ß) C12 ceramide, 12:0 sphingomyelin, glucosyl(ß) C12 ceramide, 12:0 ceramide (CER[N(12)S(18)]), 12:0 ceramide-1-P, 25:0 Ceramide (CER[N(25)S(18)]). *N*-acyl ethanolamine deuterated internal standard anandamide-*d*8 500 µg/500µl and 2-arachidonyl glycerol-*d*8 deuterated internal standard 50 µg/500µl (Cayman Chemical, US). Lipidomic standards all 10 ng/µl in ethanol; PGD_1, PGE_1, 13,14-dihydro-15-keto PGE_1, 6-keto $PGF_1\alpha$, 13,14-dihydro-15-keto $PGF_1\alpha$, $PGF_1\alpha$, PGD_2, PGE_2, 15-keto PGE_2, 13,14-dihydro PGE_1, 13,14-dihydro $PGF_2\alpha$, 13,14- dihydro $PGF_1\alpha$, 13,14-dihydro-15-keto $PGF_2\alpha$, $PGF_2\alpha$, 8-iso $PGF_2\alpha$, $PGF_3\alpha$, Δ12-PGJ_2, PGJ_2, TXB_2, TXB3, 15-deoxy-Δ12, 14-PGJ_2, 13,14-dihydro-15-keto PGE_2, PGD_3, PGE_3, RvD_1, RvD_2, MaR_1, isomer of PD1 (PDX), 9-HODE, 13-HODE, ± 4-HDHA, ± 7-HDHA, ± 10-HDHA, ± 11-HDHA, ± 13-HDHA, ± 14-HDHA, ± 17-HDHA, ± 20-HDHA, HXA3, RvE_1, LTB_4, ± 8,9-DHET, ± 11,12-DHET, ± 14,15-DHET, ± 5,6- DHET, ± 5(6)-EET, 8(9)-EET, ± 11(12)-EET, ± 14(15)-EET, ± 5-oxoETE, 15-HETrE, 5-HETE, 8-HETE, ± 9-HETE, 11-HETE, 12-HETE, 15-HETE, 20-HETE, 5-HEPE, ± 8-HEPE, ± 9-HEPE, ± 11-HEPE, 12-HEPE, ± 15-HEPE, ± 18-HEPE, 9-HOTrE, 13-HOTrE, 19(20)-DiHDPA, 9(10)-EpOME, 12(13) EpOME, 9-oxoODE, 13-oxoODE, 5(15)-DiHETE, 8(15)-DiHETE, 19(20)-EpDPE, 16(17)-EpDPE, trans-EKODE, 9,10-DiHOME, 12,13-DiHOME, AEA, 2-AG, ALEA, DHEA, OEA, EPEA, PEA, LEA, MEA, PDEA, POEA, HEA, DGLEA, DPEA, DEA, NEA, LGEA, NAT, PGF2α-EA, PGE2-EA, PGD2-EA, 15-HETE-EA,

5,(6)-EET-EA, 8,(9)-EET-EA, 11,(12)-EET-EA, 14,(15)-EET-EA, 2-PG, 2-LG, 2-OG, 1-LG, 1-OG, 2-STG (Cayman Chemical, US or Santa Cruz Biotechnology, US).

2.2.2 Equipment

Sorvall refrigerated centrifuge (RT60000B; Thermo Fischer Scientific, UK), whirl mixer (Fisher Scientific, UK), nitrogen-drying cabinet (custom to Nicolaou laboratory), solid phase extraction vacuum manifold (Phenomenex, UK), vacuum pump (1c;Vacuubrand, Germany), glass Pasteur pipettes (150 mm unplugged; Fisher Scientific, UK), round- and flat-bottomed glass tubes (Fisher Scientific, UK), Hamilton glass syringes (50, 100, 250, and 500 µl; SGE analytical sciences, Australia), solid phase extraction cartridges (C18-E, 500 mg, 6 ml), insert vials, amber glass vials, screw caps and septa (Phenomenex, UK), Elga Ultra purification water system (Pure lab, UK), pH indicator strips (2.5-4.5 narrow range; Merck, UK), compressed nitrogen gas (size W; BOC, UK), scissors and glass wool pesticide grade (Sigma Aldrich, UK). LC/ESI-MS/MS was performed on a Waters Alliance 2695 ultra high-performance liquid chromatography system (Acquity, Waters, UK) coupled to an electrospray ionisation triple quadrupole mass spectrometer Xevo TQ-S (Waters, UK).

2.2.3 Internal standards

Deuterated internal standards, available commercially, co-elute with endogenous lipid species at similar chromatographic retention times, but have different masses, as hydrogen isotopes are replaced with deuterium isotopes (2H). This allows for detection of a deuterated species as different to endogenous lipid species, by mass spectrometry. The addition of such deuterated internal standards to each sample prior to lipidomic extraction allows for adjustment of sample loss during the lipid extraction, as a known amount of the internal standard is added to each sample.

2.2.3.1 Deuterated internal standard working solution for Eico assay

This solution was added to the samples at the start of the lipid extraction to account for any loss throughout the extraction protocol and mass spectrometer analysis. 100 µl of each of the four 10 ng/µl deuterated internal standards was added to 600 µl ethanol

to a total volume of 1 ml, to create to a 1 ng/µl working solution. This was kept in an amber vial, sealed with parafilm and stored at -20°C for up to three months.

2.2.3.2 Deuterated internal standards preparation for NAE assay

There were two deuterated internal standards used in this project: deuterated anandamide (AEA-*d8*) and 2-arachidonyl glycerol (2-AG-*d8*). The stock solutions were first diluted to 10 ng/µl for addition to the working internal standard solution and further to 1 ng/µl separately, for use in the calibration line. 10 µl of the AEA-*d8* 500 µg/500µl stock was added to 990 µl ethanol to a total volume of 1 ml to create a 10 ng/µl stock solution. 100 µl of the 2-AG-*d8* 50 µg/500µl stock was added to 900 µl ethanol to a total volume of 1 ml to create a 10 ng/µl stock solution. Stock solutions were sealed with parafilm and stored at -80°C, and the 10ng/µl solutions were kept in amber vials, sealed with parafilm and stored at -20°C for up to three months. 100 µl of each of two 10 ng/µl deuterated internal standards was added to 900 µl ethanol to a total volume of 1 ml to create to a 1 ng/µl working solution each. These working solutions were kept in amber vials, sealed with parafilm and stored at -20°C for up to three months.

2.2.3.3 NAE-CER joint internal standard working solution

An internal standard working solution was created by adding 200 µl of a 25 µM commercial sphingolipid cocktail, 200 µl deuterated anandamide (10ng/µl), 400 µl deuterated 2-arachidonyl glycerol (10ng/µl), and 200 µl ethanol, to a total volume of 1 ml. The solution was kept in an amber vial, sealed with parafilm and stored at -20°C for up to three months. This solution was added to the samples at the start of the lipid extraction to account for any loss throughout the extraction protocol and mass spectrometer analysis. AEA-*d8* was used for all identified NAE species, C25-CER was used for all ceramide species (CER[NS], CER[NDS], and CER[AS]), C17-S was used for sphingoid bases (C18S and C18S1P), and C17-DS was used for the sphinganine base (C18DS).

2.2.4 Lipid extraction protocol

2.2.4.1 Plasma extraction of Eico species

Plasma samples (204 samples; 31 participant families) were profiled for Eico and related mediators. 1 ml of plasma was extracted using 3 ml ice-cold water and 700 µl ice-cold methanol, as published previously (Masoodi *et al.*, 2007, 2008; Massey *et al.*, 2013). 20 µl of the deuterated internal standard working cocktail was added to each sample (1 ng/µl, 18-HETE-*d8* for all species, 11-DHET-*d8* for DHET species). The samples were gently mix and incubated on ice in the dark for 15 min and centrifuged for 10 min at 3,000 rpm at 4°C, to separate the protein precipitate.

The supernatants were collected and a drop of the sample was placed on pH indicator paper to adjust the sample to a pH of 3 using 1 M HCL. This acidification protonates the lipid mediators and enhances the interaction with hydrophobic silica cartridges. The solid phase extraction (SPE) cartridges were washed with 6 ml of 100% methanol followed by 6 ml ultrapure water to condition the SPE cartridges under vacuum. The acidified samples were added to the cartridges using glass Pasteur pipettes and allowed to slowly run through the cartridges. The cartridges were then washed by 6 ml of 15% (v/v) methanol-water, followed by 6 ml ultrapure water, and 6 ml hexane to remove the water and impurities of differing polarity. The lipids were removed from the cartridge with the addition of 6 ml methyl formate and the organic solution was dried under nitrogen and reconstituted in 100 µl ethanol.

The lipid extracts were kept in glass conical inserts in amber vials sealed with parafilm, and stored before mass spectrometry analysis at -20°C for up to one week. Detailed notes were recorded; the number of batches of twelve samples that underwent extraction and any samples with a visually altered appearance (i.e. presence of erythrocytes or leukocytes). A schematic of the full extraction protocol is depicted in Figure 2-1.

2.2.4.2 Plasma extraction of NAE and CER species

A joint assay was used to analyse NAE and CER species. The same participants were analysed as the Eico species (204 samples) for initial assessment by a range finding study (Chapter 4). This was followed by the analysis of a total of 999 samples for the

full cohort analysis (Chapter 5). 1 ml of plasma was extracted using 6 ml 2:1 (v/v) chloroform-methanol solution. 10 µl of the joint NAE-CER working deuterated standard solution (50 pmol of sphingolipid internal standards, 20 ng anandamide-*d8* and 40ng 2-arachidonyl glycerol-*d8* deuterated internal standards) was added to each sample and incubated for 15 min with vortexing every 5 min. 1 ml ultrapure water was added to each sample and the samples were vortexed to mix before centrifugation for 5 min, 2,500 g, at 4°C to separate organic and aqueous phases. The supernatant was filtered through glass wool and dried under nitrogen gas. The resulting lipid extract was reconstituted in 100 µl ethanol.

Detailed notes were recorded; the number of batch of twelve samples extracted for lipids and if the sample had a visually altered appearance (presence of erythrocytes or leukocytes). A schematic of the extraction protocol is depicted in Figure 2-2.

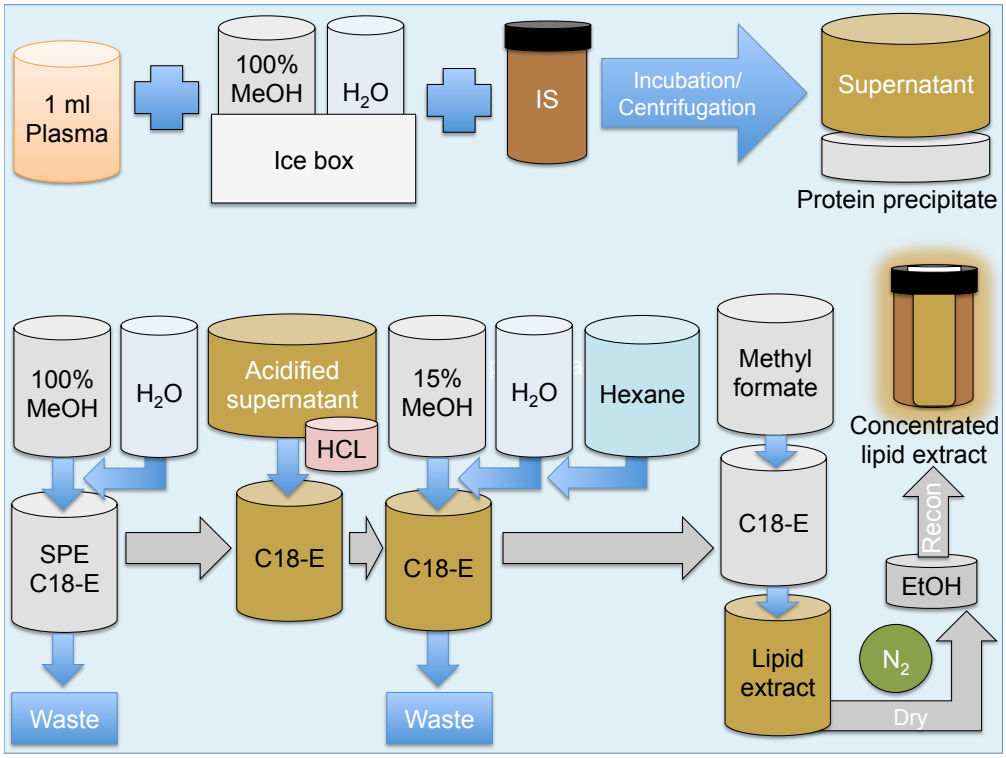


Figure 2-1: Schematic of the plasma extraction protocol for the analysis of Eico.

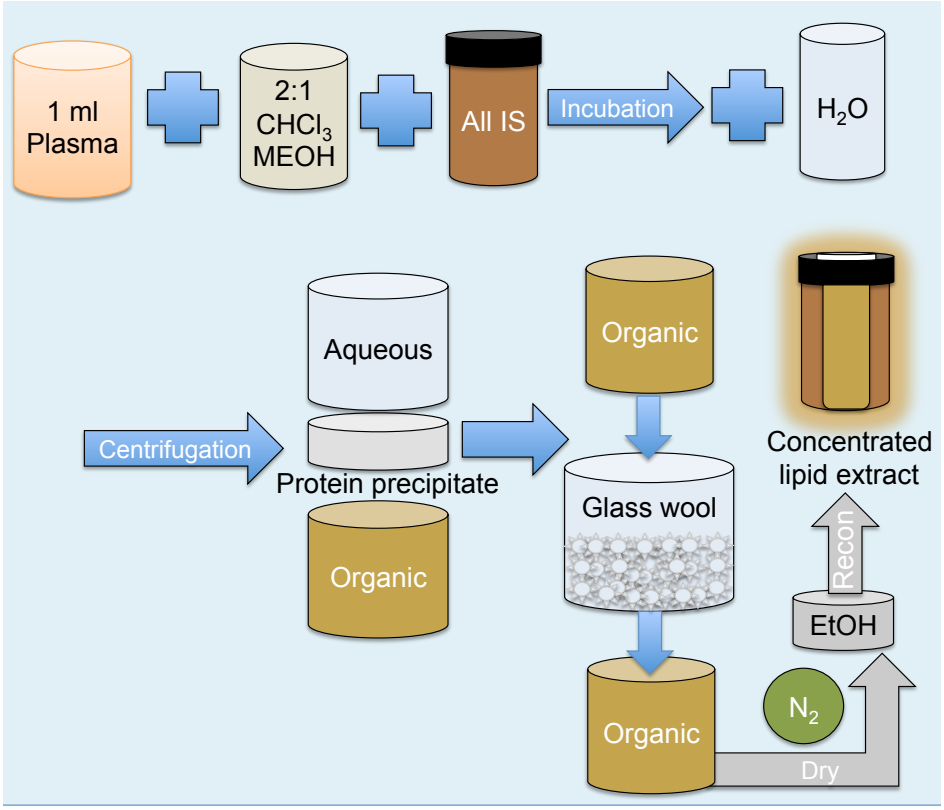


Figure 2-2: Schematic of the extraction protocol for the analysis of NAE and CER species.

2.2.5 Calibration lines

Calibration lines are constructed with commercial standards of each lipid species included in an assay and this is used to link the response of the instrument to known lipid concentrations. The use of such standards and calibration lines allows for the quantification of each lipid species. The responses of lipids in the cohort samples were compared to the known concentrations of the calibration line responses to calculate their concentration. The calibration line is created using the standards in ethanol, and the lipids are quantified in plasma; in the presence of a complex matrix. This may interfere with the signal obtained by the mass spectrometer. However, the use of the same amount of deuterated internal standards in both the calibration line samples and the plasma samples at the same concentration likely adjusts for this issue in a sample-based manner, but not a species-specific way. There is a lack of analyte-free matrix or Certified Reference Materials commercially available, and the removal of lipids from plasma in the creation of such reference materials would likely influence the endogenous plasma matrix environment, potentially adding further variability to a mass spectrometry lipidomics analysis. Creation of deuterated species-specific internal standards, although costly, would aid lipidomics studies, as they could be added to plasma samples without being influenced by endogenous species.

Commercial standards were available for each of the lipid species included in the Eico and NAE assays, but were not available at the start of this study for CER species. Thus, the concentration of plasma Eico and NAE species were quantified and reported in pg/ml, while the relative abundance of plasma CER species is reported in pmol/ml. CER abundance was calculated based on the addition of one internal standard for CER and two for sphingoid bases, due to the lack of commercially available internal standards for the CER class of lipids at the time of the analysis. Therefore, the abundances are presented in rank order, after adjustment of the mass spectrometry response (ion count) for that of the internal standard. This is useful for genetic analyses, which require a relative quantitative measure, but it is not a report of the absolute quantification of the species in plasma. This is similar to other large-scale genetic studies of CER species (Hicks et al., 2009; Demirkan et al., 2012; Tabassum et al., 2019). Relative quantification limits the precision of measurement of each unique lipid species and does not facilitate data integration among different studies

(New Ref - Kito et al 2008). More internal standards for these lipids are becoming available commercially and this will aid the studies of these bioactive lipids.

2.2.5.1 Calibration line preparation for Eico species

Two working solutions were created from stock solutions of all the commercially available standards. Commercial standards of 54 lipid species produced by lipoxygenase (LOX) and CYP450 enzymes (10 µl of each 10 ng/µl stock) were added with ethanol to a total volume of 1 ml and a final concentration of 100 pg/µl. The same was completed for the 24 lipids derived from cyclooxygenase (COX) reactions. The solutions were kept in amber vials, sealed with parafilm and stored at -80°C for up to three months. Two serial dilutions were created from the working solutions (100 pg/µl), covering the range 0.3125-10 pg/µl. 20 µl of the deuterated internal standards (1 ng/µl) were added to each tube. The tubes were dried under nitrogen and reconstituted in 100 µl ethanol. The six solutions for each enzymatic group (12 in total) were kept in amber vials, sealed with parafilm and stored at -20°C for up to one week.

2.2.5.2 Calibration line preparation for NAE species

A calibration line working solution was created by adding 10 µl of each 10 ng/µl stock solutions of 33 NAE standards to 670 µl ethanol, to a total volume of 1 ml and a final concentration of 100 pg/µl. The solution was kept in amber vials, sealed with parafilm and stored at -80°C for up to three months. A serial dilution was set up from the calibration line working solution (100 pg/µl) for quantification of the NAE species. The responses of the endogenous lipids in the cohort samples were compared to the known concentrations of the calibration line responses, to calculate their concentration. Conical inserts were set up in amber vials. 40 µl of a working solution was mixed with 160 µl ethanol. 100 µl of this solution was added into another conical insert with 100 µl ethanol. The serial dilution was continued in this way to cover the range 0.009-20 pg/µl (12 tubes). 20 µl of deuterated anandamide and 40 µl of deuterated 2-arachidonyl glycerol (1 ng/µl solutions) were added to each tube. The tubes were dried under nitrogen and reconstituted with 100 µl ethanol. The twelve solutions were kept in amber vials, sealed with parafilm and stored at -20°C for up to one week.

2.2.6 Mass spectrometry

2.2.6.1 Mass spectrometry parameters for the analysis of cyclooxygenase-derived Eico mediators

The lipids detected from the Eico group that are derived via cyclooxygenase (COX) reactions were separated using a C18 column (Acquity UPLC BEH, 1.7 μm, 2.1 x 50 mm; Waters, UK) and a flow rate of 0.6 ml/min. The mobile phases were as follows; mobile phase A was water:acetic acid (99.98:0.02%)] and mobile phase B was acetonitrile:acetic acid (99.98:0.02%). The solvent gradient is depicted in Table 2.2 and the specific parameters used to analyse the lipids of the Eico class are depicted in Table 2.4.

Table 2.2: Solvent gradients used for the LC-MS/MS analysis of COX-derived Eico species

Mobile phase A was water:acetic acid (99.98:0.02%) and mobile phase B was acetonitrile:acetic acid (99.98:0.02%).

Time (min)	A (%)	B (%)
0	80	20
0.5	80	20
0.6	60	40
2.5	60	40
4	35	65
4.1	80	20
5.8	80	20

2.2.6.2 Mass spectrometry parameters for the analysis of lipoxygenase and CYP450-derived Eico mediators

The lipids detected from the Eico group that are derived via lipoxygenase (LOX) and CYP450 reactions were separated using a C18 columns (Acquity UPLC BEH, 1.7 μm, 2.1 x 50 mm; Waters, UK) and a flow rate of 0.6 ml/min. The mobile phases were as follows; mobile phase A was water:acetic acid (99.98:0.02%) and mobile phase B was acetonitrile:acetic acid (99.98:0.02%). The solvent gradient is depicted

in Table 2.3 and the specific parameters for the analysis of the lipids of the Eico class are depicted in Table 2.4.

Table 2.3: Solvent gradients used for the LC-MS/MS analysis of LOX/CYP-derived Eico species

Mobile phase A was water:acetic acid (99.98:0.02%) and mobile phase B was acetonitrile:acetic acid (99.98:0.02%).

Time (min)	A (%)	B (%)
0	75	25
3	20	80
3.2	75	25
5	75	25

Table 2.4: Further mass spectrometry analysis parameters for Eico species

Compounds are shown that are assessed in the Eico assay. The table includes information on the multiple reaction monitoring transitions assessed (MRM), cone voltage, collision energy and indicative retention times. LOX/CYP dwell time of 0.003s, and COX dwell time of 0.007s.

Lipid	MRM	Cone voltage (V)	Collision energy (eV)	Retention time (min)
PGD1	353>317	12	12	1.33
PGE1	353>317	12	12	1.28
6-keto PGF1α	369>163	12	24	0.99
13,14-dihydro-15-keto PGF1α	355>193	20	30	1.62
PGF1α	355>311	14	24	1.17
13,14-dihydro PGE1	355>337	18	16	1.4
13,14-dihydro-15-keto PGE1	353>335	12	14	1.69
PGB2-d4	337>179	12	20	2.06
PGD2	351>271	24	16	134
PGE2	351>271	24	16	1.25
15-keto PGE2	349>113	14	20	1.4
13,14-dihydro PGF2α	355>311	14	24	1.32
13,14-dihydro-15-keto PGF2α	353>113	10	26	1.69
PGF2α	353>193	12	24	1.18
8-iso PGF2α	353>193	12	24	1.1
PGJ2	333>271	14	16	1.99
Δ12-PGJ2	333>271	14	16	2.03

15-deoxy-Δ12,14-PGJ2	315>271	12	14	3.9
TXB2	369>169	18	18	1.11
13,14-dihydro PGF1α	357>113	4	32	1.37
13,14-dihydro-15-keto PGE2	351>333	12	12	1.57
TXB3	367>169	16	14	1.03
PGD3	349>269	10	16	1.18
PGE3	349>269	10	16	1.13
PGF3α	351>193	2	22	1.08
9-HODE	295>171	16	16	2.47
13-HODE	295>195	2	18	2.49
15-HETrE	321>303	2	14	2.68
5-HETE	319>115	14	14	2.75
8-HETE	319>155	10	14	2.66
± 9-HETE	319>123	16	14	2.71
11-HETE	319>167	14	14	2.61
12-HETE	319>179	20	14	2.66
12-HETE-*d8*	327>184	20	16	2.64
15-HETE	319>175	4	14	2.54
20 HETE	319>245	4	14	2.31
± 5(6)-EET	319>191	4	10	3.03
± 8(9)-EET	319>155	10	14	2.66
± 11(12)-EET	319>167	14	14	2.61
± 14(15)-EET	319>113	4	14	2.54
5-HEPE	317>115	16	12	2.46
± 8-HEPE	317>155	26	12	2.38
± 9-HEPE	317>149	20	14	2.42
± 11-HEPE	317>167	12	12	2.35
12-HEPE	317>179	28	12	2.39
± 15-HEPE	317>175	8	14	2.33
± 18-HEPE	317>215	12	14	2.24
± 5,6-DHET	337>145	8	16	2.33
± 8,9-DHET	337>127	8	16	2.22
± 11,12-DHET	337>167	2	18	2.14
± 14, 15-DHET	337>207	18	16	2.04
5-oxo-ETE	317>203	14	18	2.94
LTB4	335>195	12	14	1.87
RvE1	349>195	14	16	0.81
RvD1	375>141	18	12	1.39
RvD2	375>175	2	22	1.26
± 4-HDHA	343>101	8	12	2.8
± 7-HDHA	343>141	6	14	2.68
± 10-HDHA	343>153	2	16	2.61
± 11-HDHA	343>193.87	2	12	2.65
± 13-HDHA	343>193.15	2	12	2.58

± 14-HDHA	343>161	12	14	2.61
± 17-HDHA	343>201	14	14	2.55
± 20-HDHA	343>241	2	12	2.48
PDX	359>206	18	16	1.81
MaR1	359>177	16	16	1.82
9 OxoODE	293>185	14	18	2.66
13 OxoODE	293>113	16	20	2.58
9 HOTrE	293>171	20	16	2.23
13 HOTrE	293>195	12	16	2.26
9(10) EpOME	295>171	16	16	2.88
12(13)EpOME	295>195	2	18	2.83
Trans EKODE	309>209	16	10	2.3
9,10 DiHOME	313>201	16	20	1.98
12,13 DiHOME	313>183	16	20	1.91
8(9) EET-*d11*	330>155	14	12	2.96
8,9 DHET-*d11*	348>127	16	24	2.21
HXA3	335>273	16	12	2.29
5,15 DiHETE	335>115	12	12	1.82
8,15 DiHETE	335>155	22	16	1.76
16(17) EpDPE	343>233	14	12	2.87
19(20) EpDPE	343>285	18	12	2.79
19,20 DiHDPA	361>273	18	16	2.04

Compounds are shown that are assessed in the Eico assay. The table includes information on the multiple reaction monitoring transitions assessed (MRM), cone voltage, collision energy and indicative retention times.

2.2.6.3 Mass spectrometry parameters for the analysis of NAE species

The lipids detected from the NAE class were separated using a C18 column (Acquity UPLC BEH, 1.7 µm, 2.1 x 50 mm; Waters, UK) and a flow rate of 0.6 ml/min. The mobile phases were as follows; mobile phase A was water:acetic acid (99.98:0.02%) and mobile phase B was acetonitrile:acetic acid (99.98:0.02%). The solvent gradient is described in Table 2.5, and the specific parameters for the mass spectrometry analysis of the lipids of the NAE class are described in Table 2.6.

Table 2.5: Solvent gradients used for the LC-MS/MS analysis of NAE species

Mobile phase A was water:acetic acid (99.98:0.02%) and mobile phase B was acetonitrile:acetic acid (99.98:0.02%).

Time (min)	A (%)	B (%)
0	78	22
3	72	28
3.1	45	55
11	20	80
14.5	20	80
14.51	78	22
17	78	22

Table 2.6: Further mass spectrometry parameters for the analysis of NAE species

Compounds are shown that are assessed in the NAE assay. The table includes information on the multiple reaction monitoring transitions assessed (MRM), cone voltage, collision energy and indicative retention times. Dwell time of 0.003s.

Lipid	MRM	Cone voltage (V)	Collision energy (eV)	Retention time (min)
AEA	348>62	35	15	5.48
DHEA	372>62	35	15	5.38
DPEA	374>62	35	14	5.69
HEA	314>62	35	12	7.15
LEA	324>62	35	15	5.41
OEA	326>62	35	16	6.58
PEA	300>62	35	13	6.15
POEA	298>62	35	14	4.97
PDEA	286>62	35	12	5.32
SEA	328>62	35	15	8.27
VEA	326>62	35	16	6.50
AEA-*d8*	356>63	35	16	5.41
2-AG	379>287	35	18	6.24
MEA	272>62	35	12	4.66
ALEA	322>62	35	14	4.68
DGLEA	350>62	35	14	6.00
EPEA	346>62	35	14	4.74
PGE2-EA	378>360	35	14	2.16
PGD2-EA	378>360	35	14	2.50
PGF2a-EA	380>344	35	16	2.10
15-HETE-EA	364>62	35	12	3.61
5(6)-EET-EA	364>346	35	16	4.29
8(9)-EET-EA	364>346	35	10	4.20
11(12)-EET-EA	364>346	35	12	4.04
14(15)-EET-EA	364>346	35	16	3.79
2-PG	331>239	35	18	7.30
2-STG	359>341	35	18	9.63
2-OG	357>265	35	18	7.71
2-LG	355>263	35	18	6.29

2.2.6.4 Ceramides and related mediators

The lipids detected from the CER class were separated using a C8 column (Acquity UPLC BEH, 1.7 µm, 2.1 x 100 mm; Waters, UK) and a flow rate of 0.3 ml/min. The mobile phases were as follows; mobile phase A was water:formic acid (99.99:0.01%) and methanol:formic acid (99.99:0.01%). The solvent gradient is depicted in Table 2.7, and the specific parameters for the analysis of the lipids of the CER class are depicted in Table 2.8. The cone energy and collision energy for this assay was 30 V and 30 eV, respectively.

Table 2.7: Solvent gradients used for the LC-MS/MS analysis of CER species

Mobile phase A was water:formic acid (99.99:0.01%) and methanol:formic acid (99.99:0.01%).

Time (min)	A (%)	B (%)
0	40	60
6	40	60
9	4	96
20	0	100
30	0	100
32	40	60
40	40	60

Table 2.8: Further mass spectrometry parameters for the analysis of CER species

Compounds are shown that are assessed in the CER assay. The table includes information on the multiple reaction monitoring transitions assessed (MRM) and indicative retention times. Dwell time of 0.2s.

Lipid	MRM	Retention time (min)
CER[A(24)H(16)]	636.593>252	12.07
CER[A(25)H(16)]	650.609>252	12.30
CER[A(24)H(17)]	650.609>266	12.07
CER[A(26)H(16)]	664.624>252	12.54
CER[A(27)H(16)]	678.64>252	12.82
CER[A(25)H(18)]	678.64>280	12.77
CER[A(22)S(18)]	638.609>264	12.06
CER[A(24)S(18)]	666.64>264	12.54
CER[A(26)S(18)]	694.672>264	13.11
CER[N(16)DS(18)]	540.536>284	11.00
CER[N(18)DS(18)]	568.567>284	11.18
CER[N(22)DS(18)]	624.63>284	11.97
CER[N(24)DS(18)]	652.661>284	12.39
CER[N(18)DS(24)]	652.661>368	12.43
CER[N(25)DS(18)]	666.677>284	12.69
CER[N(24)DS(19)]	666.677>298	12.62
CER[N(26)DS(18)]	680.692>284	12.98
CER[N(24)DS(20)]	680.692>312	12.92
CER[N(20)DS(24)]	680.692>368	12.92
CER[N(18)DS(26)]	680.692>396	12.98
CER[N(27)DS(18)]	694.708>284	13.30
CER[N(28)DS(18)]	708.724>284	13.62
CER[N(14)S(18)]	510.489>264	10.40
CER[N(16)S(18)]	538.52>264	10.66
CER[N(18)S(18)]	566.551>264	11.18
CER[N(20)S(18)]	594.583>264	11.47
CER[N(24)S(16)]	622.614>236	11.92
CER[N(22)S(18)]	622.614>264	11.86
CER[N(24)S(17)]	636.63>250	12.10
CER[N(23)S(18)]	636.63>264	12.08
CER[N(22)S(19)]	636.63>278	12.01
CER[N(24)S(18)]	650.645>264	12.32
CER[N(24)S(19)]	664.661>278	12.49
CER[N(23)S(20)]	664.661>292	12.52
CER[N(26)S(18)]	678.677>264	12.84
CER[N(24)S(20)]	678.677>292	12.78

CER[N(27)S(18)]	692.692>264	13.13
CER[N(26)S(19)]	692.692>278	13.05
CER[N(25)S(20)]	692.692>292	13.07
CER[N(28)S(18)]	706.708>264	13.45
CER[N(24)S(22)]	706.708>320	13.33
CER[N(29)S(18)]	720.724>264	13.79
C18S	300>282	8.67
C18DS	302>284	8.82
C18 S1P	380>264	9.21
C18 DS1P	382>266	9.68
CER[N(14)S(18)C1P]	590>264	11.68
CER[N(16)S(18)C1P]	619>264	12.08
CER[N(16)DS(18)C1P]	621>266	12.12
CER[N(18)S(18)C1P]	647>264	12.56

Compounds are shown that are assessed in the CER assay. The table includes information on the multiple reaction monitoring transitions assessed (MRM) and indicative retention times.

2.2.6.5 LC-MS/MS protocol

The correct column for the class of species was connected to the instrument and the respective mobile phases loaded. Columns were fitted with VanGuard pre-column filters (Acquity UPLC BEH, 1.7 μm, 2.1 x 5 mm) (Waters, UK). The wash solutions were created as follows; seal wash, methanol:water 50:50 (v/v); strong needle wash, water:methanol:acetonitrile:isopropanol 1:1:1:1 (v/v/v/v); and weak needle wash, methanol:water 60:40 (v/v). The LC lines were primed with the solvents. Fresh mobile phases and wash solutions were created at the beginning of a mass spectrometry batch. The runs were paused every 48 hours to allow for fresh water mobile phase, which can become contaminated by bacteria.

The mass spectrometer method was set up and the cone temperature was increased to the requirement of the analysis of interest. Once the LC priming was complete and the cone temperature correct, the flow of mobile phases through the instrument in slow increasing increments to 0.6 ml/min for Eico and NAE, and 0.3 ml/min for CER (due to the longer column for CER species). The flow was allowed to settle until the instrument's pressure decreased to a delta of below 30, where fluctuations in pressure over the last minute in time are below ±30 psi.

Further class-based mass spectrometry information is as follows: CER were analysed using a source temperature 100 °C, a desolvation temperature 450 °C, a capillary voltage of 3.5 kV, and a dwell time of 0.2 s. NAE were analysed using a source temperature of 150 °C, desolvation temperature of 400 °C, capillary voltage of 1.8 kV, and dwell time of 0.003 s. Eico were analysed using a source temperature of 150 °C, a desolvation temperature of 500 °C for the COX assay and 600 °C for LOX/CYP assay, a capillary voltage of 3.1 kV for COX assay and 1.5 kV for the LOX/CYP assay, and a dwell time 0.007 s for the COX assay and 0.003 for the LOX/CYP assay. Compounds were fragmented using argon gas and monitored in the positive ion mode (NAE and CER) or negative ion mode (Eico) by multiple reaction monitoring (MRM). The lipids were separated at a temperature of 25 °C for NAE and Eico species, and 30 °C for CER. The sample chamber was set to 8 °C. Injection volume was 3 µl.

Once the sample sheet was set up with the list of samples and the samples were loaded into the instrument, a blank ethanol injection was analysed and assessed for any issues. Three injections of blank ethanol samples were then analysed before any sample lipid extracts to ensure the instrument was clear of any issues. The cohort samples' lipid extracts were analysed in duplicate with a blank injection separating samples to ensure the injector and columns were rid of any crossover contaminants (Table 2.8.1).

Familial samples were extracted and analysed by mass spectrometry in the same batch, where possible. The analyses have been previously published (Eico (Astarita *et al.*, 2015), NAE (Urquhart *et al.*, 2015) and CER (Kendall *et al.*, 2017)). Detailed notes were recorded for every analysis; the batches the samples were analysed in, when the mobile phases were changed, the series the tubes were analysed in, and when there was annual commercial performance maintenance on the mass spectrometer.

Table 2.8.1: Example sample list

Blank ethanol injections were run between calibration line sample and plasma samples. All samples were run in duplicate (a,b) to a maximum of 48 plasma samples, including a pooled plasma quality control sample. (=) denotes a continuation in the sample list as previously described until all samples were analysed.

Blank ethanol injection 1
Blank ethanol injection 2
Blank ethanol injection 3
Calibration Line injection 1a
Calibration Line injection 1b
Blank ethanol injection 4
Calibration Line injection 2a
Calibration Line injection 2b
Blank ethanol injection 5
=== Calibration Line ends ===
Blank ethanol injection X
Plasma Sample injection 1a
Plasma Sample injection 1b
Blank ethanol injection X+1
Plasma Sample injection 2a
Plasma Sample injection 2b
Blank ethanol injection X+2
= Plasma Sample List ends =
Blank ethanol injection N
Mass Spectrometry Shutdown Solutions

2.2.6.6 Detection and quantification

Target Lynx software (version 4.1, Waters, UK) was used to process the mass spectrometry data and normalise the identified peak areas against the deuterated internal standards. The peaks obtained for each injection of each sample were manually assessed. Peaks were included in analysis if they had an area 10 times greater than that of the background (signal/noise ration > 10), an area greater than a value of 100, and had the same retention time as the commercially available calibration line species, where available. The area of the peak selected was divided by the area of the deuterated internal standard peak, to create a response value. The response value for each lipid measured in the cohort samples were assessed by linear regression from the equation of the line created using the calibration line of

commercial standards, to quantify the lipids as pg/µl of injected extract. The mean of the duplicate injections were calculated for accuracy and normalised against the volume of liquid used, which was 1 ml of plasma. All duplicate injections were included in the analysis. All sample injections were less than 14% variable from the mean for the Eico analysis (assessed via the area of 12-HETE-*d8*). All sample injections were less than 28% variable from the mean for the CER analysis (assessed via the area of CER-C25). 6% of NAE samples were found to have injections that were over 30% from the mean across batches (assessed via the area of AEA-*d8*). The genetic results of the removal of samples with injections that varied more than 30% for the NAE species were the same, likely due to the estimate of genetic analyses for measurement (non-genetic) errors and the removal of outliers in the statistical analyses. The only difference was that the GWAS association for PEA seen in FAAH, dipped below GWAS significance of $5x10^{-8}$, due to the lower sample size available for analysis.

The quantification of NAE and Eico species are presented in pg/ml. The relative semi-quantitation for each CER species is presented in pmol/ml.

2.2.7 Creation of pooled plasma quality control samples

Plasma samples from healthy volunteers of a previous study recruited in 2008 were pooled in a glass beaker, aliquoted equally (0.9 ml), and stored at -80°C. 16 samples of such pooled quality control (QC) plasma were extracted and analysed blindly alongside the full cohort NAE and CER analysis, and 4 were extracted and analysed blindly alongside the range finding study for the Eico species analysis. These samples were used to produce equal quality control samples that were extracted and analysed by mass spectrometry alongside each batch of cohort samples, and used for statistical adjustment of batch and processing effects in the statistical analyses.

The same QC sample was analysed by three injections in the analysis of injection variability described in 2.2.8.1, 2.2.8.2, 3.2.2, where pooled samples were analysed in triplicate to assess the injection variability for each lipid species. While the QC samples only underwent two injections during each batch and therefore cannot be analysed by coefficient of variation, the analysis of injection variability showed low

variability in sample injections, and the analysis of the percentage difference for each cohort sample injection presented in Chapter 3: 3.2.4.

2.2.8 Quality assessment of detectable plasma lipids

Where a mass spectrometry peak was identified for a lipid species in the cohort plasma samples, the species were assessed by the following quality control (QC) measures. The results are described in Chapter 3.

2.2.8.1 Quality assessment of the detected plasma NAE and Eico species

Three solutions (QC 1-3), described below, were analysed for the NAE and Eico lipid classes that had commercial standards available for analysis. Each sample was analysed in triplicate with a blank ethanol injection run in between the samples. The samples were analysed to test the following quality control criteria (Matuszewski *et al.*, 2003); plasma matrix effects (Equation 2.1), lipid extraction recovery (Equation 2.2), and process efficiency (Equation 2.3).

The solutions were as follows;

QC 1. A solution of the commercially available lipid standards (standards) in ethanol at a known concentration together with deuterated internal standards (IS): standards + IS.

QC 2. A pooled QC plasma sample with the same concentration of lipid standards as included in QC 1, added to the sample before lipid extraction: plasma + IS + standards added before extraction.

QC 3. A pooled QC plasma sample with the same concentration of lipid standards as in QC 1 added to the sample after lipidomic extraction: plasma + IS + standards added after extraction.

An estimation of sample carry over was calculated by the presence of lipid species in the blank injections; mass spectrometry-produced peaks for each lipid were assessed for a peak at the same retention time in the blank ethanol injections, and the area of such peak found in the blank injections was divided by the area of the corresponding in the plasma samples.

Mass spectrometry injection variation was analysed by calculation of the mean response and standard deviation of the triplicate injections for each of the three QC samples, assessed for each lipid species. The coefficient of variation (Equation 2.4) was calculated for lipid from each of the triplicate injections of the three samples. The mean of the coefficient of variations resulting from the three QC samples are described in Chapter 3. This robust assessment of injection variation took into account any variation in the analysis of each lipid species via commercial standards in ethanol, without matrix effects (QC 1), in a plasma matrix with extraction losses (similar setting to the cohort samples) (QC 2), and in a plasma matrix without extraction losses (QC 3).

$$\text{Matrix Effect: } \frac{Plasma + IS + Standards\ after\ extraction}{Standards + IS} = \frac{QC\ 3}{QC\ 1}$$

Equation 2.1: Calculation of matrix effect

$$\text{Extraction Recovery: } \frac{Plasma + IS + Standards\ before\ extraction}{Plasma + IS + Standards\ after\ extraction} = \frac{QC\ 2}{QC\ 3}$$

Equation 2.2: Calculation of recovery

$$\text{Process Efficiency: } \frac{Plasma + IS + Standards\ before\ extraction}{Standards + IS} = \frac{QC\ 2}{QC\ 1}$$

Equation 2.3: Calculation of process efficiency

$$Coefficient\ of\ variation = \frac{Standard\ deviation}{Mean}$$

Equation 2.4: Calculation of the coefficient of variation

2.2.8.2 Quality assessment of the detected plasma CER species

As the CER species do not have commercial standards available for all of the species analysed in this study, peaks that were detected in the cohort plasma samples in the correct transition by selected reaction monitoring (SRM) and at the indicative retention time for each CER species, were assessed via multiple reaction monitoring

(MRM) based on structure-specific fragments. These multiple transitions were assessed to confirm the identification of the correct peak for each CER species and were used for peak referencing. Multiple transitions were identified from literature for each species (Sullards, 2000; Bielawski *et al.*, 2006; Boath *et al.*, 2008; Masukawa *et al.*, 2008, 2009; Shaner *et al.*, 2009; Van Smeden *et al.*, 2011; T'Kindt *et al.*, 2012; Mercado *et al.*, 2014), or calculated based on the class-based pattern of fragmentation (Table 2.9). A pooled sample of many lipid extracts from the cohort was used for this analysis to create a concentrated lipid sample, allowing for the identification of a peak of a minor fragment. The results of this analysis are described in Chapter 3. Nearing the end of the project, structure-specific deuterated internal standards became available for each CER class (one for each of CER[NS], CER[NDS], and CER[AS]) and these confirmed the identified peaks were correct.

Sample carry over was also assessed as was completed for the NAE and Eico species, by the presence of lipid species in a blank ethanol sample. Injection variation was assessed by the analysis of three pooled QC plasma samples with internal standards added, mirroring the cohort plasma extraction protocol, and the mean response and standard deviation for each lipid of the three samples analysed by triplicate injection was used to calculate the coefficient of variation (Equation 2.4). The mean coefficient of variation of all three QC samples is presented in Chapter 3.

Table 2.9: MRM transitions used to confirm the presence of the CER class

Each CER species detected was confirmed via multiple reaction monitoring (MRM) from identified transitions in literature or calculated via class-based fragment patterns. NA, no further transitions described in literature.

Ceramide	Precursor	Fragment 1	Fragment 2	Fragment 3
CER[A(22)S(18)]	638.609	252.4	264.4	282.4
CER[A(24)S(18)]	666.640	252.4	264.4	282.4
CER[A(26)S(18)]	694.672	252.4	264.4	282.4
CER[N(16)S(18)]	538.520	252.4	264.4	282.4
CER[N(20)S(18)]	594.583	252.4	264.4	282.4
CER[N(22)DS(18)]	624.630	266.4	284.4	302.4
CER[N(22)S(18)]	662.614	252.4	264.4	282.4
CER[N(22)S(19)]	636.630	266.4	278.4	296.4
CER[N(23)S(18)]	636.630	252.4	264.4	282.4
CER[N(23)S(20)]	664.661	310.4	292.4	280.4
CER[N(24)DS(18)]	652.661	266.4	284.4	302.4
CER[N(24)DS(19)]	666.677	280.4	289.4	316.4
CER[N(24)DS(20)]	680.692	294.4	312.4	330.4
CER[N(24)S(16)]	622.614	224.4	236.4	254.4
CER[N(24)S(17)]	636.630	238.4	250.4	268.4
CER[N(24)S(18)]	650.645	252.4	264.4	282.4
CER[N(24)S(19)]	664.661	266.4	278.4	296.4
CER[N(24)S(20)]	678.677	310.4	292.4	280.4
CER[N(24)S(22)]	706.708	338.4	320.4	308.4
CER[N(25)DS(18)]	666.677	266.4	284.4	302.4
CER[N(25)S(20)]	692.692	310.4	292.4	280.4
CER[N(26)DS(18)]	680.692	266.4	284.4	302.4
CER[N(26)S(18)]	678.677	252.4	264.4	282.4
CER[N(26)S(19)]	692.692	266.4	278.4	296.4
CER[N(27)S(18)]	692.692	252.4	264.4	282.4
CER[N(28)S(18)]	706.608	252.4	264.4	282.4
CER[N(29)S(18)]	720.724	252.4	264.4	282.4
C18S	300.000	252.4	264.4	282.4
C18DS	302.000	254.4	266.4	284.4
C18S1P	380.000	264.4	82.4	NA

2.3 Statistical analyses

2.3.1 Covariate adjustment

Systematic error was considered from several places along the lipidomics experimental pipeline; the number of lipidomic extraction batches, the number of mass spectrometry batches, the presence of erythrocytes or leukocytes in the plasma samples, the number of mobile phase changes during mass spectrometry, the number of annual performance maintenance that were undertaken for the mass spectrometer during the cohort analysis (increasing the sensitivity of the mass spectrometer), and the series the samples were analysed in. Integer traits were created for each potential predictor.

The trait for the number of mass spectrometry batches (e.g. MSBatch) correlated strongly with extraction batch, mobile phase changes, performance maintenance assessments, and the series the samples were run in ($R>0.81$, Table 2.10 and Table 2.11). Therefore, only the trait for mass spectrometry batches and a trait for sample abnormality (the presence of erythrocytes or leukocytes), were included in covariate analyses to avoid collinearity and over-adjustment. Such samples that were haemolysed were recorded and a binary trait to highlight such samples was included in the stepwise linear regression analysis as a potential predictor in the statistical analysis. The trait for haemolysis was not found to be a significant predictor influencing the measurements of the lipid classes studied (Appendix Table 0.4).

The lipid-specific abundances of the pooled QC plasma samples analysed with every mass spectrometry batch (e.g. 13-HODEQC) were added to the list of potential covariates for each lipid species. This trait added further information to that of the MSBatch trait as there was a unique value for each lipid studied in each batch, taking into account lipid species-specific mass spectrometry differences, such as ionisation, matrix effects of plasma on a specific lipid, and any variation in peak processing for the specific lipids over the study progression. There was not a strong relationship (i.e. $R>0.80$) with MS Batch and the pooled QC plasma values for every species (Appendix Table 0.1), therefore both MS Batch and a lipid-specific pooled QC plasma value were included in the assessments of covariance. Stepwise multiple linear regression allowed for the additive assessment of the influence of predictors on

a lipid species measurement, therefore assessing MSBatch and the values from the pooled plasma samples independently. Of all covariates, batch effects were identified as the most reoccurring predictors, identified for most of the lipid species (Appendix), in particular the trait providing lipid specific values from the pooled plasma samples was the most included trait in the model. The binary trait describing 1-24 batches (MSBatch) did not have a strong correlation with the lipid species (i.e. 0.6<R<-0.6) (Table 2.9.1).

However, the trait of lipid-specific values from the pooled quality control samples analysed alongside each batch of cohort samples, was the most significant predictor for all classes of lipids, with the relationship depending on lipid species (Table 2.9.2). The relationship increased with this trait from analysing the range-finding study samples that were analysed within four batches (n=204) compared to the full cohort analysis of twenty further batches (n=1016; Table 2.9.2). Variability is expected from undertaking the analysis over this large number of batches, which was required to analyse this moderate number of samples, and this highlights the importance of the statistical adjustment of such batch effects in the discovery of the genetic influences underlying the measurement of these lipid species.

During the time taken to analyse the samples, annual preventative maintenance (PM) was undertaken on the liquid chromatography and mass spectrometry instrumentation. This includes changing parts of the instrument, cleaning the quadruples, and ensuring the system is has a high standard of sensitivity, in order to prevent any future issues. These occurred during this study before batches 7 and 22. An increase in sensitivity (and therefore relative abundance) is clearly depicted for CER[N(26)S(18)], the most abundant CER found in plasma (highlighted in green, panel E, Figure 2.2.1). The repetitive use of the standardised pooled plasma analysed alongside each batch has allowed for the adjustment of such systematic errors for the genetic analysis. However, reporting the relative abundance in plasma, which is required by many lipidomic journals, requires the adjustments for batch effects in a large lipidomics cohort analysed over multiple batches, such as in this project. This association of lipid relative abundances with batch effects highlights the importance of adjustments for batch effects in bioanalysis, where future use of high-throughput robotics and the

allocation of multiple instruments will be desirable as bioactive lipidomics expands to analyse larger cohorts and biobanks.

Table 2.9.1: Correlation of lipid species concentration with batch effect

Correlation coefficient (R) is depicted in the column described "Batch" for n=204 samples for all three class of lipids (left), and n=1016 for NAE and CER (right).

Lipid	Batch (n=204)
A22_S18	-0.21
A24_S18	-0.19
A26_S18	-0.43
C18_DS	0.23
C18_S	-0.43
C18_S1P	-0.40
N16_S18	-0.17
N20_S18	-0.37
N22_DS18	-0.30
N22_S18	-0.34
N22_S19	-0.29
N23_S18	-0.15
N23_S20	-0.11
N24_DS18	-0.33
N24_DS19	-0.27
N24_DS20	-0.19
N24_S16	-0.42
N24_S17	-0.17
N24_S18	-0.23
N24_S19	-0.22
N24_S20	-0.12
N24_S22	-0.01
N25_DS18	-0.18
N25_S20	-0.15
N26_DS18	-0.26
N26_S18	-0.24
N26_S19	-0.19
N27_S18	-0.19
N28_S18	-0.25
N29_S18	-0.04

Lipid	Batch (n=1016)
A22_S18	-0.46
A24_S18	-0.15
A26_S18	0.12
C18_DS	0.35
C18_S	0.12
C18_S1P	0.11
N16_S18	0.04
N20_S18	-0.30
N22_DS18	-0.43
N22_S18	-0.35
N22_S19	-0.44
N23_S18	-0.51
N23_S20	-0.17
N24_DS18	-0.33
N24_DS19	-0.19
N24_DS20	-0.04
N24_S16	-0.46
N24_S17	-0.44
N24_S18	-0.58
N24_S19	-0.34
N24_S20	0.01
N24_S22	0.13
N25_DS18	-0.07
N25_S20	0.14
N26_DS18	0.04
N26_S18	0.00
N26_S19	0.01
N27_S18	0.11
N28_S18	0.09
N29_S18	0.00

AEA	-0.20
DHEA	-0.08
DPEA	-0.07
HEA	0.10
LEA	-0.05
OEA	0.18
PEA	0.17
POEA	0.05
PDEA	0.28
STEA	0.12
VEA	0.06
DHET1112	-0.26
HETE11	-0.09
EpOME1213	0.05
DiHOME1213	-0.05
HETE12	-0.11
HODE13	0.02
HOTrE13	-0.04
OxoODE13	-0.04
DHET1415	-0.10
HETE15	-0.10
DiHDPA1920	-0.02
HDHA4	0.03
HETE5	-0.01
EpOME910	0.02
DiHOME910	-0.33
HODE9	0.03
HOTrE9	0.24
OxoODE9	-0.20
Trans_EKODE	-0.11

AEA	0.12
DHEA	0.28
DPEA	0.33
HEA	0.45
LEA	0.16
OEA	-0.19
PEA	0.22
POEA	0.36
PDEA	0.43
STEA	-0.18
VEA	-0.23

Table 2.9.1: Correlation of lipid species concentration with batch effect.

Correlation coefficient (R) is depicted in the column described “Batch” for n=204 samples for all three class of lipids (left), and n=1016 for NAE and CER (right).

Table 2.9.2: Correlation of lipid species concentration with quality control sample values from each batch

Correlation coefficient (R) is shown in the columns for n=204 samples for all three class of lipids, and n=1016 for NAE and CER. Cells highlighted in red are substantially correlated (-0.6<R>0.6), while cells in white showed a more modest relationship with the quality control sample values.

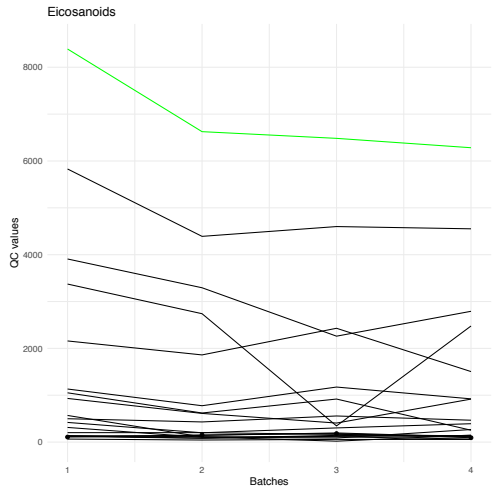
Lipid	R (n=204)	R (n=1016)
A22_S18	0.13	0.57
A24_S18	0.38	0.64
A26_S18	0.40	0.33
C18_DS	0.24	0.82
C18_S	0.68	0.66
C18_S1P	0.65	0.60
N16_S18	0.14	0.51
N20_S18	0.31	0.61
N22_DS18	-0.08	0.43
N22_S18	0.09	0.57
N22_S19	0.24	0.47
N23_S18	0.01	0.55
N23_S20	-0.06	0.15
N24_DS18	0.12	0.48
N24_DS19	0.35	0.33
N24_DS20	0.03	0.13
N24_S16	0.40	0.57
N24_S17	-0.01	0.46
N24_S18	0.10	0.64
N24_S19	0.21	0.35
N24_S20	0.04	0.17
N24_S22	0.08	0.26
N25_DS18	0.10	0.30
N25_S20	0.10	0.30
N26_DS18	0.06	0.13
N26_S18	0.36	0.31
N26_S19	0.20	0.23
N27_S18	0.19	0.30
N28_S18	0.26	0.38
N29_S18	0.11	0.28
AEA	0.03	-0.10
DHEA	-0.06	0.57
DPEA	-0.09	0.70
HEA	0.16	0.72
LEA	0.22	0.64

OEA	0.31	0.62
PEA	0.14	0.68
POEA	0.00	0.63
PDEA	0.52	0.43
STEA	0.26	0.33
VEA	0.20	0.67
DHET1112	-0.10	
HETE11	0.13	
EpOME1213	0.25	
DiHOME1213	-0.08	
HETE12	0.03	
HODE13	0.09	
HOTrE13	0.07	
OxoODE13	-0.05	
DHET1415	-0.05	
HETE15	0.11	
DiHDPA1920	-0.15	
HDHA4	-0.18	
HETE5	0.33	
EpOME910	0.05	
DiHOME910	-0.33	
HODE9	0.40	
HOTrE9	0.06	
OxoODE9	-0.03	
TransEKODE	0.39	

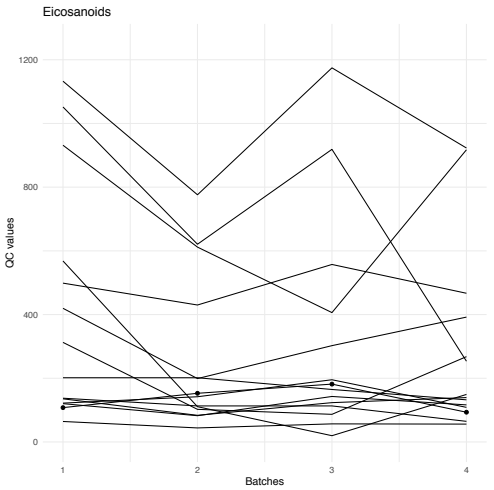
Table 2.9.2: Correlation of lipid species concentration with quality control sample values from each batch

Correlation coefficient (R) is shown in the columns for n=204 samples for all three class of lipids, and n=1016 for NAE and CER. Cells highlighted in red are substantially correlated (-0.6<R>0.6), while cells in white showed a more modest relationship with the quality control sample values.

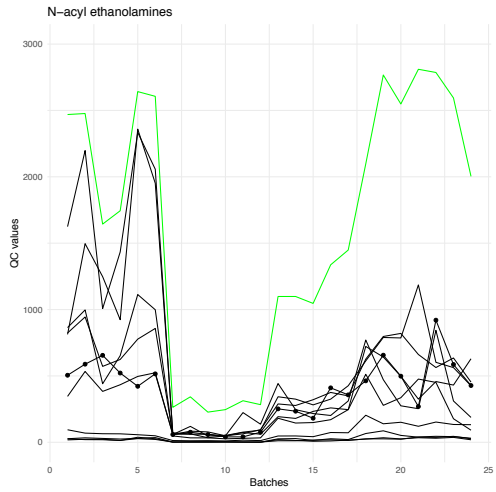
A


Eicosanoids
QC values
Batches

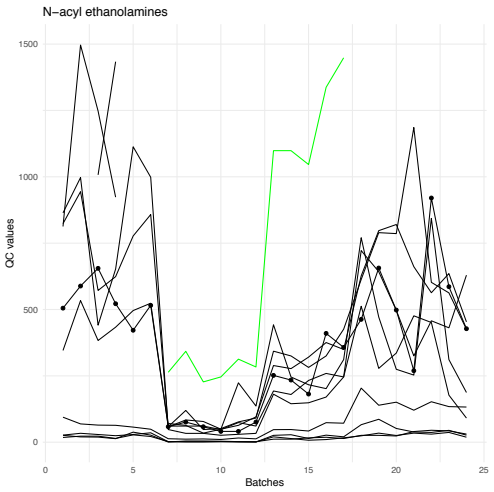
B


Eicosanoids
QC values
Batches

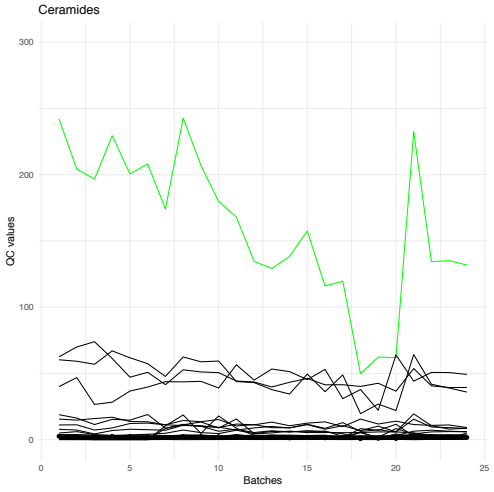
C


N−acyl ethanolamines
QC values
Batches

D


N−acyl ethanolamines
QC values
Batches

E


Ceramides
QC values
Batches

F

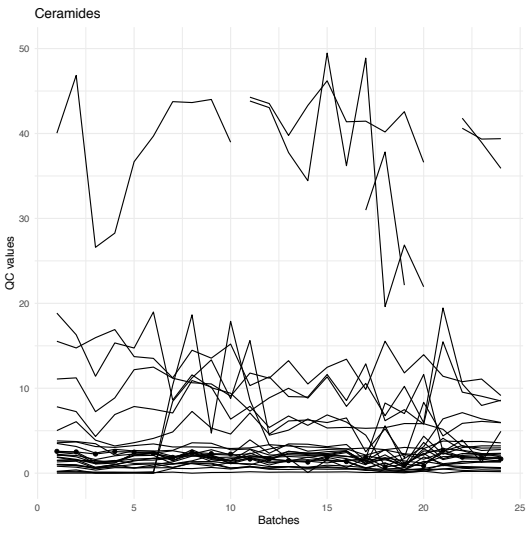

Ceramides
QC values
Batches

Figure 2.2.1: Plots of pooled plasma lipid concentration over all batches

Plots of the levels obtained for the pooled plasma samples for each of the three classes of lipid are shown (A, C, E) complimented by a "zoomed in" version of the plot using a shorter Y-axis also depicted (B, D, F). Values found higher than the Y-axis in the zoomed plots (B, D, F) are depicted as broken lines. The Eico species (A, B) were analysed over 4 batches. The NAE (C, D) and CER (E, F) species were analysed over 24 mass spectrometry batches. During the time taken to analyse the samples, annual preventative maintenance (PM) were undertaken at two stages, before batches 7 and 22, which depicts an increase in mass spectrometry sensitivity for CER[N(26)S(18)], the most abundant CER found in plasma (highlighted in green, panel E). The most abundant NAE species, PEA, is highlighted in green in panels C and D. The most abundant Eico species, 13-HODE is highlighted in green in panels A and B.

Cholesterol measures and BMI were added to the model. While cholesterol is produced from a different lipid biosynthetic pathway to the lipids studies here, the measure represents a total lipid insult and is regularly adjusted for in CER literature to identify associations with CER species beyond cholesterol (Demirkan *et al.*, 2012; Laaksonen *et al.*, 2016). BMI was used to adjust for a potential dietary insult influencing the lipid levels. Ascertainment selection status was assessed by a binary hypertension trait. Sex, age, and age^2 (the non-linear effect of age) were included in the model to adjust for differences in age and gender on the lipid levels. The final model contained the variables MS Batch, sample abnormality, age, age^2, sex, hypertension status, BMI, cholesterol, and pooled QC sample measures.

The set of potential predictors were assessed by stepwise multiple linear regression analysis to identify the predictors with greatest influence over the lipid levels. This allows for the assessment of the influence a trait has over the lipid levels and adds other traits in a stepwise manner, to find the best set of influencing predictors. MS Batch and QC samples were therefore assessed additively.

The lipid measurements were assessed for effect of potential covariates using stepwise multiple linear regression to identify the best set of predictors, using the caret package and 'leapSeq' method in R (version 3.5.2). Initially this was completed in SPSS software, but due to the numerous lipid traits being analysed, a high throughput analysis was required, therefore this was completed using a more rapid software, that of standard R programming language approaches, which can complete the analysis in a more automated manner.

Multiple linear regression analysis of the best predictors was undertaken using the 'lm' function in R. The best set of predictors are presented in Chapter 4 for the range finding study, and Chapter 5 for NAE and CER analysis in the full cohort. Residuals from the covariate-adjusted regression models were standardized to have a mean of 0 and a variance of 1. Outliers were assessed using the R package 'car' for a Bonferroni p-value of $P<0.05$, created for each observation by testing them as a mean-shift outlier, based on studentized residuals, to remove the most extreme observations. The impact of outlier removal is explored in Chapter 3. Missing values were coded as missing such in the genetics analyses.

Table 2.10: Correlation between potential covariates to assess for collinearity.

The recorded experimental traits were assessed for collinearity by assessment of correlation (Pearson correlation with two-tailed P-values, Graph Pad Prism 7 software); the value under the trait name is the correlation coefficient (R), assessing the strength of the relationship, and the P-value is depicted, assessing the significance of the test. Extraction batch, the number of batches of 12 plasma samples that underwent lipidomic extraction; MS batch, the number of mass spectrometry batches required to complete the total number of samples; Sample abnormality, the presence of erythrocytes or leukocytes; Solvent change (CER), the number of changes of the aqueous mobile phase during a CER mass spectrometry run; Solvent change (NAE), the number of changes of the aqueous mobile phase during a NAE mass spectrometry run; Sample series, the order the samples were run in; PM calibration, the number of annual performance maintenance calibrations on the mass spectrometer, potentially increasing the sensitivity of the instrument. A Graph Pad Prism P-value of <0.0001 is depicted here as 0.00.

	Extraction Batch	**P-value**	**MS Batch**	**P-value**	**Sample Abnormality**	**P-value**
Extraction Batch	X	X	1.00	0.00	-0.10	0.00
MS Batch	1.00	0.00	X	X	-0.10	0.00
Sample Abnormality	-0.10	0.00	-0.10	0.00	X	X
Solvent Change (CER)	1.00	0.00	1.00	0.00	-0.11	0.00
Solvent Change (NAE)	1.00	0.00	1.00	0.00	-0.10	0.00
Sample Series	1.00	0.00	1.00	0.00	-0.10	0.00
PM Calibration	0.85	0.00	0.86	0.00	-0.08	0.01

Table 2.11: Correlation between potential covariates to assess for collinearity.

The table is as depicted in Table 2.10, but for the 200 plasma samples analysed for Eico. There was no variation in mobile phase or performance maintenance during the analyses of these samples due to the speed of the assay.

	Extraction Batch	**P-value**	**MS Batch**	**P-value**	**Sample Abnormality**	**P-value**
Extraction Batch	X	X	0.97	0.00	0.04	0.57
MS Batch	0.97	0.00	X	X	0.01	0.85
Sample abnormality	0.04	0.57	0.01	0.85	X	X
Sample series	1.00	0.00	0.97	0.00	0.09	0.57

2.3.2 Genome-wide genotyping quality control

At the commencement of the project, the genotyping data was obtained as a set of binary PLINK genotyping files (.bed, .bim, .fam) containing information on 1,234 individuals (580 male, 654 female, 0 unspecified sex), including 248 founders and 986 non-founders for 557,124 SNPs measured to 0.993318 genotyping rate. The following quality control thresholds were undertaken using PLINK version 1.90 (Purcell *et al.*, 2007) with graphing in R software, using the University of Manchester's high performance computing cluster environment. The quality control steps were completed as described by Marees *et al.* (2018) in line with thresholds completed on a previously published cohort using the same genotyping chip (Cordell *et al.*, 2013), with adaptations for family relatedness (Zaitlen *et al.*, 2013). Quality control assessment is required as genotyping data errors can arise from poor quality DNA, poor hybridization to the array, poor genotyping probes, sample mix-ups, and contamination (Marees *et al.*, 2018).

2.3.2.1 Heterozygous haploid genotypes

As the sex chromosomes were included at the initial stage of the analysis, and are later removed from the GWAS analyses, heterozygous haploid genotypes existed (n=12,753) and these were set to missing (--set-hh-missing). The standard GWAS exclusion of sex chromosomes from analysis is due to the lack of power in a sample size of 580 males to analyse the Y-chromosome, and the distorted allele frequency by inclusion of an X- or the Y-chromosome. Sex chromosome variants are comparatively lower to the number of variants assessed on autosomes by standard genotyping arrays. Moreover, the effect of X-chromosome variants identified by GWAS is unknown, due to the phenomenon of random X-inactivation in women (Nature, 2017). Future whole genome sequencing efforts will aid genetic analyses of sex chromosomes.

2.3.2.2 Mendelian inconsistencies

Mendelian inconsistencies are genotypes that are not seemingly inherited from parents (e.g. a mother with the genotype AA at a particular locus, a father with the same AA genotype, and an offspring with a different Aa genotype, depicted in Figure 2 3). DNA sequencing has estimated that the germ line de novo mutation rate in

humans is between $1.0x10^{-8}$ - 1.8×10^{-8} per nucleotide per generation, with substantial variation among families, of which 44 to 82 de novo single-nucleotide mutations are found in the genome of the average individual (Acuna-Hidalgo et al., 2016). As there is ~3.2 billion nucleotides per human genome, the frequency of a de novo mutation is 0.00000003. This cohort assesses 557,124 SNPs, so about 0.014 per individual would be detected, or 18 such SNPs in the entire cohort.

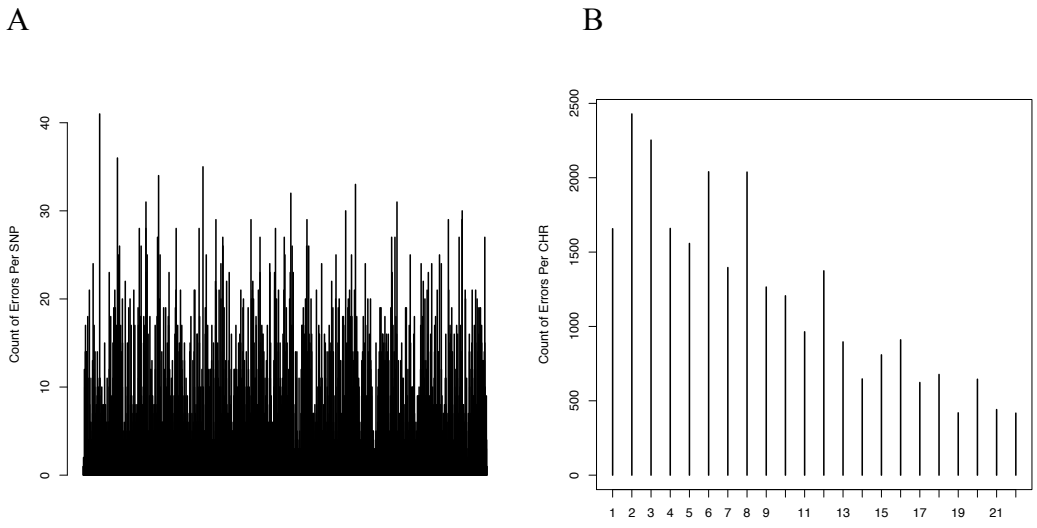
The presence of variants with Mendelian inconsistencies was assessed (--mendel-multigen) and 26,323 errors were identified over the 1,234 individuals (~21 per individual) from 105 families and 9,683 SNPs (Figure 2 3). As this was inflated compared to the predicted presence of de novo SNPs in this population (i.e. 18 SNPs), the variants identified are likely genotyping errors (Turner et al., 2011). Exome or whole genome sequencing studies are required to assess rare and de novo variants as genotyping arrays are used to study common DNA variants only.

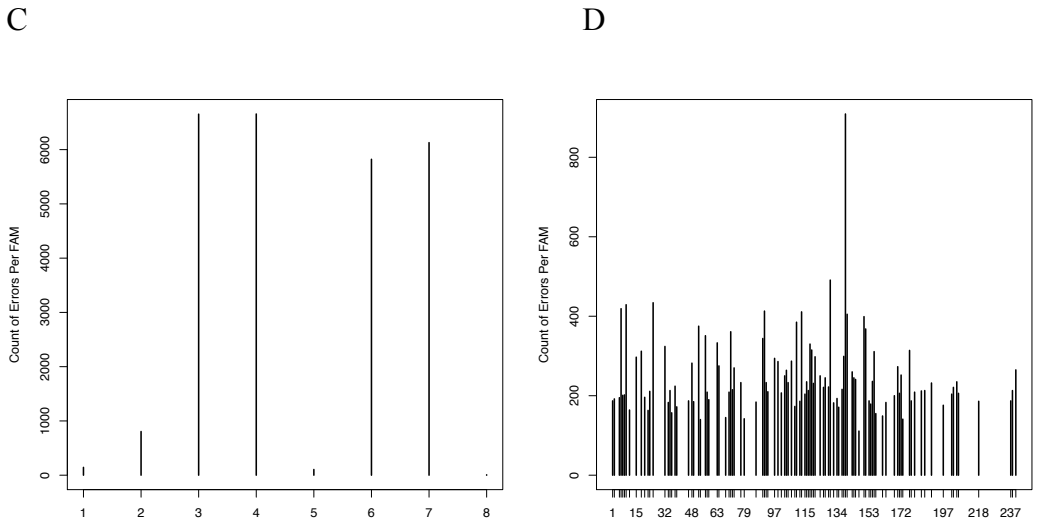
The resulting error codes identified by the analysis showed that the most prominent errors were due to the presence of missing values, where other error codes include alterations between homozygous parents and their offspring. To avoid false positive results, all inconsistencies were set to missing (--set-me-missing), in the aim of identifying only SNPs that were inherited in the cohort and confidently genotyped. This did not alter the number of SNPs available for analysis, as the inconsistencies were spread out over SNPs and individuals.

2.3.2.3 Missingness

Missingness was assessed for individuals and SNPs using the command --missing, and results are depicted in Figure 2-4. SNPs were removed that are missing in a large proportion of the participants, which can lead to biases, and individuals were removed if they had high rates of genotyping missingness, which can be due to poor DNA quality or technical issues (Marees *et al.*, 2018). SNPs and individuals with more than 5% missing values were removed (--geno 0.05 and --mind 0.05), first removing missing SNPs before individuals (Figure 2-4). 538,771 SNPs and 1,230 people (579

males, 651 females/247 founders, 983 non-founders) remained (3% SNPs were excluded and 4 individuals were excluded).





E

Code	Pat. genotype	Mat. genotype	Child genotype	Samples implicated
1	11	11	12	all
2	22	22	12	all
3	22	11/12/missing	11	father, child
4	11/12/missing	22	11	mother, child
5	22	22	11	child
6	11	12/22/missing	22	father, child
7	12/22/missing	11	22	mother, child
8	11	11	22	child
9	(Xchr male)	11	22	mother, child
10	(Xchr male)	22	11	mother, child

Figure 2-3: Assessment of Mendelian inconsistencies.

The figure depicts the count of Mendelian errors A) by SNP, B) by chromosome, C) by error code, and D) by family ID. Panel E describes the error codes identified in panel C, where Mat., maternal and Pat., paternal (taken from the PLINK1.9 website, cog-genomics.org/).

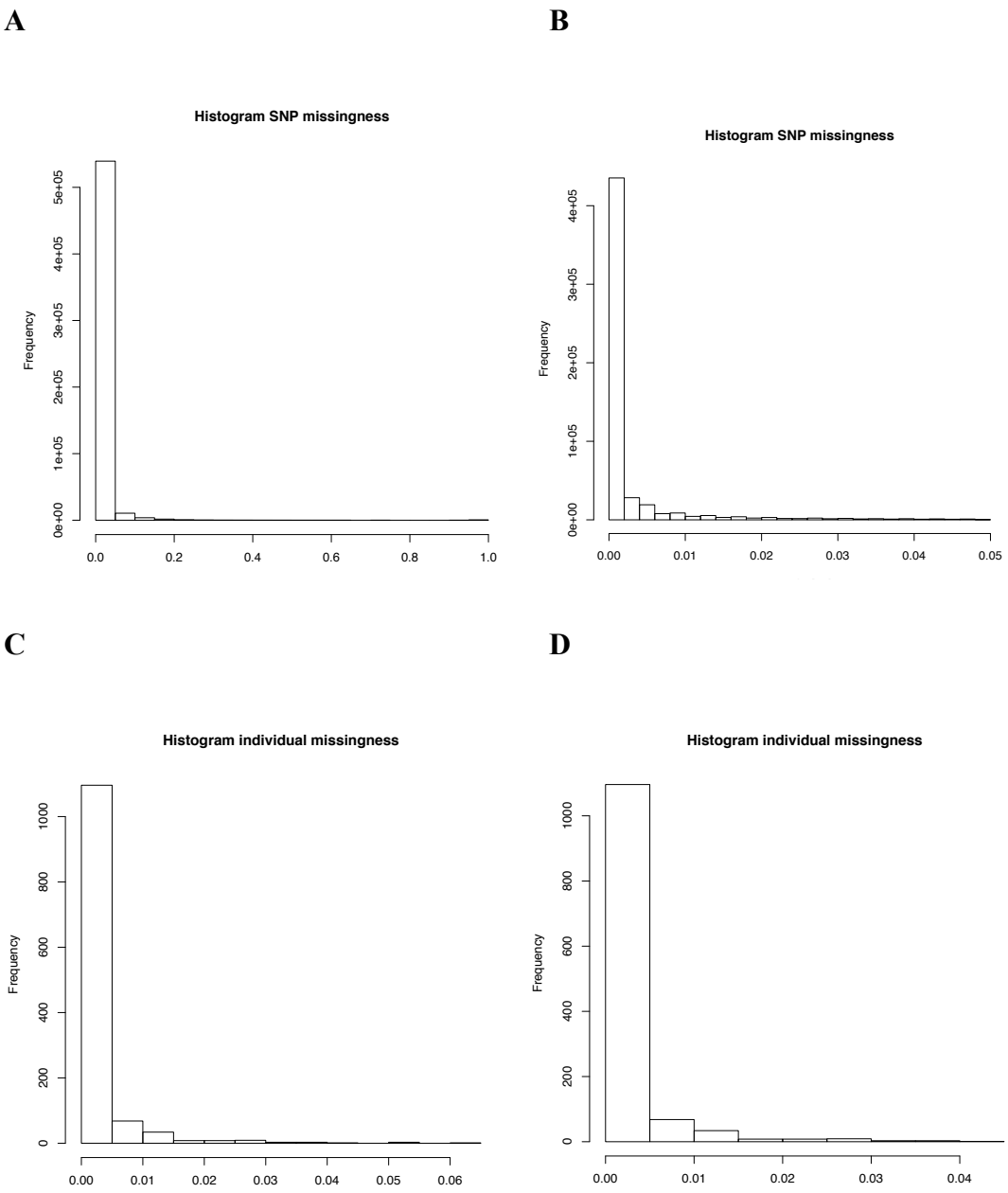


Figure 2-4: The effect of quality control on SNP- and individual-based missingness.

The figure depicts the frequency of missingness (0.00 - 1.00). The threshold of less than 5% missingness was used for genotype calls on SNPs (A and B) and individual-based missingness (C and D). Panels A and C depict the raw data before genotyping quality control and panels B and D depict the data post genotyping control.

2.3.2.4 Gender assessments

Assigned gender was assessed (--check-sex) by checking for discrepancies between the recorded gender for each participant and their gender based on F value produced from the X chromosome inbreeding (homozygosity) estimate. As all females had a F value <0.2 (due to the presence of two X-chromosomes) and all males had an F value >0.8 (due to presence of only one X-chromosome), no changes were required (Figure 2-5). This step can indicate the presence of sample mix-ups (Marees *et al.*, 2018).

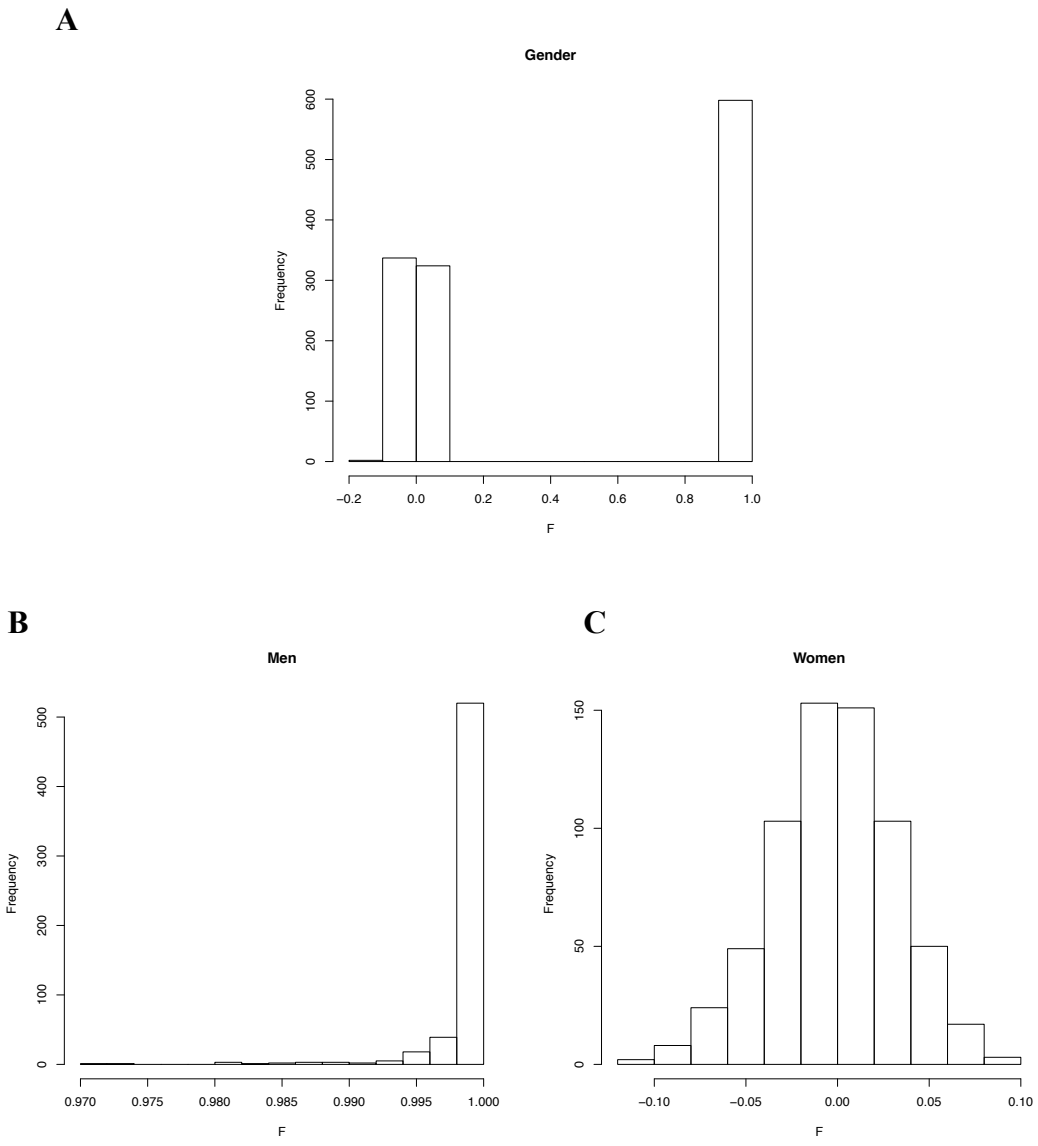


Figure 2-5: Assessment of gender.

Gender was assessed based on the F value produced from the X chromosome inbreeding (homozygosity) estimate. As in panels A and C, females had a F value <0.2 as expected (low homozygosity, two X-chromosomes) and males had an F value >0.8 (high homozygosity, only one X-chromosome), depicted in panels A and B. No changes were required.

2.3.2.5 Minor Allele Frequency

To assess the distribution of minor allele frequency (MAF), one X-chromosome was retained and a list of SNPs from chromosome 1-23 (chromosomes 1-22 plus one X-chromosome) were extracted from the data (--extract). 538,754 SNPs were available for analyses. Rare SNPs lack the power to detect GWAS associations (Marees *et al.*, 2018) and this study was not created to identify rare SNPs and thus would be underpowered in doing so. The --maf command filters out all SNPs with a minor allele frequency below the provided threshold. After applying the --maf 0.01 filter to only include SNPs above a 1% MAF as in Cordell *et al.* (2013), rare SNPs which were not genotyped consistently were removed, and 514,772 SNPs remained for analysis (Figure 2-6).

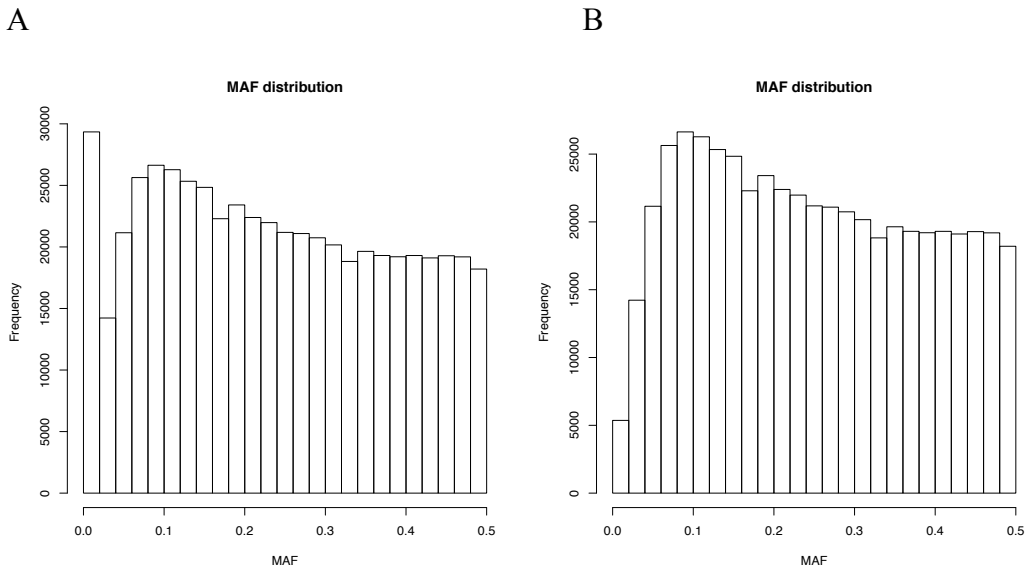


Figure 2-6: Assessment of allele frequency.

The minor allele frequency in panel A depicts that of the raw data. Any SNPs that were not found in at least 1% of the cohort were removed (panel B).

2.3.2.6 Duplicate SNPs

The genotyping data was assessed for duplicate SNPs using the --list-duplicate-vars command. No duplicated SNPs were identified in the files.

2.3.2.7 Hardy-Weinberg equilibrium

SNPs were removed that deviated from Hardy-Weinberg equilibrium (HWE). This step assessed founders only (--hardy). Under Hardy-Weinberg assumptions, allele and genotype frequencies can be estimated, and are constant from one generation to the next, assuming that the fictitious indefinitely large population doesn't undergo selection, mutation, or migration. Departure from this equilibrium and thus violation of the HWE law by significantly different observed genotype frequencies than expected, can be indicative of potential genotyping errors, population stratification, or potentially evolutionary selection. Typically, HWE deviations toward an excess of heterozygotes reflect a technical problem in the assay, such as non-specific amplification of the target region (Marees *et al.*, 2018). A threshold of $1x10^{-8}$ (--hwe 1e-8 include-nonctrl, Figure 2-7) was applied as in a similar study using the same genotyping chip (Cordell *et al.*, 2013) and all SNPs passed.

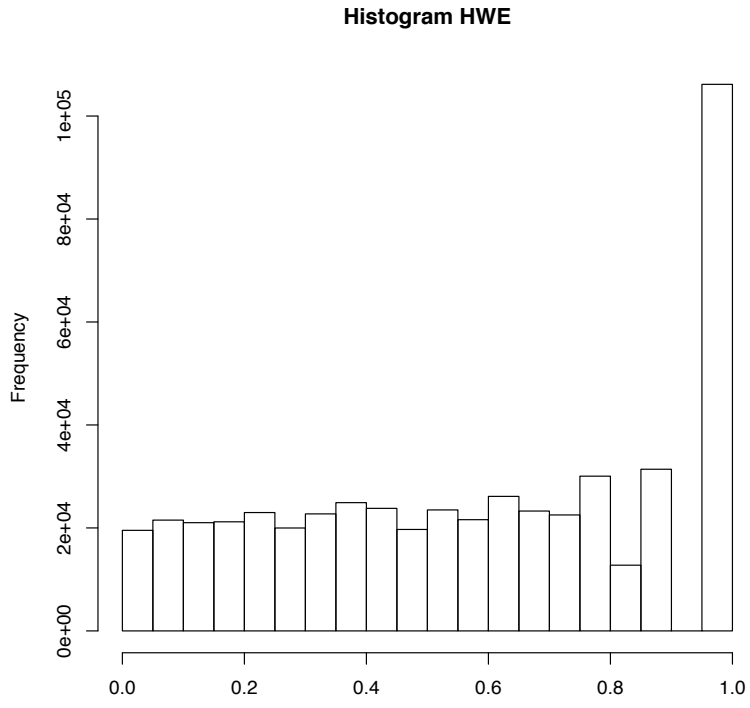


Figure 2-7: Assessment of Hardy Weinberg Equilibrium.

The figure shows the minor allele frequency of the SNPs (Y-axis) and the level of HWE (X-axis).

2.3.2.8 Heterozygosity

Heterozygosity was assessed to remove extreme individuals with a heterozygosity rate deviating more than four standard deviations from the mean. Low heterozygosity, excessive homozygosity, can be due to population bottlenecks or stratification, metapopulation dynamics with a reduced level of genetic variation relative to that expected or found in comparable humans, inbreeding, assortative mating, or sample contamination. High heterozygosity, lots of genetic variability, can be due to potentially mixing of two isolated populations, or low sample quality (Marees *et al.*, 2018). To assess for this, SNPs were pruned to select for those in similar linkage equilibrium; a set of SNPs were excluded (--exclude) that are not correlated (SNPs in high linkage disequilibrium regions; high inversion regions) provided by Marees *et al.* (2018) in areas of inversion (Table 2.12).

Table 2.12: Areas of inversion in the genome excluded from the analysis of heterozygosity

Chr, chromosome; Start Position, starting position of the locus in base pairs; End Position, ending position of the locus in base pairs; Name, name for the region of high inversion.

Chr	Start Position	End Position	Name
6	25500000	33500000	8 HLA
8	8135000	12000000	Inversion8
17	40900000	45000000	Inversion17

SNPs were assessed using the parameters 50 for window size, 5 for the number of SNPs to shift the window at each step, and 0.2 for multiple correlation coefficient of a SNP being regressed on all other SNPs simultaneously (--range --indep-pairwise 50 5 0.2). Individuals that deviated outside more than four standard deviations were removed (Figure 2-8). The heterozygosity rate, the proportion of heterozygous genotypes of an individual, was calculated as in Equation 2.5.

$$Heterozygosity\ rate = \frac{N(NM) - O(HOM)}{O(HOM)}$$

Equation 2.5: Calculation of heterozygosity rate

N(NM) is the number of total observations and O(HOM) is the count of observed homozygosity.

For each individual, O(HOM) ranged from 65,152 - 68,907 (mean of 68,224) and N(NM) ranged from 96,239 - 100,926 (mean of 100,682). The heterozygosity rate was 0.2998 - 0.3372, and after correction for four standard deviations it was 0.3135 - 0.3322. 1,221 participants (574 males, 647 females) remained (9 individuals were excluded). PLINK1.9 provided an estimate of method-of-moments F coefficient as in Equation 2.6.

$$F\ coefficient = \frac{O(HOM) - E(HOM)}{N(NM) - E(HOM)}$$

Equation 2.6: Calculation of F coefficient

E(HOM) is the expected homozygosity loaded from --read-freq or via imputed minor allele frequencies in small sample sets.

Both heterozygosity rate and F coefficient are similar and can be used to assess heterozygosity. In this study the heterozygosity rate was used to estimate heterozygosity as advised by Marees *et al.* (2018). The F coefficient had a strong, inverse relationship with the heterozygosity rate, as expected from the calculations (R = -0.99, Figure 2-9).

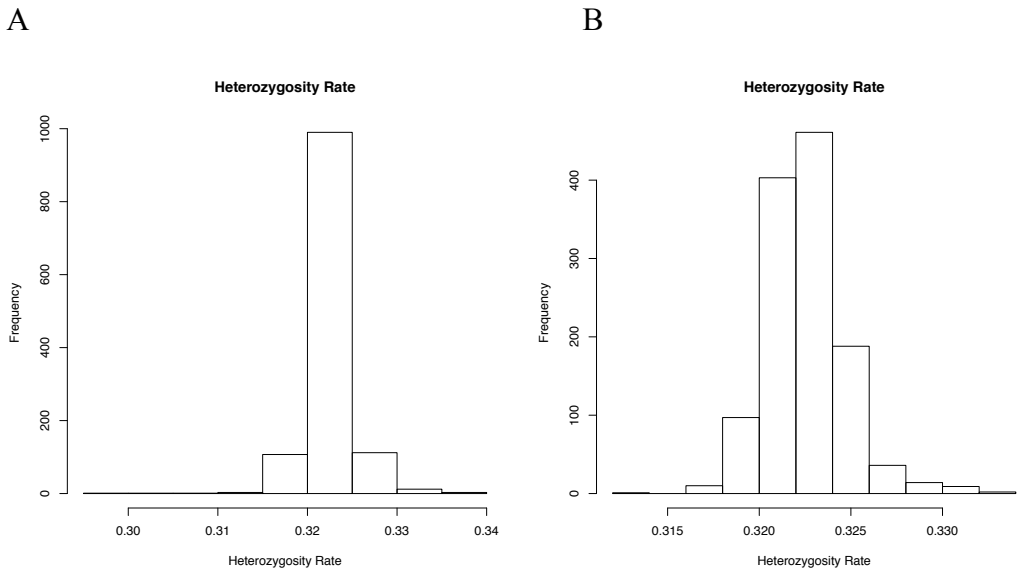


Figure 2-8: Assessments of heterozygosity.

Outliers of heterozygosity were removed; individuals who were outside four standard deviations of the mean. Panel A depicts the raw results, with panel B depicting the post quality control results.

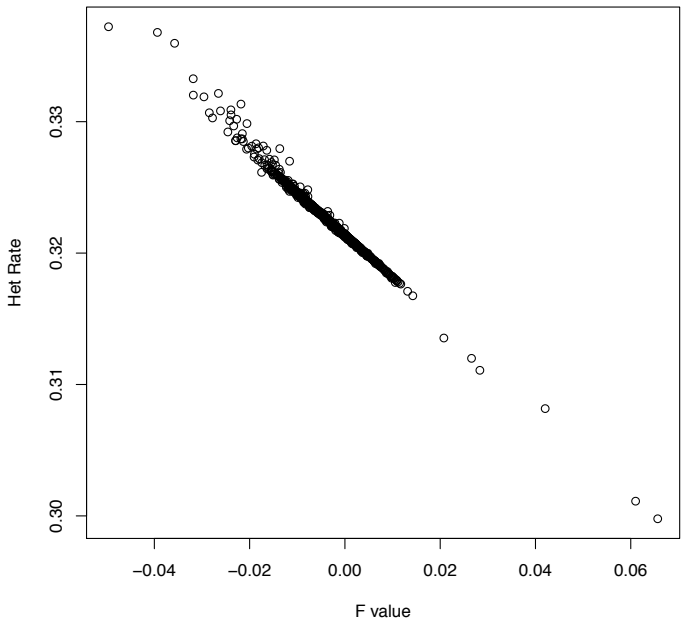


Figure 2-9: Assessment of the relationship between F coefficient and heterozygosity rate calculations, showing an inverse relationship (R=-0.99).

2.3.2.9 Relatedness

During the curation of the cohort, DNA samples that showed an altered pedigree structure to that of the reported family structure were removed from analyses. This family-based cohort was assessed for genetic relatedness in several ways, depicted in Figure 2-10. While this assessment was not used to exclude relatedness as in case/control studies, a proof-of-principal approach was undertaken to confirm the reported relatedness. The command --rel-check was used to identify overall relatedness in the cohort. The cohort was assessed for overall cohort relatedness (--rel-check) and a pi-hat>0.2, which is commonly used to identify cryptic relatedness (second degree relatives) (Marees *et al.*, 2018). As expected, relatedness was identified but this initial assessment was mostly uninformative in this family-based analysis.

The command --genome was used to calculate identity by descent (IBD) of all sample pairs using the same set of SNPs as Section 2.3.2.8 for heterozygosity, i.e. excluding high inversion regions (Figure 2-10). Pairs of individuals were then plotted by their degree of relatedness by the proportion of loci shared by the pair at one allele IBD (Z1) and the proportion of loci shared by the pair at no alleles (Z0). Two half sibling and one full sibling relationships were found in the intermediate space between full and half siblings. Their relationship was assessed (Tables 2.13 and 2.14). Individual identified as 221-4 was implicated twice in these outlier relationships and was therefore removed. Individual identified as 63-8 had more missing values than individual identified as 63-9, and so the former was removed. Before and after adjustment plots are depicted in Figure 2-10. While the two individuals could have been left in the analyses as they were not extreme outliers, it was preferred to remove them to ensure the quality of the results achieved in this study. At this point in the quality control pipeline, 514,772 variants remained (chromosomes 1-23; 503,221 variants for chromosomes 1-22) for 1,219 individuals (243 founders and 976 non-founders) from 216 families.

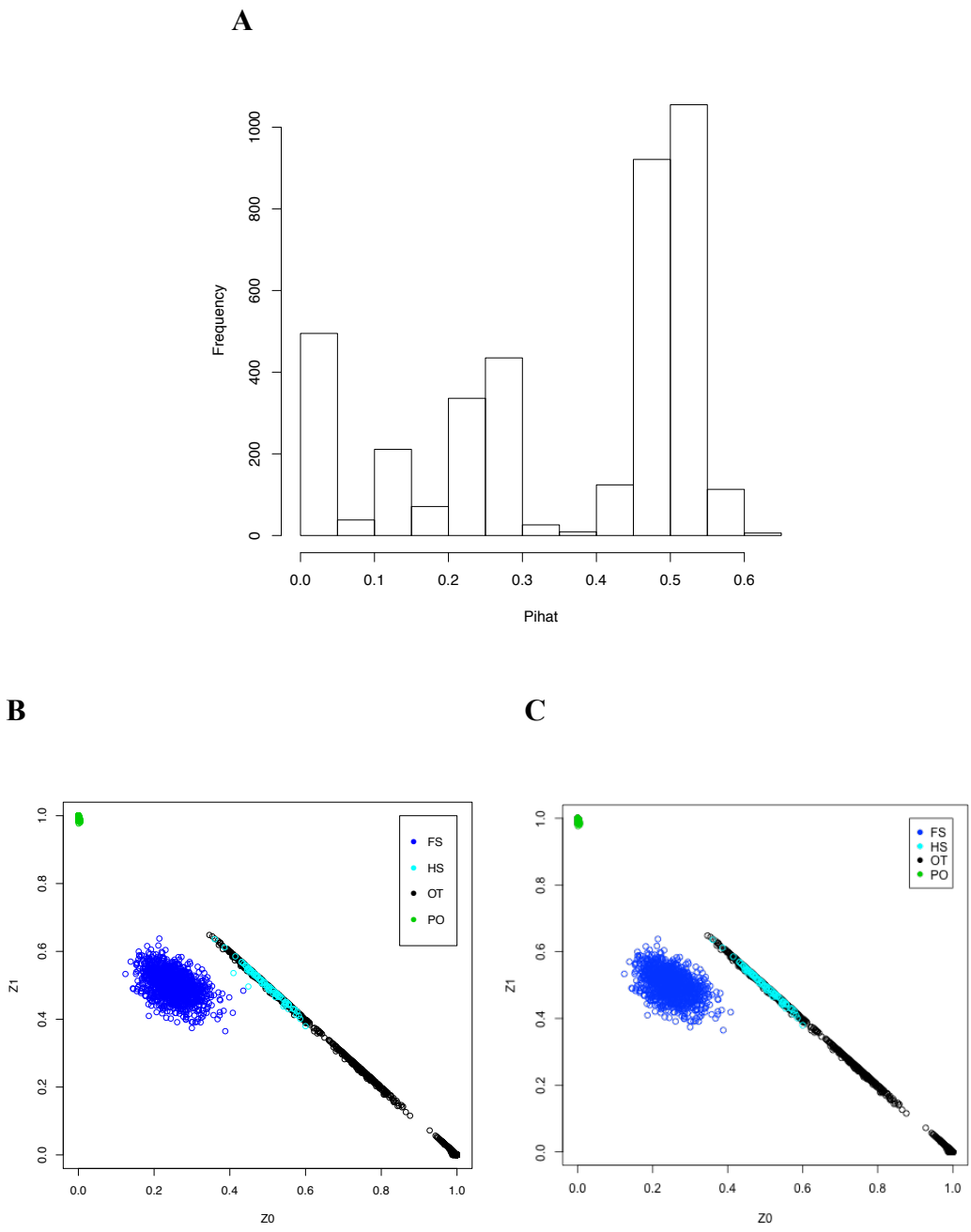


Figure 2-10: Assessments of family-based relatedness.

A) A plot of the frequency of individuals at different relationships in the cohort. --rel-check identified overall relatedness in the cohort. 0.5 is high levels of IBD sharing (first degree relatives), with 0.2 describing lower levels (second degree relatives). B) A plot of pairs of individuals plotted by their degree of relatedness by Z1, the proportion of loci shared by the pair at one allele IBD, and Z0, the proportion of loci shared by the pair at no alleles. The colour code is determined by their pedigree information (.fam file). Two half sibling relationships and one full sibling relationship was found in the intermediate space between full and half siblings (panel B). These were removed, depicted in panel C. FS, full siblings; HS, half siblings; OT, other; PO, parent-offspring.

Table 2.13: Assessment of the relationships found in the intermediate space between half siblings and full siblings.

The same individual 221-4 was identified in two outlier relationships (bold). FID1, family ID for the first singleton of the pair; IID1, individual ID for the first singleton of the pair; FID2, family ID for the second singleton of the pair; IID2, individual ID for the second singleton of the pair; RT, relationship type inferred from the pedigree information (.fam file), where FS is full sibling and HS is half sibling; EZ, IBD sharing expected value based on the inferred relationship; Z0, P(IBD=0); Z1, P(IBD=1); Z2, P(IBD=2); PI_HAT, Proportion of IBD (P(IBD=2) + 0.5*P(IBD=1)); DST, IBS distance ((IBS2 + 0.5*IBS1) / (IBS0 + IBS1 + IBS2)); PPC, IBS binomial test.

FID1	IID1	FID2	IID2	RT	EZ	Z0	Z1	Z2	PI_HAT	DST	PPC
63	9	63	8	FS	0.5	0.4285	0.4859	0.0856	0.3285	0.812	1
221	**4**	221	3	HS	0.25	0.4173	0.5244	0.0583	0.3205	0.808	1
221	6	**221**	**4**	HS	0.25	0.445	0.5046	0.0504	0.3027	0.804	1

Table 2.14: Assessment of the individuals identified from Table 2.13.

Individual 63-8 was removed for having more missing values in comparison to individual 63-9. FID, family ID; IID, individual ID; N_MISS, the number of missing SNPs for that individual; N_GENO, the number of total genotyped SNPs for the cohort at this stage of the quality control pipeline; F_MISS, frequency of missingness for that individual.

FID	IID	N_MISS	N_GENO	F_MISS
221	6	106	514772	0.0002059
221	4	103	514772	0.0002001
221	3	4172	514772	0.008105
63	9	2751	514772	0.005344
63	8	6055	514772	0.01176

2.3.2.10 Population stratification

The population was assessed via principal components analysis with the 1,000 Genome Project (Auton *et al.*, 2015), which confirmed all participants were of homogenous European ancestry (Figure 2-11). Allele frequencies can differ between populations and lead to false positive or false negative results (Marees *et al.*, 2018). The SNPs that overlapped between this cohort and the 1,000 genomes project were assessed in this analysis; using --extract to extract the cohort SNPs from the 1,000 Genomes data, and further extract the corresponding SNPs from the cohort data; --recode to .map and .ped format; --update-map to change the 1,000 Genomes data to the genome build of the cohort data; --reference-allele to update the reference allele of the cohort data to the new build of the 1,000 Genomes data; --flip to resolve any strand issues; --exclude to remove any further uncorresponding SNPs; --bmerge to merge the binary cohort and 1,000 Genomes files; --extract to prune the analysis for only autosomal chromosomes; --genome to calculate regions of IBD of all sample pairs; --read-genome, --cluster, and --mds-plot to create a multidimensional scaling (MDS) plot, performed on an inter-sample distance matrix.

462,853 variants were included in the analysis of 1,848 individuals made up of 629 1,000 Genomes participants and 1,219 from the cohort data. The analysis calculates the genome-wide average proportion of alleles shared between pairs of individuals within the sample to generate quantitative indices/components of the genetic variation for each individual. It produces a k-dimensional representation (10 dimensions) of any substructure in the data based on IBS, to identify individuals that are more genetically similar to each other than expected.

The 1,000 Genomes data included populations of known ethnic structure, of European (EUR), Asian (ASN), mixed American (AMR), and African (AFR) ancestry, where the cohort genomes (OWN) were anchored using the other genomes, and shown to all gather around those of the European ancestry group as was expected (Figure 2-11). While the top 10 principal components were calculated from the multidimensional scaling report, the family and relatedness-specific heritability and GWAS software used in this study take into account population structure and thus no population stratification was required. Following the full quality control pipeline, 503,221

autosomal SNPs (514,772 SNPs from chromosomes 1-23) from 1,219 individuals (216 families) were available for SNP-based heritability and imputation analyses.

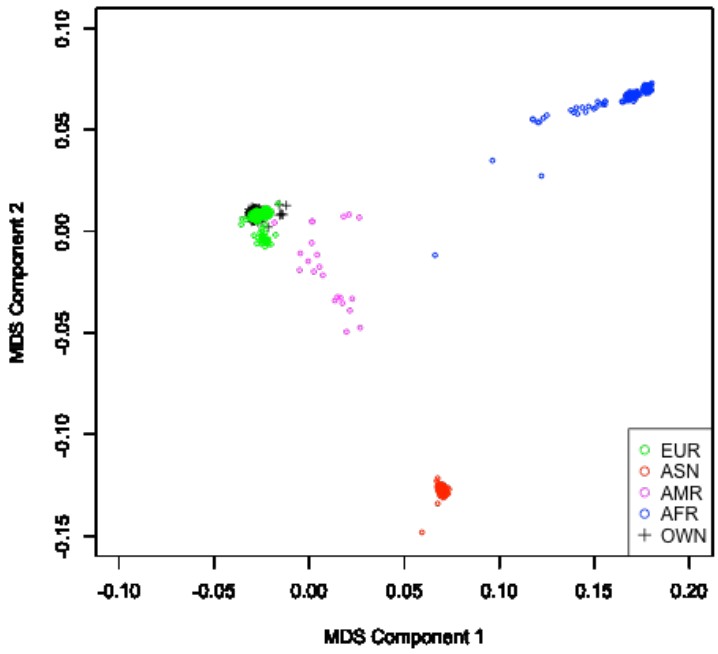


Figure 2-11: Assessment of population stratification with the 1,000 Genomes Project.

The figure depicts genomes of European (EUR), Asian (ASN), mixed American (AMR), and African (AFR) ancestry, where the cohort genomes studied here (OWN) were all gathered around those of the European ancestry group, as expected.

2.3.3 Heritability estimates

Heritability was estimated in this project using variance components analysis with statistical software QTDT (Quantitative Transmission Disequilibrium Tests; version 2.6.1). The software analyses the levels of quantitative traits (i.e. plasma lipid concentrations) from each family member in the reported pedigrees of extended families, and assesses the similarities between individuals, non-shared environment (unique to the individual and measurement error), polygenic (relatedness and polygenes), additive major gene effect (one allele causes the effect of a phenotype, having two doubles this effect), and common environment (shared between families), which are used to produce a variance-covariance matrix (Abecasis *et al.*, 2000).

QTDT compares two models to estimate the additive genetic variance; an environmental only model and a model of environmental and polygenic variances (-we and -veg commands). This evaluates the significance of individual components of variance specific to alternative variance models, and an estimated size of the polygenic effect. The software produces parameter estimates of the polygenetic variance (Vg) and the environmental variance (Ve) of a trait. Heritability estimates were then calculated using Equation 2.7. QTDT has been used abundantly in literature to estimate the narrow-sense heritability of quantitative traits (see Keavney *et al.*, 1998; Peeters *et al.*, 2005; Rhodes *et al.*, 2008; Goldinger *et al.*, 2013).

$$h^2 = \frac{Vg}{(Vg + Ve)}$$

Equation 2.7: Calculation of heritability

Vg is variance estimated due to genetic factors, Ve is the variance estimated due to environmental factors, and (Vg+Ve) is the total variance.

A complementary estimation of heritability was undertaken using GCTA software (version 1.26.0) (Yang *et al.*, 2010; Zaitlen *et al.*, 2013) on the un-imputed genotyping data to assess SNP-based heritability (h^2_{SNP}); the fraction of phenotypic variance explained by all genotyped SNPs. A genetic relationship matrix (GRM) was created from the genotyping data (--make-grm-bin) and the --reml command was used to estimate variance of the traits explained by the genotyped SNPs, using the restricted maximum likelihood statistical approach. Had the GRM been estimated

from imputed SNPs, the estimate of variance explained by the SNPs would depend on the imputation quality control cut off (R^2) as this correlates with MAF, so the selection threshold would affect the estimate of variance explained by the SNPs, therefore the estimates here were undertaken on the unimputed genotyping data.

2.3.4 Genotyping imputation

Imputation was completed using the 39,235,157 genotyped SNPs from 64,976 participants of the 20 predominately European studies of the Haplotype Reference Consortium data (McCarthy *et al.*, 2016) through the Michigan Imputation Server (version v1.0.4) (Das *et al.*, 2016) for both pre-phasing with Eagle (version 2.3) (Loh *et al.*, 2016) and imputation using Minimac3 (Das *et al.*, 2016) with the European population of the reference panel (version hrc.r1.1.2016).

2.3.4.1 File set up for imputation

From the genotyping data, a set of files were created of autosomes only (--chr 1-22), due to the assessment of autosomes only in the genetic software, and contained information on 503,221 SNPs. The files were recoded into .map and .ped files (--recode) and the build was changed from hg18 (NCBI36) to hg19 (NCBI37) using LiftOverPlink.py [github.com/sritchie73/liftOverPlink/], adapted from the original LiftOver script (Hinrichs, 2006). 105 SNPs failed this step and were disregarded as is standard. The files were then recoded back to binary format (--make-bed) containing 503,116 SNPs. The frequency information of the files was analysed (--freq) and used to assess compatibility with the Human Reference Consortium (HRC) using Wrayner Tools [www.well.ox.ac.uk/~wrayner/strand/] from the McCarthy Group toolset [http://mccarthy.well.ox.ac.uk/software-tools/].

The command "perl HRC-1000G-check-bim-NoReadKey.pl -b mydata.bim -f plink.frq -r HRC.r1-1.GRCh37.wgs.mac5.sites.tab -h" was used to assess compatibility with version 4.2.11 of the perl script. The output showed that the majority of SNPs matched the IDs and positions of the HRC dataset, including DNA strands. The script then implicated changes to any unmatched SNPs, as follows; 502,028 SNP ID matches to the HRC data, 1,002 SNP IDs didn't match, 503,030 total SNP position matches, 1 different SNP position to HRC, and 85 SNPs had no match

to HRC. The script then implicated changes to the unmatched SNPs; the SNPs identified to require a change to the reference allele was 348,796 SNPs. DNA strand issues and other issues were also assessed by the script; 79 SNPs required strand changes, 154 were removed for allele frequency differences > 0.2, 5 palindromic SNPs were identified with a frequency of > 0.4, 1,148 SNPs didn't have matching alleles, ID and allele mismatching was identified for 614 SNPs, and no duplicates were identified.

The python script Run-plink.sh was created by the perl script, ready to remove inconsistencies, using --exclude, --update-map, --flip, --a2-allele command and creating VCF files for each chromosome (--recode vcf), which is the required format for the imputation servers. The VCF files were then sorted and zipped using bgzip and tabix. The checkVCF.py python file (version 1.4) was used to check the created VCF files for errors against the imputation server's reference file (human_g1k_v37.fasta). No errors were identified. The files were then uploaded to the Michigan Imputation Server containing 501,724 SNPs.

2.3.4.2 Post imputation analyses

The Michigan Imputation Server quality control (QC) assessed the allele frequency correlation between the cohort and the HRC reference panel, as in Figure 2-12. The quality control identified only 92 mismatched frequencies; markers where chi-squared is greater than 300.

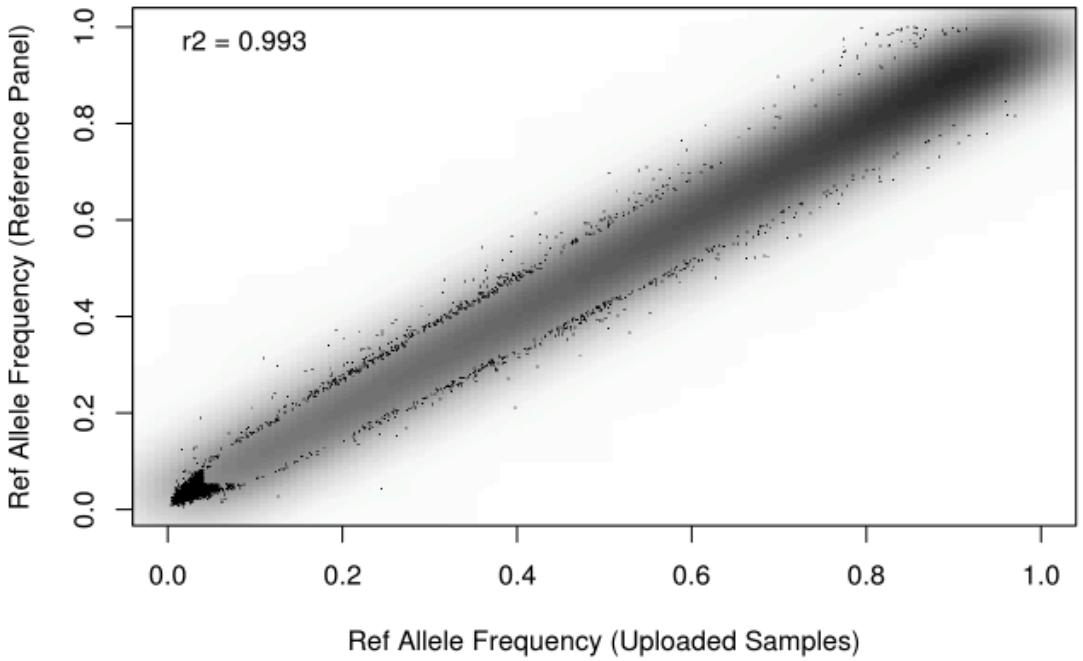


Figure 2-12: Allele frequency correlation from the Michigan Server QC Report.

The plot shows the densities of reference (Ref) allele frequencies. The first 5,000 points from areas of lowest regional densities are plotted.

The imputed files were downloaded and assessed using McCarthy Tools post imputation data checking program, *ic* [http://www.well.ox.ac.uk/~wrayner/tools/Post-Imputation.html], depicted in Figure 2-13. The results showed that there was a strong relationship between the cohort allele frequency and the HRC allele frequency, as described in the server's QC plot. There was a large percentage of SNPs with the frequency of the alternative allele imported from the HRC reference panel at 0.0. The QC Info score had an inflexion point at 0.30. There is a relationship between MAF of a variant in the imputed dosage data with Info score, as mentioned previously, where those with a high MAF have a higher info score, as common SNPs are imputed to a better quality, and this is depicted in Figure 2-13. The quality across the chromosome was good, and the whole chromosome was covered other than the centromere, as is standard for current technologies.

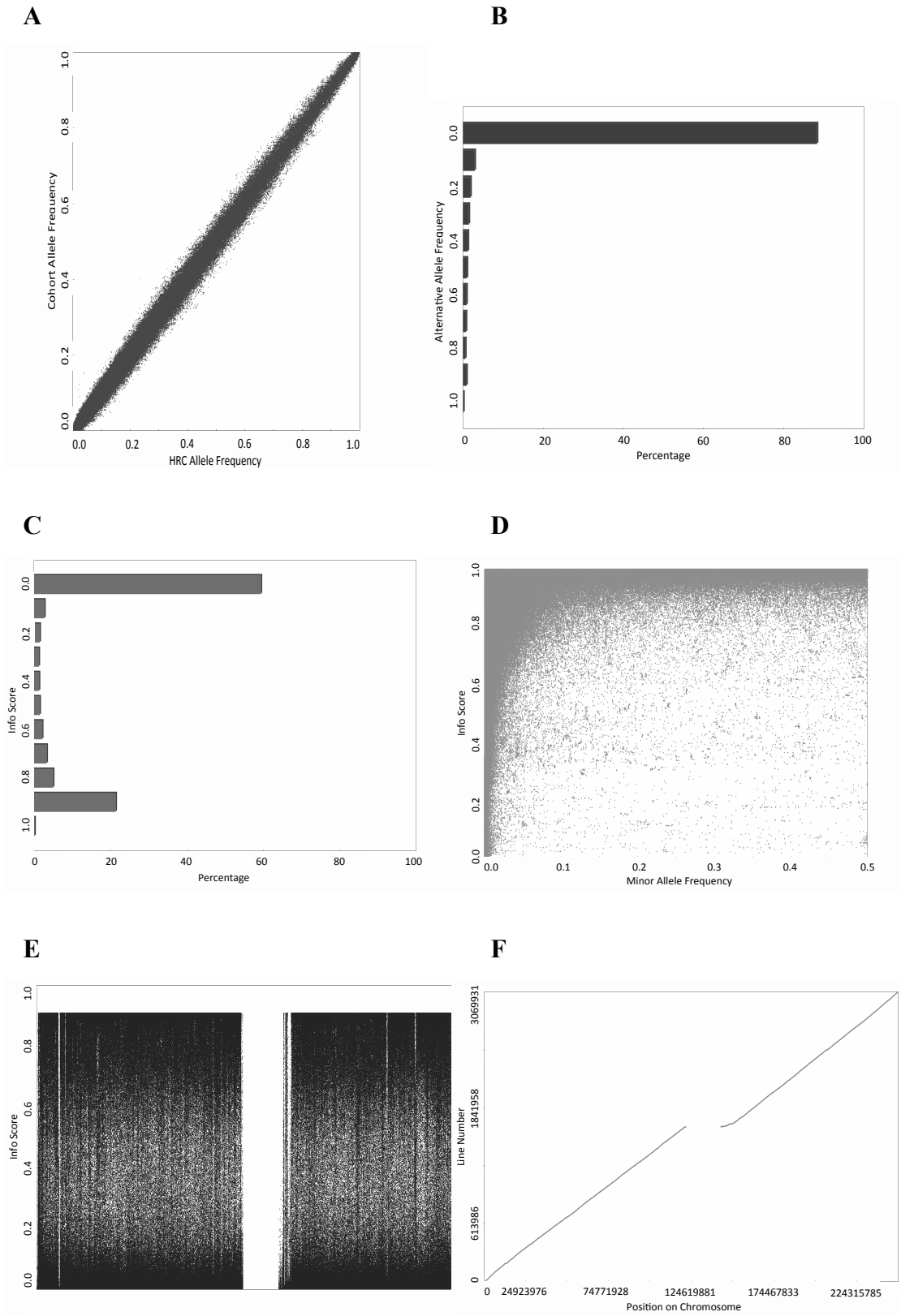

A
Cohort Allele Frequency
HRC Allele Frequency
0.0
0.2
0.4
0.6
0.8
1.0
B
Alternative Allele Frequency
Percentage
0
20
40
60
80
100
C
Info Score
Percentage
D
Info Score
Minor Allele Frequency
0.0
0.1
0.2
0.3
0.4
0.5
E
Info Score
F
Line Number
0
613986
1841958
3069931
24923976
74771928
124619881
174467833
224315785
Position on Chromosome

Figure 2-13: Assessment of imputation using *ic* software.

The images depict the results for chromosome one as an example of the data output. The software compares the results to that of the Human Reference Consortium reference genome (HRC). A) Comparison of allele frequency to HRC B) Percentage of SNPs at each allele frequency C) Percentage of SNPs at each Info Score D) The relationship between minor allele frequency and info score E) the coverage of chromosome 1 by info score, the black colour representing a good imputation score F) the coverage of chromosome one by imputed SNPs. The gaps present in the middle of panels E and F represent the centromere of the chromosome, and area not well genotyped by current short read technologies.

All quality control information can be found in Table 2.7. The mean MAF of the variant in the imputed dosage data was assessed over the chromosomes and the genome-wide mean per chromosome was 0.035. The average SNP call, the probability and certainty of observing the most likely allele for each haplotype, was 0.999; depicting good confidence in the most-likely genotypes.

The mean R^2 varied across chromosomes from 0.31-0.36, with a genome-wide mean of 0.33. The R^2 is an estimated value of the squared correlation between imputed genotypes and true, unobserved genotypes; a measure of the confidence in imputed dosages. It is based on the theory that poorly imputed genotype counts will shrink towards their expectations based on population allele frequencies alone (Das *et al.*, 2016).

The number of genotyped SNPs across chromosomes varied in relation to the size of the chromosome, and ranged from 7,643 for chromosome 21 to 41,641 for chromosome 2, with a genome-wide mean of 22,806. The imputed SNPs across chromosomes varied similarly, from 516,747 for chromosome 22 to 3,350,596 for chromosome 2, with a genome-wide mean of 1,755,245 per chromosome. In total, the imputed files contained information on 39,117,105 SNPs before quality control.

The file was converted from VCF format to PLINK binary format (--make-bed). Duplicate variants were identified and removed (n=109,097 across the genome, --exclude) and quality was assessed by R^2 score. As described, and in Figure 2-13 and Table 2.7, the mean score was 0.33 across chromosomes, at the point of the inflexion

point between the high levels of SNPs with R^2 scores of 0.0 and 0.9. Assessment of the correct thresholds for GWAS analysis are discussed in Chapter 3. A threshold of $R^2 > 0.80$ was used in this study, following the Chapter 3 analysis of current thresholds.

The variant IDs were in Chromosome:Position format, which was changed to SNP rsIDs using a downloaded reference file from dbSNP for the hg19 (NCBI37) build and --update-name PLINK command. The imputed files were then merged with the genotyped files (--bmerge, --merge-mode 2) to include genotyped SNPs. 93,374 genotyped SNPs were retained that were mismatched with the imputed SNPs. The final set of files for each chromosome was merged into one set of files (--merge-list, --make-bed) for quality control.

Table 2.15: Quality control assessment of Imputation data.

Chr, chromosome; MAF, minor allele frequency; Avg Call, average imputation call; R2, correlation between imputed and genotyped SNPs; Dups, duplicates; Mergefails_keptgenotyped, where the merge between genotyped SNPs and imputed data failed and the genotyped were kept.

									R^2>0.3		R^2>0.8		
Chr	**Mean MAF**	**Mean Avg Call**	**Mean R2**	**N Genotyped**	**N Imputed**	**Total Imputed**	**N Dups**	**Total No Dups**	**Poor Quality**	**Good Quality**	**Poor Quality**	**Good Quality**	**Mergefails keptgenotyped**
1	0.034	0.999	0.326	39114	3030817	3069931	6768	3056381	1955181	1110536	2252180	815142	6421
2	0.033	0.999	0.326	41641	3350596	3392237	7952	3376323	2168866	1218587	2478746	910494	7991
3	0.035	0.999	0.334	34680	2787214	2821894	6709	2808472	1780308	1037375	2040791	778363	6564
4	0.035	0.999	0.336	30393	2757188	2787581	6641	2774292	1751774	1031722	2009776	775330	5079
5	0.034	0.999	0.328	31318	2556850	2588168	6122	2575921	1649548	934903	1882167	703644	5375
6	0.037	0.999	0.351	33487	2426624	2460111	5996	2448113	1513432	942621	1728961	728483	5567
7	0.035	0.999	0.331	27850	2261455	2289305	5490	2278316	1449877	836005	1662206	624930	5095
8	0.034	0.999	0.330	28558	2214147	2242705	6110	2230479	1423426	815402	1627554	612663	5645
9	0.034	0.999	0.336	24629	1651269	1675898	6251	1663382	1053580	617434	1206785	465796	5332
10	0.036	0.999	0.345	26906	1900597	1927503	6958	1913572	1194900	727008	1371687	552122	4978
11	0.035	0.999	0.333	24899	1912091	1936990	6951	1923069	1224972	706555	1400224	533179	4822
12	0.035	0.999	0.335	25055	1823062	1848117	6322	1835464	1162453	680680	1335357	509510	4377
13	0.036	0.999	0.340	19206	1366227	1385433	4643	1376141	867504	514176	990982	391994	3243
14	0.034	0.999	0.330	16898	1249638	1266536	4433	1257662	805005	458040	920817	343425	3044
15	0.034	0.999	0.322	15287	1123928	1139215	4090	1131026	729863	406197	840555	296680	3215
16	0.033	0.999	0.313	15585	1265712	1281297	4022	1273245	832863	445990	959131	320652	3688
17	0.034	0.999	0.310	13428	1076644	1090072	3823	1082411	707172	380022	826476	261914	2602
18	0.035	0.999	0.338	15332	1089423	1104755	2676	1099399	690473	412554	798391	305311	2849
19	0.035	0.999	0.319	8926	859628	868554	2261	864025	552606	314505	655363	212412	1759
20	0.035	0.999	0.326	13090	871893	884983	2236	880506	563469	320142	650286	233827	2615
21	0.037	0.999	0.325	7643	523633	531276	1268	528739	338440	192055	392359	138458	1468
22	0.036	0.999	0.328	7797	516747	524544	1375	521787	331335	192288	387174	136808	1645
GW	**0.035**	**0.999**	**0.330**	**22806**	**1755245**	**39117105**	**109097**	**38898725**	**24747047**	**14294797**	**28417968**	**10651137**	**93374**

2.3.4.3 Imputed data quality control

The imputed files did not contain information on the full family pedigree nor gender assignments, so this was reinstated into the file (--update-parents, --update-sex). Information on 1,219 individuals (573 males, 646 females) including 243 founders and 976 non-founders were in the files, including 10,652,600 SNPs. Mendelian inconsistencies were assessed (--mendel-multigen) and 3,200,651 SNPs were affected, depicted in Figure 2-14. All were set to missing (--set-me-missing).

As imputation was completed on the whole cohort (n=1,219) and the maximum overlap with lipid results was 999 individuals, the --keep flag was used to create a set of genotyping files that were specific for the lipidomics analysis, including 10,652,600 variants for 999 people (198 founders, 801 non-founders; 476 males, 523 females) for the final analysis, and 196 people (59 founders and 137 nonfounders; 93 males, 103 females) for the range finding study.

The same thresholds were used as before for the genotyping data, except with the inclusion of a stricter MAF threshold of 0.05, as there was an increase in low frequency variants. The final QC thresholds were as follows; --geno 0.05, --mind 0.05, --maf 0.05, --hwe 1e-8, depicted in Figure 2-15. 36 variants were removed using --geno 0.05, no further individuals were removed using --mind 0.05, 5,372,027 low frequency variants were removed using --maf 0.05, and 78 variants were removed due to Hardy-Weinberg exact test (--hwe 1e-8). The final set of files contained 5,280,459 variants for 999 individuals.

For the range finding study, 11,255 variants were removed using --geno 0.05, no further individuals were removed using --mind 0.05, 5,388,226 low frequency variants were removed using --maf 0.05, and 23 variants were removed due to Hardy-Weinberg exact test (--hwe 1e-8). The final set of files contained 5,253,096 variants for 196 individuals.

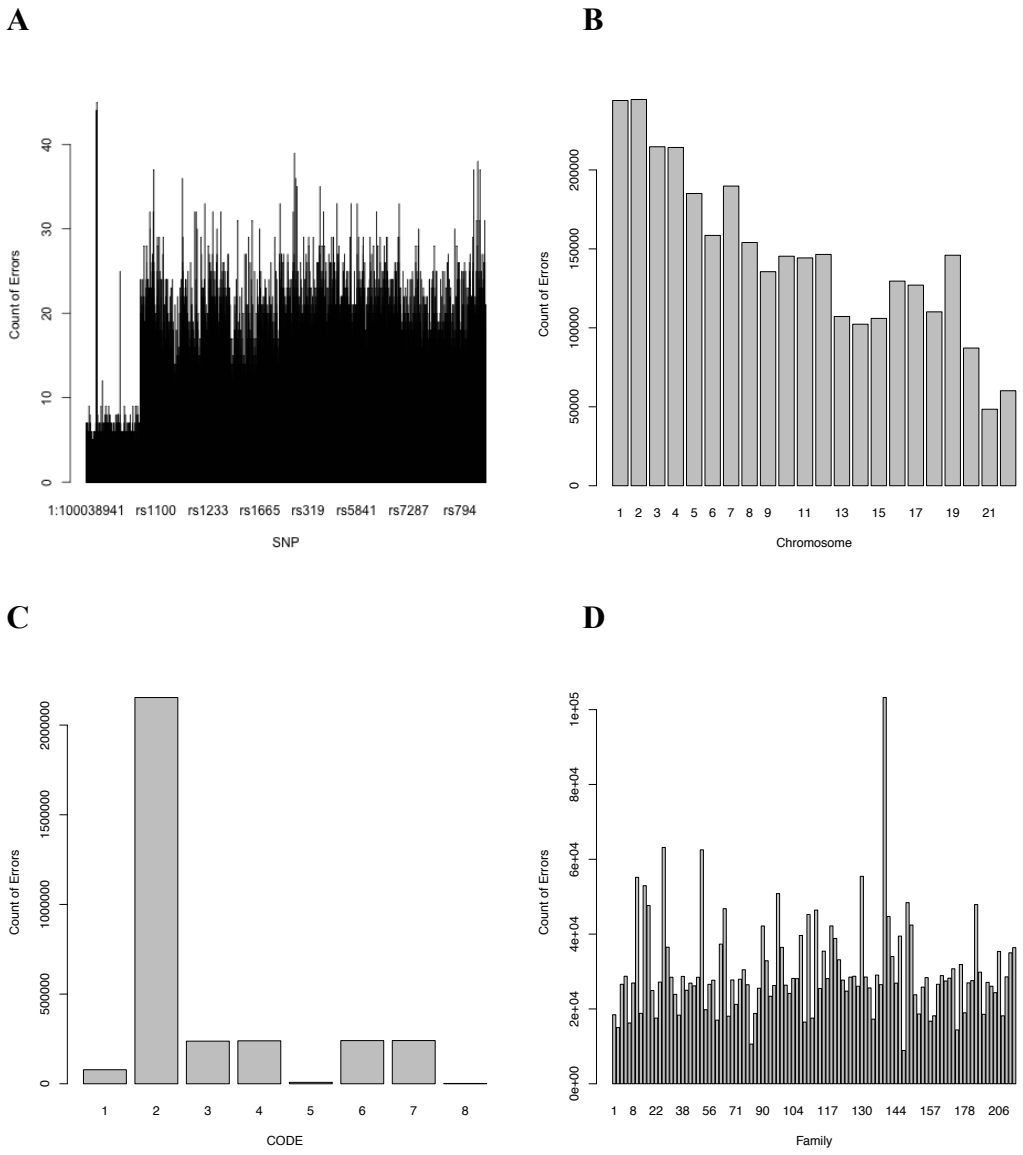


Figure 2-14: Imputation data Mendelian inconsistencies analysis.

The figure depicts the assessment of Mendelian inconsistencies (errors) identified for the cohort. A) the number of errors (Y-axis) for each SNP (X-axis). B) the number of errors (Y-axis) for each chromosome (X-axis). C) the number of errors (Y-axis) by code (X-axis). Codes are described in Figure 2.3. D) the number of errors (Y-axis) by participant family (X-axis).

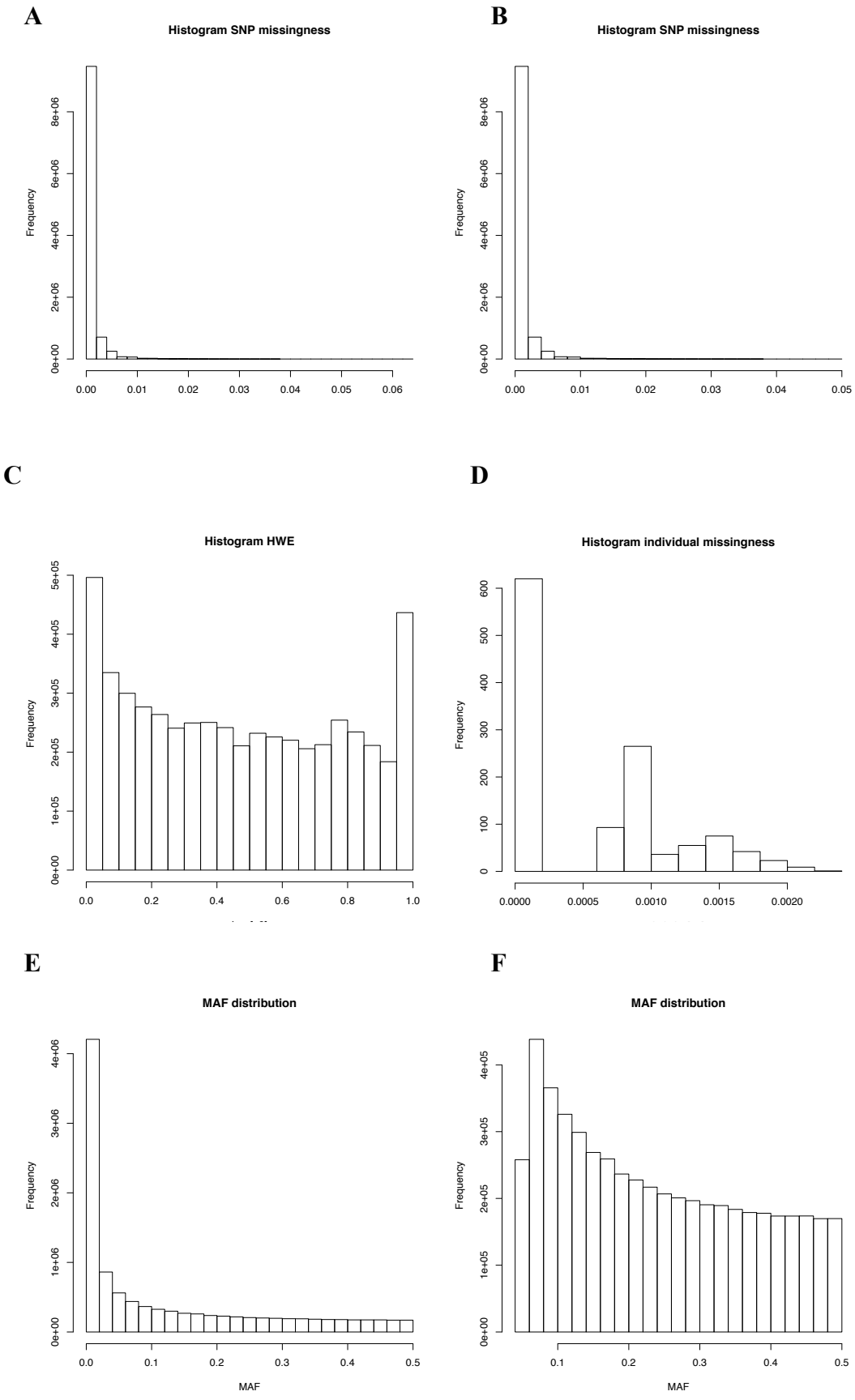

A
Histogram SNP missingness
Frequency
0e+00
2e+06
4e+06
6e+06
8e+06
0.00
0.01
0.02
0.03
0.04
0.05
0.06
B
Histogram SNP missingness
Frequency
0e+00
2e+06
4e+06
6e+06
8e+06
0.00
0.01
0.02
0.03
0.04
0.05
C
Histogram HWE
Frequency
0e+00
1e+05
2e+05
3e+05
4e+05
5e+05
0.0
0.2
0.4
0.6
0.8
1.0
D
Histogram individual missingness
Frequency
0
100
200
300
400
500
600
0.0000
0.0005
0.0010
0.0015
0.0020
E
MAF distribution
Frequency
0e+00
1e+06
2e+06
3e+06
4e+06
0.0
0.1
0.2
0.3
0.4
0.5
MAF
F
MAF distribution
Frequency
0e+00
1e+05
2e+05
3e+05
4e+05
0.1
0.2
0.3
0.4
0.5
MAF

Figure 2-15: Imputation quality control for the full cohort analysis

The figure depicts the quality control steps; the changed in SNP missingness after the 5% missingness threshold was applied (A, before quality control; B, after quality control), a histogram of the Hardy-Weinberg exact test where 78 variants were removed (C, no change in figure so only one shown), the individual missingness was less than 5% and therefore no individual was removed at this stage (D), and the distribution of the minor allele frequency before quality control (E), and after quality control (F).

2.3.5 Family-based genome-wide association studies

Family-based GWAS analyses were completed with knowledge gained from attendance at the 2016 Wellcome Trust Genome Campus Advanced Course on Genome-wide Association Studies, Hinxton, UK. Linear mixed modelling approaches were used to account for population substructure and relatedness in GWAS. Family-based GWAS were assessed using three software; PLINK (version 1.9) (Purcell *et al.*, 2007), FaST-LMM (version 2.07.20140304) (Lippert *et al.*, 2011), and GCTA (version 1.26.0) (Yang *et al.*, 2010; Zaitlen *et al.*, 2013), with results described in Chapter 3. The result of the Chapter 3 analysis concluded that GCTA software would be used in this project, specifying mixed linear model association analyses (--mlma).

Population structure can lead to false positive GWAS associations. To adjust for population structure, matching can be completed across cases and controls, or family data can be collected but requires two parents. However, inferences can be made across the genome to adjust for population ancestry; population structure has been hypothesized to inflate the distribution of Cochran-Armitage trend tests genome-wide, by a constant multiplicative factor, lambda (λ). This is estimated by the median Chi-squared divided by 0.456. If the resulting test statistic is greater than 1, population structure or genotyping error is present. A test statistic less than 1 represents an underpowered analysis. This can be visualised by Quantile-Quantile plot. Such genomic inflation factors (GIF) were estimated by the following equations in R software:

$$chi = qchisq(1 - Pvalue, 1)$$

Equation 2.8: Calculation of chi

$$lambda = \frac{median(chi)}{0.456}$$

Equation 2.9: Calculation of lambda, the genomic inflation factor

where qchisq, is the chi-squared calculation in R; P-value, is the GWAS P-value of association; lambda, genomic inflation factor; the result was calculated for 1 degree of freedom.

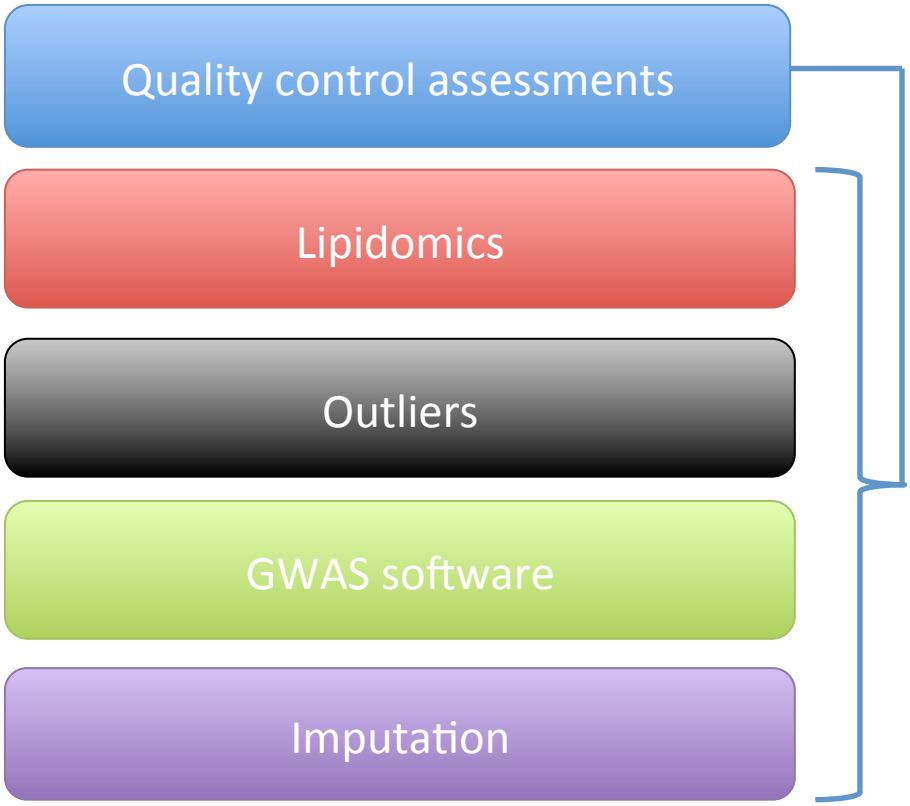
Manhattan plots, QQ plots, lists of significant SNPs and calculations of GIFs was undertaken in R for each GWAS result. Significantly associated SNPs were analysed by Ensembl API Client (version 1.1.5 on GRCh37.p13) to identify neighbouring genes. Further analyses were undertaken of the significantly associated SNPs; expression quantitative trait loci (eQTL) were identified using the GTEx eQTL Browser (version 8) [RRID:SCR_001618], assessment of previously identified SNPs from GWAS was undertaken using the GWAS: Catalog of Published Genome-Wide Association Studies [RRID:SCR_012745], review of variants on OMIM [RRID:SCR_006437], visualisation of variants using UCSC Genome Browser [RRID:SCR_005780], and assessment of PheWAS with the UK Biobank (Sudlow *et al.*, 2015) was undertaken using the Gene Atlas Browser [RRID:SCR_017577]. The least significant P-values of the significantly associated SNPs ($P<5x10^{-8}$) are depicted as P<X in Chapter 5.

2.3.6 Two-sample Mendelian randomisation analyses

Two sample Mendelian randomisation (2SMR) analysis was undertaken in R following the guidelines provided by Davey Smith et al [https://mrcieu.github.io/TwoSampleMR/] (Smith *et al.*, 2014). Selected examples of the significant associations from the full cohort analysis, identified for each class of lipid, were analysed by 2SMR for a number of previously published GWAS of interest; the GWAS significant associations ($P<5x10^{-8}$) identified for NAE species PEA, and CER traits CER[N(22)S(19)], CER[N(24)S(16)], and CER[N(24)S(19)]/CER[N(24)DS(19)] ratio, were assessed for association with coronary artery disease (Nikpay *et al.*, 2015) (all traits assessed), Type-2 Diabetes (Mahajan *et al.*, 2014) (CER[N(22)S(19)] only), and blood cell counts (Astle *et al.*, 2016) (CER[N(24)S(16)] and CER[N(24)S(19)]/CER[N(24)DS(19)] ratio were assessed). Details on the published GWAS used as outcomes are presented in Appendix Table 0.2. As many GWAS associated SNPs were in linkage disequilibrium (LD), the following SNPs remained in the analysis after the data clumping step that removes SNPs in LD; rs324420 (*FAAH*; PEA), rs438568 (*SPTLC3*; CER[N(22)S(19)]), rs7160525 (*SGPP1*; CER[N(24)S(16)]), and rs4653568 (*DEGS1*; CER[N(24)S(19)]/CER[N(24)DS(19)] ratio).

Chapter 3

Quality control assessments



3.1 Introduction, aim and objectives

The aim of the quality control studies undertaken in this Chapter, was to identify a robust set of lipid species detectable in the study plasma samples, and the most appropriate bioinformatic techniques and thresholds, which would allow for high quality and throughput analyses of the genetics of the lipid species identified.

Lipidomics

Certain lipids of the Eico class have been shown to be altered in concentration upon stimulation, such as during coagulation and platelet clotting in blood (Ishikawa *et al.*, 2014). To insure that the study plasma samples were not affected by coagulation and clotting of platelets, and to identify species that require stimulation to be detected in blood, the lipidomic results of the study plasma samples were compared to test serum samples and test plasma samples from another cohort, both analysed in the same laboratory. The identified lipid species that were significantly different between serum and plasma were thromboxane B_2 (TXB_2) and 12-HETE, which have been described previously (Ishikawa *et al.*, 2014).

Commercial standards are not available for each species of CER, therefore further assessments were required to confirm the identity of the plasma CER species through an additional analysis of multiple reaction monitoring (MRM). This incorporated literature-based and class-based calculated mass spectrometry transitions. Identification of the peak of a lipid species at the same retention time in three chromatographs (MRMs) of varying product transitions selected for each specific lipid species, allowed for confirmation of the detection of specific CER species, and for retention time referencing. Only CER species with multiple MRMs confirming their specific retention time were included in the genetic analyses.

To assess the reliability of the lipidomic results for confirmation of the integrity of the final genetic results in this study, a quality control analysis was undertaken for each lipid species analysed, assessing; mass spectrometry injection variation, lipid recovery, process efficiency, matrix effects, and sample carryover. This allowed for the inclusion of only lipid species that are reliably measured in the plasma samples.

Only lipid species that met defined quality control criteria were included in future genetic analyses.

Statistics and Genetics

Lipidomic values were assessed after adjustment for covariates using the R package 'car' to identify extreme outlier values by a Bonferroni p-value of $P<0.05$ that was created for each observation by testing them as a mean-shift outlier, based on studentized residuals. To assess the effect of the inclusion and exclusion of outliers on final genetic analyses, an exemplar GWAS was undertaken to compare the results, which presented with improved genetic analyses quality when extreme values were removed.

As many lipidomic species were measured from the same biochemical pathways, including product and precursor species; summation and ratio traits were created in the aim of identifying GWAS associations with SNPs influencing the genes of proteins involved in each reaction of the respective metabolic pathways. Analysis of ratios between related metabolites reduces the overall biological variability, increases statistical power, experimental errors are cancelled out, the overall noise is reduced, and metabolite ratios approximate the reaction rate under idealized steady state assumptions, representing the flux through a biochemical pathway (Petersen et al., 2012).

Underlying population structure can have an effect on genetic analyses, and close relatedness violates the assumptions of conventional GWAS analyses, i.e. no pair of individuals are more closely related than second-degree relatives (Marees *et al.*, 2018). Standard errors of SNP effect sizes are biased if this assumption is violated. To assess the effect of the family-based cohort used in this study on GWAS results, three software were assessed, those that adjust for population structure (GCTA, FastLMM) and a comparison software that doesn't (PLINK). Of the two software that adjusted for the family-based cohort, the results were very similar, differing only by speed of analysis. GCTA was used for the final GWAS assessments of the lipid traits due to the fast speed of its analysis.

Genotyping chip data in this project contained ~0.5M SNPs. Upon imputation using the Human Reference Consortium (McCarthy *et al.*, 2016), this increased to ~7.5M SNPs. This caused a structure to present in the quantile-quantile plots resulting from a test GWAS analysis. Two thresholds were therefore assessed; the effect of R^2 imputation quality control threshold and the effect of minor allele frequency (MAF) on the resulting GWAS analyses. Low thresholds of both parameters allowed for the identification of low frequency variants and badly imputed variants, therefore high thresholds were used for the final GWAS analyses (R^2>0.8; MAF>0.05), and the structure of the resulting Quantile-Quantile plots replicated those of the genotyping-only analyses.

The objectives were as follows:

1. Identify lipids detectable in the cohort plasma samples
2. Examine the quality of the lipidomics assay results
3. Calculate extra lipidomic traits from the lipid species
4. Assess the impact of outlier lipid measures on the genetic results
5. Identify the most appropriate family-based GWAS software
6. Identify optimal quality control thresholds for imputed genetic data

3.2 Lipidomics data

3.2.1 Detection of lipid mediators in plasma

Lipids were deemed detected if they were identified above a signal-to-noise (S/N) ratio of 10:1, a peak area of 150, found at a constant retention time (allowing for persistent shifting) comparable to the commercial standards, and identified in a substantial percentage of the samples analysed (>60%). The S/N ratio is the ratio of the peak height to that of the baseline noise of the chromatogram. The limit of detection defined by Waters Corporation is a signal-to-noise ratio of 3:1, and the limit of quantification is set at a signal-to-noise ratio of 10:1 (Waters Corporation, 1998). Species were only included in analysis if found above this limit of quantification as this is the smallest concentration of a lipid that can be reliably measured by the Waters mass spectrometer (Armbruster *et al.*, 2008).

3.2.1.1 Eico species

Of the assay of 83 prostanoids, eicosanoids, dodecanoids, octadecanoids, and resolvins in the Eico array, 19 species of eicosanoids, dodecanoids, and octadecanoids were detected above the limit of quantification in the cohort plasma. The levels of some lipid mediators of the Eico class are affected by the process of coagulation and platelet aggregation (Ishikawa *et al.*, 2013). Serum is blood that has been allowed to clot, while clotting is prevented with the addition of an anticoagulant in plasma samples. A comparison was made with biobanked serum samples (n=10) and plasma samples from separate cohorts (n=21) for comparison, to ensure the cohort plasma samples were not affected by coagulation or platelet activation.

Multiple t-tests were undertaken using Prism Graph Pad (version 8) to identify statistically different levels of 30 Eico lipids that were detected in serum or plasma. Two species were identified as significantly increased (P_{adj}<0.000001); TXB_2 and 12-HETE were substantially raised in the serum samples compared to cohort (HTO) plasma (Figure 3-1). The p-value was controlled using FDR and adjusted for multiple comparisons (30 tests) using Benjamini and Hochberg methods; the results remained P_{adj}<0.001 (q=<0.000001 for both; Q=1%). TXB_2 plasma levels were measured at 2 ± 5 pg/ml (mean ± SD) in the cohort plasma (n=204), 68 ± 228 pg/ml in comparison

plasma (n=21), and substantially raised in serum (9,384 ± 11,533 pg/ml; n=10). 12-HETE plasma levels were measured at 117 ± 82 pg/ml in the cohort plasma (n=204), 254 ± 153 pg/ml in comparison plasma (n=21), and also substantially raised in serum (14,999 ± 15,773 pg/ml; n=10).

As depicted in Figure 3-1, there were three participants with extreme serum values; their levels of 12-HETE and TXB_2 are linked via coloured bars in the diagram. When these individuals were excluded from analyses, the results remained significant ($P_{adj}<0.000001$). TXB_2 is the inactive derivative of TXA_2, which has roles in platelet aggregation, and 12-HETE is produced primarily from platelets (Petroni *et al.*, 1995; Yoshimoto *et al.*, 2002). This highlighted the lack of affect of platelet activation or coagulation on the cohort plasma samples.

3.2.1.2 NAE species

Of the 29 n-acyl ethanolamines (NAE) and glycerols included in the NAE assay, 11 NAEs were detected in the cohort plasma samples.

3.2.1.3 CER species

Of the 53 sphingolipids included in analyses (CER[NS], CER[NDS], C18S species, C1P, CER[AH], CER[AS], CER[ADS]), 39 species of CER[NS], CER[NDS], C18 species, and CER[AS] were detected in the cohort plasma samples.

3.2.1.4 Comparison with published lipid concentrations in literature

Comparisons were undertaken with published concentrations of CER, NAE and "Eico" lipids. The cohort plasma analyses of n=1,016 plasma samples is dubbed as "Cohort". Reported values and lipid nomenclature are presented before conversion to pmol/ml for comparison to the literature. No other study was found to cover the large range of species analysed in this study, for all three groups of lipids.

The plasma concentrations of NAE and Eico species, which have species-specific commercial standards available of analysis and calibration lines are used in the analysis of these species, allowed for the values to be comparable to literature where these standards were used for quantification. The CER species ranged in

concentration between publications, including those publications that analysed the same pooled plasma reference material, as CER concentrations are relative to a fixed amount of internal standard. Since this study, the laboratory the samples incorporated an extraction protocol that incorporates an SPE clean up step to reduce matrix effects. Thus, for comparison to literature, the plasma CER levels have been adjusted for such matrix effects.

This was undertaken by linear regression using a set of 20 healthy plasma samples analysed under the new extraction protocol, whose mean CER lipid levels correlated strongly with those found in this study (R=0.95; Figure 3.1.1). The conversion factor was used to normalise the standard deviation (SD) and standard error of the mean (SEM). The normalised concentrations, presented in Table 3.1.1, were only used for comparison with the literature, as the values used in the genetic analyses were not based on pmol/ml levels, but adjusted for covariates and mean standardised (presented in Section 3.2.1.5). In some publications depicted in this section, this may have led to the lack of reporting the summary statistics of the concentrations for CER species in plasma (Hicks et al. 2009).

Differences in instrumentation have been described to influence the relative abundance of reported CER measures as sensitivity for CER species and standards can differ by orders of magnitude, shown by a study analysing the same plasma sample on different instruments (Shaner et al. 2009). Differences in the extraction protocols, including reagents (not chloroform-methanol extractions), volumes (use of 96 well plates), and handling of the samples (use of sonication, blending, collection in non-EDTA anticoagulants); or differences in quantitation through the use of standards, which ranged from using one standard for rank-based quantitation to using one standard per CER subgroup also influence the reported values between studies. However, in all cases, comparison with publications that have reported values for a substantial number of CER species, showed that the most abundant plasma CER of those CERs studied here was CER[N(24)S(18)], and it was at much larger abundance than the other studied CER species. The strong relationship described in the following section between the raw CER levels found in this study and those reported in publications, highlights the similarity between the measurements in plasma (R>0.75).

Finally, an assessment was undertaken to confirm that the CER concentrations found for the pooled plasma and Cohort samples analysed in this study were in a similar pmol/ml range to other plasma samples (Cohorts 1-3; n=10-80) that were analysed in the same laboratory. The use of the pooled QC samples for adjustments allowed the samples to be normalised for batch effects in a species-specific manner. As genetic studies only require quantitative values to identify patterns across families and not exact plasma abundances, the use of the values after adjustment for covariates allowed for heritability estimates and identification of genetic loci of the genes of known CER metabolic enzymes that are involved in influencing CER lipid species.

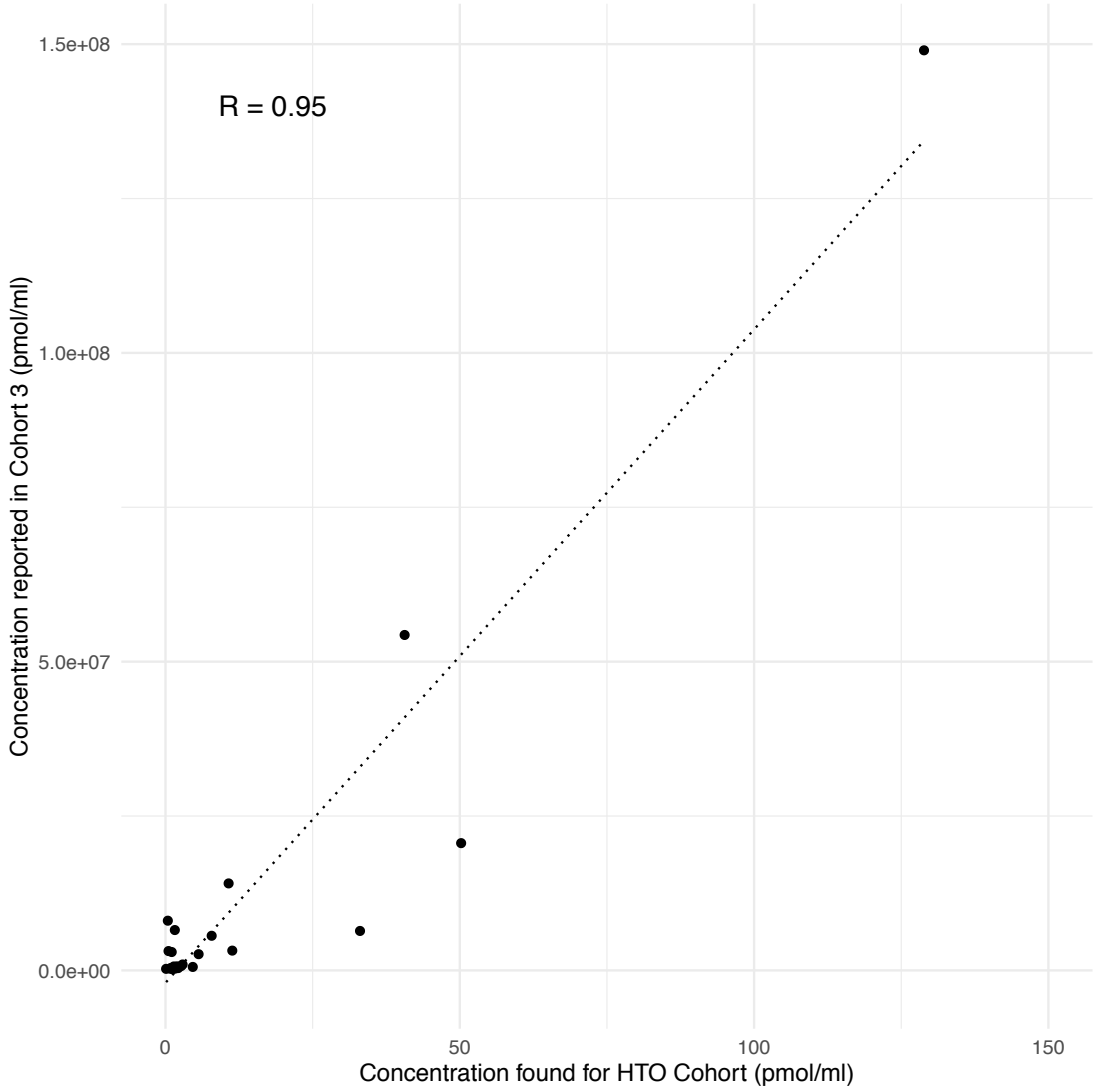


Figure 3-1.1: Correlation between the plasma CER levels of healthy volunteers with adjustment for matrix effects and the HTO CER levels
A correlation was undertaken with plasma CER levels from this study (HTO Cohort) and 20 healthy volunteers (Cohort 3) which incorporated an SPE step in extraction to remove matrix effects. The correlation coefficient (R) depicted a strong correlation of 0.95.

Table 3.1.1: Normalised plasma CER values for matrix effects
The summary statistics of the raw cohort values for CERs (Cohort Raw) were normalised by an adjustment for matrix effects (Cohort Normalised). Standard deviation, SD; standard error of the mean, SEM.

Ceramide	Cohort Raw (pmol/ml) Mean	Cohort Raw (pmol/ml) SD	Cohort Raw (pmol/ml) SEM	Cohort Normalised (pmol/ml) Mean	Cohort Normalised (pmol/ml) SD	Cohort Normalised (pmol/ml) SEM
CER[N(20)S(18)]	0.4	0.27	0.01	18.32	12.49	0.38
CER[N(22)DS(18)]	0.52	0.38	0.01	20.85	15.10	0.48
CER[N(22)S(18)]	5.63	3.25	0.10	128.48	74.25	2.31
CER[N(22)S(19)]	1.06	0.73	0.02	32.22	22.18	0.68
CER[N(23)S(18)]	40.63	17.14	0.57	865.65	365.16	12.06
CER[N(23)S(20)]	1.96	0.62	0.02	51.18	16.29	0.54
CER[N(24)DS(18)]	7.85	5.03	0.17	175.23	112.21	3.69
CER[N(24)DS(19)]	2.67	1.69	0.05	66.13	41.81	1.32
CER[N(24)DS(20)]	1.42	0.78	0.03	39.80	21.84	0.72
CER[N(24)S(16)]	1.86	1.15	0.04	49.07	30.24	0.95
CER[N(24)S(17)]	10.73	4.81	0.16	235.89	105.83	3.45
CER[N(24)S(18)]	128.87	61.00	1.96	2724.18	1289.51	41.52
CER[N(24)S(19)]	50.23	22.86	0.66	1067.85	486.08	13.97
CER[N(24)S(20)]	11.35	4.24	0.14	248.95	93.07	3.06
CER[N(24)S(22)]	1.69	1.16	0.04	45.49	31.32	1.05
CER[N(25)DS(18)]	1.07	0.55	0.02	32.43	16.79	0.55
CER[N(25)S(20)]	1.34	0.67	0.02	38.12	19.10	0.65
CER[N(26)DS(18)]	0.77	0.41	0.01	26.11	13.90	0.46
CER[N(26)S(18)]	33.05	10.26	0.35	706.00	219.21	7.37
CER[N(26)S(19)]	4.64	3.18	0.11	107.62	73.66	2.45
CER[N(27)S(18)]	2.16	1.56	0.05	55.39	39.97	1.39
CER[N(28)S(18)]	0.88	0.56	0.02	28.43	17.94	0.62
CER[N(29)S(18)]	1.21	1.40	0.05	35.38	40.90	1.39

3.2.1.4.1 Plasma ceramide concentrations in literature

Quehenberger et al. (2010) analysed pooled plasma in triplicate of 22L of plasma from 100 fasting individuals. The samples were collected in lithium heparin, which is described in Section 6.4.2 to enhance plasma phospholipase A2, the enzyme that releases lipids from membrane phospholipids, and has been shown to interfere with mass spectrometry analyses of CER. Differences with the experimental of this study include the blending blood to create the pooled aliquots and extraction with multiple rounds of sonication. The CER species showed a strong correlation with the reported values (R=0.77), and highlighted CER[N(24)S(18)] as the most abundant CER species of those studied (Table 3.2.1).

Table 3.2.1: Comparison of plasma lipid concentrations reported here with Quehenberger et al. (2010)
Reported measures from Quehenberger et al. (2010) (Reported Lipid; Reported Mean; Reported SEM), conversion of the reported values to pmol/ml (Calculated Mean; Calculated SEM), and the cohort values found in this study (Cohort Mean; Cohort SD; Ceramide) are depicted. Mean and standard error of the mean (SEM) are depicted.

Reported Lipid	Reported (nmol/ml) Mean	Reported (nmol/ml) SEM	Calculated (pmol/ml) Mean	Calculated (pmol/ml) SEM	Cohort (pmol/ml) Mean	Cohort (pmol/ml) SEM	Ceramide
d18:1/C20:0	0.145	0.007	145	7	18	0.38	CER[N(20)S(18])
d18:0/C22:0	0.685	0.003	685	3	21	0.48	CER[N(22)DS(18)]
d18:1/C22:0	1.22	0.046	1220	46	129	2.31	CER[N(22)S(18)]
d18:1/C23:0	0.281	0.033	281	33	866	12.06	CER[N(23)S(18)]
d18:0/C24:0	1.22	0.004	1220	4	175	3.69	CER[N(24)DS(18)]
d18:1/C24:0	3.00	0.107	3000	107	2724	41.52	CER[N(24)S(18)]
d18:0/C26:0	0.041	0.002	41	2	26	0.46	CER[N(26)DS(18)]
d18:1/C26:0	0.061	0.0007	61	0.7	706	7.37	CER[N(26)S(18)]

Bowden et al. 2017 reports the analysis of pooled plasma by a consortium of investigators, using the same plasma source as Quehenberger et al. (2010). In this study, the samples were analysed in different laboratories, with each laboratory using their own CER extraction pipeline and instrumentation. Only the concentrations for CER[N(24)DS(18)] and CER[N(26)S(18)] were reported (Table 3.3.1). The concentration reported for CER[N(26)S(18)] (63 pmol/ml) was similar to the Queheneberger et al. (2010) study (61 pmol/ml). However, the concentrations reported for CER[N(24)DS(18)] were 1220 pmol/ml reported by Quehenberger et al.

(2010), Bowden et al. (2017) reported 280 pmol/ml, and this study found 175 pmol/ml. This may be due to the differences in the reported rank-based concentrations of CER species due to different instrumentation or the stability of the CER species. The future availability of CER species-specific commercial standards to construct calibration lines would aid this analysis.

Table 3.3.1: Comparison of plasma lipid concentrations reported here with Bowden et al. (2017)
Reported measures from Bowden et al. (2017) (Reported Lipid; Reported Median; Reported Uncertainty), conversion of the reported values to pmol/ml (Calculated Median; Calculated Uncertainty), and the cohort values found in this study (Cohort Mean; Cohort SD; Ceramide) are depicted.

Reported Lipid	Reported (nmol/ml) Median of lab means	Reported (nmol/ml) Standard Uncertainty	Calculated (pmol/ml) Median of lab means	Calculated (pmol/ml) Standard Uncertainty	Cohort (pmol/ml) Mean	Cohort (pmol/ml) SD	Ceramide
CER d42:0	0.28	0.18	280	180	175	112	CER[N(24)DS(18)]
CER d44:1	0.063	0.031	63	31	706	219	CER[N(26)S(18)]

Kauhanen et al. (2016) analysed the plasma from 42 healthy participants and extracted CERs using ethyl acetate:isopropanol. The value presented in Table 3.4.1 was obtained from a figure in the article, as summary statistics were not presented.

Table 3.4.1: Comparison of plasma lipid concentrations reported here with Kauhanen et al. (2016)
Reported measure for CER[N(24)S(18)] from Kauhanen et al. (2016) (Reported Median), the final pmol/ml values after conversion (Calculated Median) and the cohort values found in this study (Cohort Mean) are presented. The column "Ceramide" uses the nomenclature used in this study, while "Reported Lipid" is the species reported in Kauhanen et al. (2016).

Reported Lipid	Reported (pmol/µl) Median	Calculated (pmol/ml) Median	Cohort (pmol/ml) Mean	Ceramide
Cer d18:1/24:0	2.23	2230	2724	CER[N(24)S(18)]

Three studies, Tabassum et al. (2019), Hicks et al. (2009), and Bellis et al. (2014), completed similar genetic analyses of plasma CERs. Tabassum et al. (2019) analysed

2,181 participants and reported only median values. No units were presented in the paper or supplementary, thus it is assumed that the units are the same as another paper published by the same group (Surma et al. 2015), where CER range was reported in μmol/L. The samples that fasted for less than 6 hours are summarised in Table 3.5.1. One of the papers presenting genetic analyses of plasma CER did not publish reference ranges (Hicks et al. 2009).

Table 3.5.1: Comparison of plasma lipid concentrations reported here with Tabassum et al. (2019)
Reported measures from Tabassum et al. (2019) (Reported Median), the final pmol/ml values after conversion (Calculated Median) and the cohort values found in this study (Mean Cohort) are depicted. The column "Ceramide" uses the nomenclature used in this study, while "Reported Lipid" is the species name reported in Tabassum et al. (2019).

Reported Lipid	Reported (assume μmol/L) Median	Calculated (pmol/ml) Mediam	Cohort (pmol/ml) Mean	Ceramide
CER(40:1;2)	0.68	680	129	CER[N(22)S(18)]
CER(42:1;2)	1.83	1830	2724	CER[N(24)S(18)]

Bellis et al. (2014) also published a genetic analysis of plasma CER levels in family samples (n=1,212). The extraction procedure included the addition of anti-oxidant butyl hydroxytoluene (Weir et al. 2013) and multiple rounds of sonication (Meikle et al. 2013). There was a strong correlation between the reported values and those of this study (R=0.96), with CER[N(24)S(18)] at highest concentration (Table 3.6.1).

Table 3.6.1: Comparison of plasma lipid concentrations reported here with Bellis et al. (2014)
Reported measures from Bellis et al. (2014) are shown (Paper Mean, Paper SD; standard deviation), alongside the cohort values found in this study (Cohort). The column "Ceramide" uses the nomenclature used in this study, while "Reported Lipid" is the species reported in Tabassum et al. (2019).

Reported Lipid	Reported (pmol/ml) Mean	Reported (pmol/ml) SD	Cohort (pmol/ml) Mean	Cohort (pmol/ml) SD	Ceramide
dhCer 22:0	145.59	66.53	20.85	15.10	CER[N(22)DS(18)]
dhCer 24:0	207.86	98.63	175.23	112.21	CER[N(24)DS(18)]
Cer 20:0	149.35	53	18.32	12.49	CER[N(20)S(18)]
Cer 22:0	1002.37	336.57	128.48	74.25	CER[N(22)S(18)]
Cer 24:0	2771.72	855.11	2724.18	1289.51	CER[N(24)S(18)]

In conclusion, there is a strong relationship between the unadjusted CER values measured in the cohort studied in this project and those published ($R>0.75$). The incorporation of a more advanced lipidomics pipeline with SPE would have provided similar concentrations of CER species as those species that have been published. The strong relationships found provide confidence that while the plasma CER species were found at lower concentration in this study, the pattern remains the same, which is required for the genetic analyses. The differences between the published studies are likely due to the concentrations being estimated without the availability of species-specific commercial standards.

3.2.1.4.2 Plasma *N*-acyl ethanolamine concentrations in literature

Jones et al. (2014) analysed NAE in plasma samples of 36 participants. The study was split into three groups, so it is assumed the control group consisted of 12 participants. No summary statistics were provided, so the values presented in Table 3.7.1 were estimated from a figure in the publication. The values were similar to that of the cohort plasma analyses. The values are presented in pg/ml due to the quantification of the plasma NAE species using species-specific calibration lines.

Table 3.7.1: Comparison of plasma lipid concentrations reported here with Jones et al. (2014)
Reported measures from Jones et al. (2014) are shown (Reported Mean), the conversion to pg/ml (Calculated Mean), alongside the cohort values found in this study (Cohort Mean).

NAE	Reported (ng/ml) Mean	Calculated (pg/ml) Mean	Cohort (pg/ml) Mean
PEA	2	2000	1886.27
OEA	2	2000	571.93
LEA	1	1000	619.23
AEA	0.8	800	351.39
DHEA	0.6	600	350.24

Joosten et al. (2010) analysed the plasma NAE species of 22 fasting women. The concentrations reported were much higher in concentration than the study NAE concentration likely due to the addition of phenylmethanesulphonyl fluoride to the plasma samples to inactivate fatty acid amide hydrolase (*FAAH*), which degrades NAE species in the body and is the genetic loci we identify in this study to influence plasma NAE species concentrations (Table 3.8.1).

Table 3.8.1: Comparison of plasma lipid concentrations reported here with Joosten et al. (2010)
Reported measures from Joosten et al. (2010) are shown (Reported), the conversion to pg/ml (Calculated), alongside the cohort values found in this study (Cohort).

NAE	Reported (nmol/L) Mean	Reported (nmol/L) SEM	Calculated (pg/ml) Mean	Calculated (pg/ml) SEM	Cohort (pg/ml) Mean	Cohort (pg/ml) SEM
AEA	6.8	0.7	2363	243	351	10.95
OEA	43.8	3.3	14257	1074	572	17.54
PEA	40	3.7	11900	1101	1886	43.39
STEA	16.3	1.9	5338	622	497	14.66

Fanelli et al. (2018) analysed the plasma NAE species of 184 premenopausal women. The resulting plasma NAE concentrations were similar to that of the cohort NAE concentrations, but at higher concentration (Table 3.9.1). This may be due to the use of toluene for extraction, which they compared to other extraction reagents and found it had the best recovery (Fanelli et al. 2012).

Table 3.9.1: Comparison of plasma lipid concentrations reported here with Fanelli et al. (2018)
Reported measures from Fanelli et al. (2018) are shown (Reported), the conversion to pg/ml (Calculated), alongside the cohort values found in this study (Cohort). Mean and standard deviation (SD) are presented.

NAE	Reported (pmol/ml) Mean	Reported (pmol/ml) SD	Calculated (pg/ml) Mean	Calculated (pg/ml) SD	Cohort (pg/ml) Mean	Cohort (pg/ml) SD
AEA	1.13	0.37	392.675	128.575	351.39	333.45
PEA	16.4	3.8	4879	1130.5	1886.27	1359.26
OEA	5.24	1.39	1705.62	452.445	571.93	536.21

No concentrations were provided for the lipid-genetic analysis of 1,054 individuals that identified an association with OEA (Long et al. 2017). 48 healthy participants were analysed in another genetic study of NAE species (Sipe et al. 2010). Lipid levels were reported as mean and 95% confidence interval (lower, L; upper, U). The authors calculated pmol/ml for each lipid based on a ratio to the deuterated internal standards added to each sample and therefore did not use a species-specific calibration line, so

the values for plasma NAE species are much higher than those published in other studies (Table 3.10.1).

Table 3.10.1: Comparison of plasma lipid concentrations reported here with Sipe et al. (2010)
Reported measures from Sipe et al. (2010) are shown (Reported), the conversion to pg/ml (Calculated), alongside the cohort values found in this study (Cohort).

NAE	Reported (pmol/ml) Mean	Reported (pmol/ml) LCI	Reported (pmol/ml) UCI	Calculated (pg/ml) Mean	Calculated (pg/ml) LCI	Calculated (pg/ml) UCI	Cohort (pg/ml) Mean	Cohort (pg/ml) SD
AEA	12.8	11.4	14.5	4448	3961.5	5038.75	351.39	333.45
DHEA	9.2	8.3	10.1	3418.72	3084.28	3753.16	350.24	290.33
PEA	213.3	190.1	239.3	63456.75	56554.75	71191.75	1886.27	1359.26
STEA	80.1	70.5	92.7	26232.75	23088.75	30359.25	496.64	447.72
OEA	107.9	97.2	119.8	35121.45	31638.6	38994.9	571.93	536.21
LEA	38.9	34.9	43.4	12584.15	11290.15	14039.9	619.23	509.73

In conclusion for NAE species, the studies published that measure NAE species in plasma are all of very small sample size (n<184). While the studies differ in comparison of the NAE levels, the use of commercial internal standards and calibration lines in this project provides confidence in the reported measures. There is not a standard approach to extraction and analysis of NAE species and therefore the creation of a collaboration of multiple investigators studying these species would aid this analysis.

3.2.1.4.3 Plasma eicosanoids, octadecanoids, and docosanoids ("Eicos") concentrations in literature

Quehenberger et al. (2010) also analysed the plasma concentration of Eico species. The plasma was pooled from 22L of 100 fasting individuals in lithium heparin. The values are in a similar range to that found in this study, with the Eico species found in this study at increased concentration other than the HETE species that we show are increased by inflammation such as coagulation and clotting of blood (Table 3.11.1).

Table 3.11.1: Comparison of plasma lipid concentrations reported here with Quehenberger et al. (2010)
Reported measures from Quehenberger et al. (2010) are shown (Reported), the conversion to pg/ml (Calculated), alongside the cohort values found in this study (Cohort).

Lipid	Reported (ng/dl) Mean	Reported (ng/dl) SEM	Calculated (pg/ml) Mean	Calculated (pg/ml) SEM	Cohort (pg/ml) Mean	Cohort (pg/ml) SD
9-HODE	201	50	2010	500	4054	3469
13-HODE	314	38	3140	380	7317	6631
13-OxoODE	14.4	1.2	144	12	488	305
9-HOTrE	14.8	1.5	148	15	259	268
13-HOTrE	14.5	0.5	145	5	394	354
5-HETE	381	45	3810	450	131	134
12-HETE	135	9	1350	90	117	82
15-HETE	25.6	0.7	256	7	111	93
12,13-DiHOME	158	6	1580	60	3596	2872
19,20-DiHDPA	44.6	14.1	446	141	846	429

The article by Bowden et al. (2017) which analysed the same plasma as that analysed in the Quehenberger et al. (2010) paper, described three Eico species only, due to "only six laboratories provided eicosanoid concentrations (two laboratories were not able to measure any eicosanoids in the reference material)… In total, 143 eicosanoids were measured by at least one laboratory; however, only three (5-HETE, 12-HETE, and 15-HETE) were measured by at least five laboratories." The three species that were identified in the analysis were at increased concentration (Table 3.12.1), which

in this study and other publications have shown are the HETE lipids that increase with blood clotting and coagulation, such as in serum samples. This may be due to the many litres of plasma being blended together.

Table 3.12.1: Comparison of plasma lipid concentrations reported here with Bowden et al. (2017)
Reported measures from Bowden et al. (2017) are shown (Reported), the conversion to pg/ml (Calculated), alongside the cohort values found in this study (Cohort).

Lipid	Reported (pmol/ml) Median of Means	Reported (pmol/ml) Standard uncertainty	Calculated (pg/ml) Median of Means	Calculated (pg/ml) Standard uncertainty	Cohort (pg/ml) Mean	Cohort (pg/ml) SD
5-HETE	10	1.3	3205	416.65	131.45	133.96
12-HETE	6.8	1.5	2179.4	480.75	116.54	81.56
15-HETE	2.4	0.64	769.2	205.12	110.62	92.51

Miller et al. (2018) measured Eico in plasma from healthy control subjects before treatment (the number of participants was not described). The values were increased for the 12-HETE species, potentially due to the plasma sample preparation (Table 3.13.1).

Table 3.13.1: Comparison of plasma lipid concentrations reported here with Miller et al. (2018)
Reported measures from Miller et al. (2018) are shown (Reported), the conversion to pg/ml (Calculated), alongside the cohort values found in this study (Cohort).

Lipid	Reported (ng/ml) Mean	Reported (ng/ml) SD	Calculated (pg/ml) Mean	Calculated (pg/ml) SD	Cohort (pg/ml) Mean	Cohort (pg/ml) SD
9,10-DiHOME	2.01	1.88	2010	1880	2995.49	2856.31
12-HETE	8.08	4.2	8080	4200	116.54	81.56
15-HETE	0.18	0.08	180	80	110.62	92.51
11,12-DHET	0.18	0.05	180	50	135.41	87.46
14,15-DHET	0.28	0.09	280	90	137.63	83.70

In conclusion for the Eico species, analysis of the published literature has shown that the cohort plasma has lower levels of HETE species than those published. We show in this project alongside other publications that increased inflammation can raise HETE species, such as the clotting process in serum samples, so this may point to plasma sample preparation differences between the studies. Otherwise, the rest of the bioactive lipids that were included in the comparisons were similar to the cohort samples, likely due to the availability of species-specific commercial standards for these lipids.

3.2.1.5 Presentation of adjusted values used in the genetic analyses

The means and standard deviations reported for the concentration of the lipid species is before adjustment for covariates (Chapter 4 and Chapter 5), which includes accountable variability (i.e. batch effects) as well as unmeasured clinical variability (diet, individual variability). The standardised values (mean of 0) used for the genetic analyses incorporated adjustments for significant covariates such as batch effects and removed extreme outliers. With the mean of 0, the standard deviations were found at a range of 0.41-1.02 (Table 3.14.1). It is likely that the large standard deviations noted for the values of lipid concentration without adjustment for covariates is due to the presence of batch effects and the inclusion of outliers.

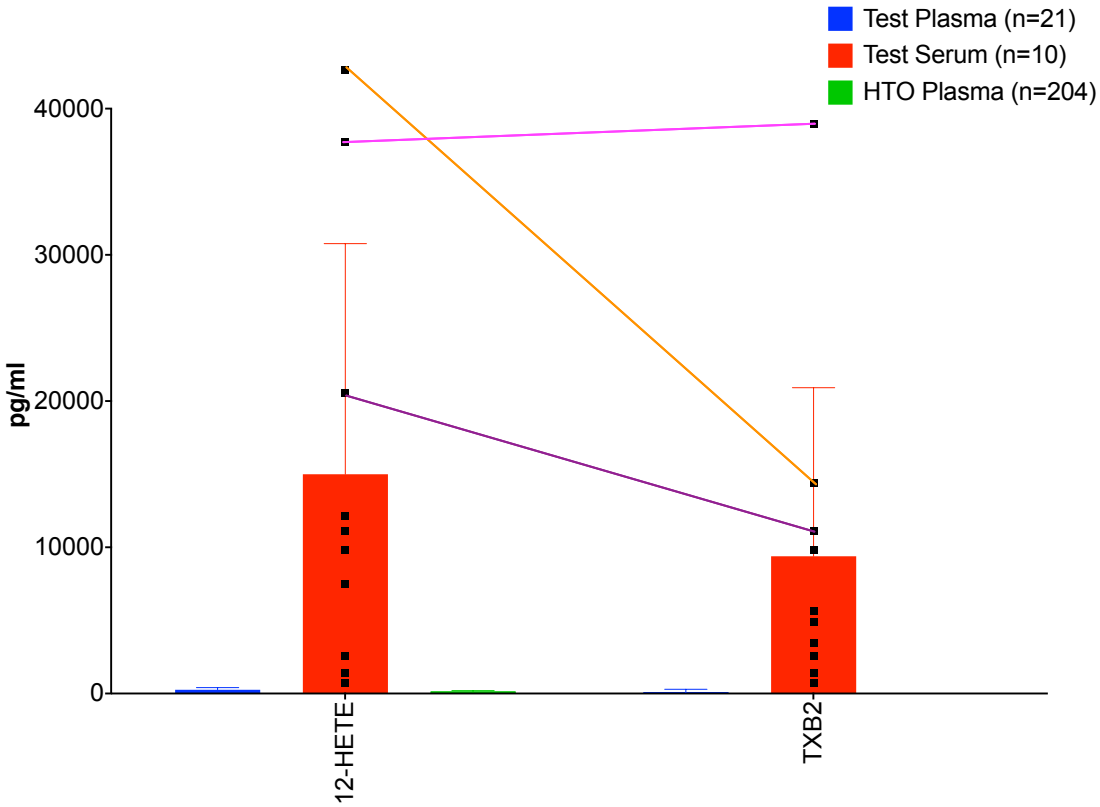


Figure 3-1: Levels of TXB_2 and 12-HETE are higher in serum than plasma samples

Multiple t-testing was used to identify lipid species that were substantially different between the serum and plasma samples. HTO; Hypertension Oxford family cohort analysed in this study. Data is shown as mean (pg/ml) and SD, with each value indicated by a black square. Coloured lines highlight three extreme participants in the serum samples.

Table 3.14.1: Standard deviation for each lipid species after adjustments and outlier removal

Depicted is the standard deviation used in the genetic analyses for each lipid species. The mean was standardised to zero after adjustments and outliers were removed.

Lipid	SD
AEA	0.82
DHEA	0.90
DPEA	0.96
HEA	0.95
LEA	0.95
OEA	0.89
POEA	0.78
PEA	0.88
PDEA	0.97
STEA	0.94
VEA	0.84
A22_S18	0.78
A24_S18	0.41
A26_S18	0.86
C18_DS	0.93
C18_S	0.66
C18_S1P	0.92
N16_S18	0.84
N20_S18	0.87
N22_DS18	0.81
N22_S18	0.89
N22_S19	0.87
N23_S18	0.79
N23_S20	0.94
N24_DS18	0.84
N24_DS19	0.91
N24_DS20	0.93
N24_S16	0.86
N24_S17	0.90
N24_S18	0.85
N24_S19	0.84
N24_S20	0.97
N24_S22	0.89
N25_DS18	0.90
N25_S20	0.94
N26_DS18	0.93
N26_S18	0.97
N26_S19	0.88
N27_S18	0.90

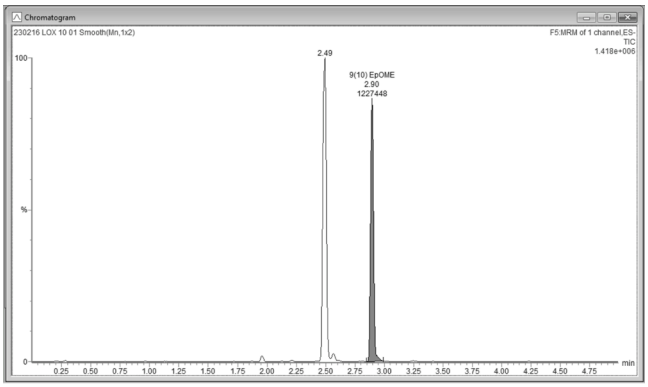
N28_S18	0.93
N29_S18	0.84
DHET1112	0.85
DHET1415	0.96
DiHDPA1920	0.94
DiHOME1213	0.71
DiHOME910	0.78
EpOME1213	0.78
EpOME910	0.84
HDHA4	0.75
HETE11	0.85
HETE12	0.86
HETE15	0.87
HETE5	0.68
HODE13	0.72
HODE9	0.85
HOTrE13	0.85
HOTrE9	0.88
OxoODE13	1.02
OxoODE9	0.91
TransEKODE	0.92

Table 3.0: Standard deviation for each lipid species after adjustments and outlier removal
Depicted is the standard deviation used in the genetic analyses for each lipid species. The mean was standardised to zero after adjustments and outliers were removed.

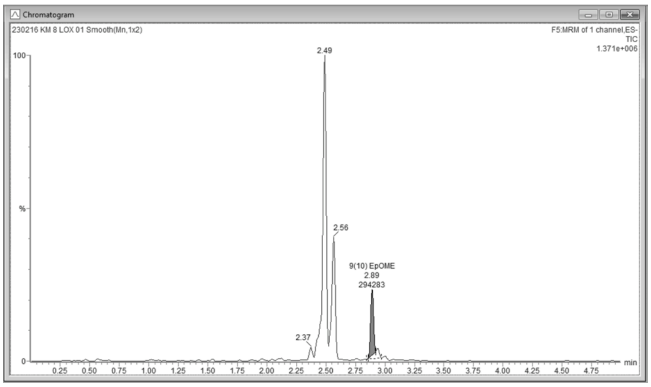
3.2.2 Assessments of injection variability

The variability of the lipidomic analysis was assessed for each lipid species. The response of the identified peaks (area normalised by internal standard) from the same sample by three injections of the LC-MS/MS was assessed. Any variability in this analysis is likely due to either the lipidomics pipeline or further processing steps. While no threshold is standard, a coefficient of variation under 20% is recommend by Food and Drug Authority guidelines for precision (FDA, 2001) and variation under 30% is deemed acceptable for publication (Checa *et al.*, 2015). All lipid species were found less than 20% variable other than Eico species 9,10-EpOME (22%), CER[A(26)S(18)] (23%) and CER[N(27)DS(18)] (30%; this species is removed from analysis in the CER MRM confirmation step 3.2.3.3). The peaks for both 9,10-EpOME commercial standard and plasma CER[A(26)S(18)] were masked by peaks (potential impurities) that likely lead to the increased analytical variability (Figure 3-2).

A



B



C

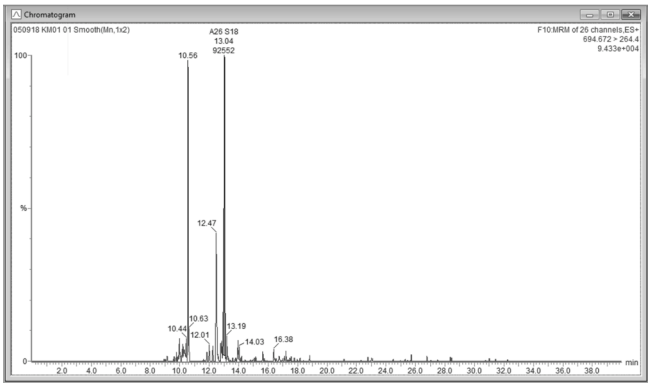


Figure 3-2: LC-ESI-MS/MS reconstructed chromatograms of A) 9,10-EpOME commercial standard B) plasma 9,10-EpOME C) plasma CER[A(26)S(18)]

The figure depicts the chromatograms used to quantify the lipid species that were identified as more variable in comparison to the rest of the measured lipids. The peaks selected for the species are annotated. Both of the chromatograms for 9,10-EpOME (A, B) and CER[A(26)S(18)] (C) were inhibited by foreign, prominent, presenting peaks, which likely affected their assessment of variability by masking their identification.

3.2.3 Recovery, efficiency, and matrix effects

3.2.3.1 Eico species

As the commercial synthetic standards were available for all species of the Eico class, process efficiency, extraction recovery and mass spectrometry matrix effects, were established for each lipid species of the class. As described in Chapter 2, process efficiency compared the identified mass spectrometry peak created from commercial standards for each species in ethanol, to the peaks created when the standards undergo lipid extraction from plasma. Process efficiency identifies losses or gains of species throughout the entire assay, such as; the impact of matrix (plasma) effects on the extraction of the lipids, the chromatography, and detection by the mass spectrometer (Matuszewski *et al.*, 2003).

Of the Eico species, some were identified with a process efficiency below 50% (12,13-EpOME, 9,10-EpOME, 9-HOTrE), therefore there was loss of the lipid that was not recoverable during the process. Nine species were found with a process efficiency above 100%, as expected (5-HETE, 12-HETE, 13-HODE, 9-OxoODE, 13-HOTrE, 13-OxoODE, 12,13-DiHOME, 9,10-DiHOME, 19,20-DiHDPA) as the species were detected at concentrations above that of the concentration of synthetic standard added, due to the addition of the plasma concentrations of the lipids to the synthetic standard's abundance.

Extraction recovery is assessed by lipid extraction of synthetic standards in plasma and compared with the results when standards are added at the end of the extraction process (i.e. not extracted from plasma). This allows for the calculation of the estimated loss of the lipid species through the extraction procedure. 9,10-EpOME, 12,13-EpOME, and 9-HOTrE were found at less than 20% of their original levels and therefore these species had major loss resulting from the extraction process, which is comparative to their respective low process efficiency results.

Matrix effect is defined as the effect of the tissue being assessed (plasma) on the measurement of an analyte. A matrix effect value of less than 100 indicates ion suppression; the reduced ionisation efficiency for an analyte due to the presence of foreign analytes in a tissue sample matrix which compete for ionisation. Nine of the

Eico lipids had matrix effects over 100 as the concentration of the lipids in the plasma added to the synthetic standard signal. However, the DHET species presented with evidence of ion suppression (80-90%), where the species were affected by the presence of other ions in the sample.

Blank ethanol samples were run in between the QC samples to assess carryover effects. The Eico species were found at minimal levels in the blank ethanol samples (<3%); there was very little carryover in the analysis. All 19 Eico lipids were taken forward for genetic analyses (Table 3.1).

Table 3.1: Quality control assessment of plasma Eico species.

The table depicts the QC results and mean injection coefficient of the analysis of variation of three samples for all Eico species. The percentage process efficiency, extraction recovery, and the effect of plasma as a matrix on lipidomic mass spectrometry, were assessed for the Eico as commercial standards were available. The carryover, the proportion of the sample identified in a blank ethanol sample, is described for all lipid species.

Lipid	Mean Injection CV (%)	Process Efficiency (%)	Extraction Recovery (%)	Matrix Effects (%)	Proportion in blank (%)
4-HDHA	9	58	41	142	0
14,15-DHET	6	58	64	90	0
11,12-DHET	8	52	65	80	0
12-HETE	4	102	76	134	0
5-HETE	20	100	71	141	0
11-HETE	6	71	51	138	0
15-HETE	8	66	43	156	0
9-HODE	7	83	33	253	3
13-HODE	6	145	41	356	0
9-OxoODE	10	187	80	235	2
9-HOTrE	7	22	17	130	0
13-HOTrE	3	107	58	183	0
13-OxoODE	6	144	86	169	1
9,10-EpOME	22	27	19	143	0
12,13-EpOME	14	20	18	113	0
TransEKODE	7	50	36	138	0
12,13-DiHOME	3	301	82	368	0
9,10-DiHOME	5	291	74	395	0
19,20-DiHDPA	6	223	94	236	0

3.2.3.2 NAE species

As the commercial synthetic standards were also available for all NAE species, process efficiency, extraction recovery, and matrix effects were established for each lipid species of the class (Table 3.2). The process efficiency ranged from 61% for POEA to 124% for PEA, which is expected as PEA and STEA (104%) are the species at highest concentration in plasma. The extraction recovery estimated for the NAE species ranged from 96% - 118%, with STEA an outlier at 160%. Matrix effects showed suppression (<100%) for all NAE species, other than high concentration PEA, which highlights that the analysis in plasma had an effect on the lipid measurements. The NAE species were found at minimal levels in the blank ethanol samples (<1%); there was very little carryover in the analysis. All 11 NAE species were taken forward for genetic analyses.

Table 3.2: Quality control assessment of plasma NAE species

The table depicts the mean injection coefficient of variation over three samples for all species. The percentage process efficiency, extraction recovery, and the effect of plasma as a matrix on lipidomic mass spectrometry was assessed for the NAE lipids as commercial standards were available. The carryover, the proportion of the sample identified in a blank ethanol sample is described for all lipid species. VEA wasn't included in the analysis as it uses the OEA commercial standard for analysis.

Lipid	Mean Injection CV (%)	Process Efficiency (%)	Extraction Recovery (%)	Matrix Effects (%)	Proportion in blank (%)
POEA	3	61	101	61	0
PDEA	2	93	105	89	0
PEA	4	124	114	109	1
HEA	4	90	114	79	0
STEA	15	104	160	65	0
OEA*	6	103	112	92	0
LEA	3	88	102	86	0
AEA	4	71	96	74	0
DPEA	2	92	115	80	0
DHEA	6	99	118	84	0

3.2.3.3 CER species

Commercial standards are not available for each species of CER, therefore further assessments were required to confirm the identity of the plasma CER species by an additional analysis of multiple reaction monitoring (MRM), completed through literature-based and class-based calculated mass spectrometry transitions. This was completed for three product ions per CER species (except for C18S1P where two were used). The presence of a peak for each of the species in all three transitions confirmed that the peak identified is the lipid of interest. Nine CER species were excluded from analysis at this step (Table 3.3). An example of the analysis of the chromatographic traces is depicted in Figure 3-3.

On assessment of carryover, the CER species were found at minimal levels in the blank ethanol samples (<8%), other than CER[N(16)DS(18)] (18%) and CER[N(18)S(18)] (45%) which were excluded from analyses at the MRM confirmation step (3.2.3.3), likely to be foreign plasma metabolites. 30 CER were taken forward for genetic analyses.

Table 3.3: Quality control assessment of plasma CER species.

The table depicts the mean injection coefficient of variation over three samples for all species. The carryover, the proportion of the sample identified in a blank ethanol sample is described for all lipid species. The CER and related sphingolipid species underwent MRM analysis to identify those lipids where a peak was identified in at the correct retention time in three (✓), two (2), or 1 (X) chromatograms.

Lipid	Mean injection CV (%)	Proportion in blank (%)	MRM Analyses
CER[A(22)S(18)]	6	0	✓
CER[A(24)S(18)]	2	0	✓
CER[A(26)S(18)]	23	0	✓
CER[N(14)S(18)]	10	0	X
CER[N(16)DS(18)]	8	18	X
CER[N(16)S(18)]	4	0	✓
CER[N(18)DS(18)]	6	0	X
CER[N(18)DS(24)]	3	0	X
CER[N(18)DS(26)]	10	0	X
CER[N(18)S(18)]	8	45	X
CER[N(20)DS(24)]	8	2	X
CER[N(20)S(18)]	9	8	✓
CER[N(22)DS(18)]	3	0	✓
CER[N(22)S(18)]	3	3	✓
CER[N(22)S(19)]	11	0	✓
CER[N(23)S(18)]	5	0	✓
CER[N(23)S(20)]	3	0	✓
CER[N(24)DS(18)]	2	0	✓
CER[N(24)DS(19)]	6	0	✓
CER[N(24)DS(20)]	7	2	✓
CER[N(24)S(16)]	5	3	✓
CER[N(24)S(17)]	5	0	✓
CER[N(24)S(18)]	6	0	✓
CER[N(24)S(19)]	7	0	✓
CER[N(24)S(20)]	6	0	✓
CER[N(24)S(22)]	9	0	✓
CER[N(25)DS(18)]	5	0	✓
CER[N(25)S(20)]	7	0	✓
CER[N(26)DS(18)]	5	0	✓
CER[N(26)S(18)]	3	0	✓
CER[N(26)S(19)]	4	0	✓
CER[N(27)DS(18)]	30	0	X
CER[N(27)S(18)]	7	0	✓
CER[N(28)DS(18)]	14	0	X
CER[N(28)S(18)]	4	0	✓
CER[N(29)S(18)]	6	0	✓

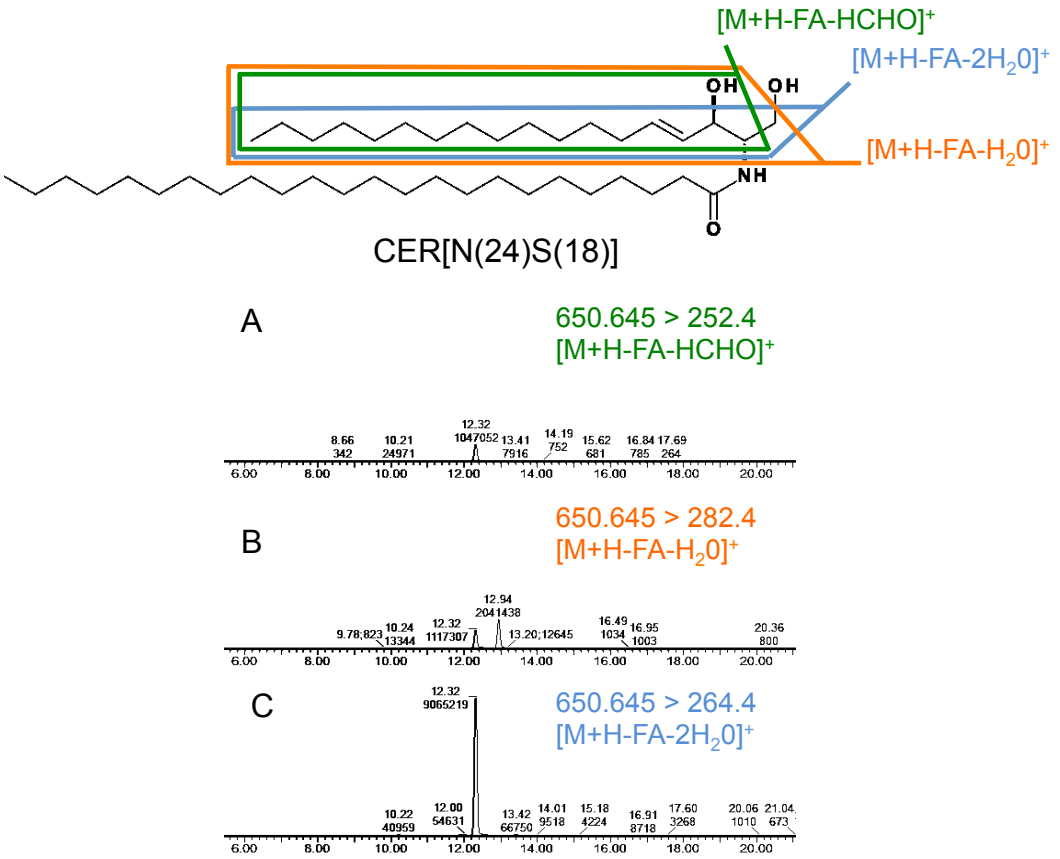


Figure 3-3: LC-ESI-MS/MS reconstructed chromatograms for CER[N(24)S(18)], which was detected by multiple reaction monitoring (MRM) of the following transitions A) 650 > 252 B) 650 > 282 C) 650 > 264.

The figure depicts the assessment of multiple reaction monitoring of an exemplar CER species. The precursor mass was 650.645 which was targeted in the first quadrupole of the triple quadrupole mass spectrometer. Fragment ions of 252.4, 282.4, and 264.4 were assessed for presence of a peak at the same retention time in all three transitions. The SRM analysis used to analyse the cohort plasma uses the transition 264.4, which has the greatest peak. The figure depicts the fragmentation patterns of the lipid; $[M+H-FA-2H_20]^+$ describes the mass of the lipid (M) calculated by summation of the mass of carbon, oxygen, hydrogen, and nitrogen atoms, the addition of a hydrogen/proton during the electrospray ionisation process (+H), the loss of the fatty acid chain during fragmentation in the second quadrupole (-FA), and the loss of two water molecules ($-2H_20$) to create the particular positively charged ($^+$) lipid fragment that is depicted as the peak. The respective coloured boxes depicting the fragmentation patterns of the lipid chemical structure are for visualisation only, and may not be the exact atom that is lost.

3.2.4 Comparison of cohort plasma samples to pooled quality control samples

Sixty lipids (n=1016 samples analysed for CER and NAE, n=204 analysed for Eico) were compared to replicate pooled plasma samples, created from plasma of fasting, unrelated individuals that was collected and stored in 2009. 24 pooled samples were ran alongside the CER and NAE analysis, and four pooled samples were analysed with the Eico samples). Comparison by assessment of the percentage difference in means was undertaken. Only six lipid species showed greater than 30% difference in their mean abundance compared to the pooled plasma samples; three CER species, four NAE species, and three Eico species (Table 3.3.1.1).

The results of the final genetic analysis show that one of the NAE species (LEA) that was identified to differ more than 30% in samples' mean to that of the pooled plasma samples' mean, associated with the genetic locus of the NAE degradation enzyme (FAAH). This may be an example of a lipid species showing different to that of the pooled plasma samples due to an underlying genetic influence in their measurement, identifiable in this family-based cohort.

However, the final genetic analyses also did not find GWAS significance for CER species with a sphingosine base (-S18). Here, such CER species are decreased in the cohort plasma samples compared to the pooled plasma, which may indicate their instability in plasma during long-term storage.

Overall, unique species differed in each class compared to the pooled plasma samples without a definitive pattern. This difference between the sample means could be due to sample storage or collection differences, the effect of the physical mixing of plasma in the creation of the pooled plasma samples, or the ascertainment for the cohort samples; non-fasting, family-based samples. All lipids were taken forward for genetic analyses as only 10 were found to be substantially different to that of the pooled plasma samples, of which there is the potential that such species are genetically-influenced in the family-based cohort.

Table 3.3.1.1: Comparison of pooled plasma samples to cohort plasma
The percentage change was calculated for the mean concentration of each lipid species between the cohort plasma samples and the repeatedly measured, pooled plasma samples. 10 lipid species were found to have a mean greater than 30% different to that of the QC samples (highlighted in yellow). 1016 samples were analysed for CER (pmol/mL of plasma) and NAE species (pg/mL of plasma); 204 samples were analysed for Eico species (pg/ml plasma).

Lipid	Sample mean	Sample SD	QC mean	QC SD	% mean change
A22_S18	1.64	0.74	1.84	0.55	11
A24_S18	2.91	0.50	3.19	0.30	9
A26_S18	0.11	0.09	0.13	0.05	17
C18_DS	0.28	0.18	0.38	0.22	27
C18_S	2.05	1.93	5.46	4.09	62
C18_S1P	3.95	4.59	5.01	5.50	21
N16_S18	1.59	1.24	1.76	0.94	10
N20_S18	0.40	0.27	0.60	0.27	33
N22_DS18	0.52	0.38	0.68	0.28	23
N22_S18	5.63	3.25	6.23	2.30	10
N22_S19	1.06	0.73	0.94	0.35	-13
N23_S18	40.63	17.14	46.00	13.42	12
N23_S20	1.96	0.62	1.96	0.18	0
N24_DS18	7.85	5.03	10.62	3.38	26
N24_DS19	2.67	1.69	2.61	0.57	-2
N24_DS20	1.42	0.78	1.99	0.50	29
N24_S16	1.86	1.15	2.23	0.82	17
N24_S17	10.73	4.81	11.34	3.16	5
N24_S18	128.87	61.00	155.00	54.40	17
N24_S19	50.23	22.86	47.22	9.49	-6
N24_S20	11.35	4.24	11.34	1.74	0
N24_S22	1.69	1.16	1.92	0.72	12
N25_DS18	1.07	0.55	1.45	0.33	27
N25_S20	1.34	0.67	1.80	0.55	26
N26_DS18	0.77	0.41	2.08	0.57	63
N26_S18	33.05	10.26	43.96	9.80	25
N26_S19	4.64	3.18	5.29	1.53	12
N27_S18	2.16	1.56	2.67	0.82	19
N28_S18	0.88	0.56	1.21	0.43	27
N29_S18	1.21	1.40	1.27	0.45	4
AEA	351.39	333.45	355.83	241.16	1
DHEA	350.24	290.33	316.44	210.38	-11
DPEA	21.54	17.38	19.30	13.73	-12
HEA	23.97	18.50	16.34	11.62	-47

LEA	619.23	509.73	439.03	290.12	-41
OEA	571.93	536.21	658.25	662.55	13
PEA	1886.27	1359.26	1538.21	953.60	-23
POEA	42.85	55.60	73.87	55.48	42
PDEA	34.55	26.17	26.00	21.76	-33
STEA	496.64	447.72	522.06	571.81	5
VEA	254.75	266.28	316.75	318.80	20
DHET1112	135.41	87.46	138.04	36.04	2
HETE11	47.47	27.03	54.11	6.96	12
EpOME1213	2242.46	2175.75	1975.86	358.96	-13
DiHOME1213	3595.88	2872.39	2955.95	568.64	-22
HETE12	116.54	81.56	172.25	28.82	32
HODE13	7316.95	6630.78	6768.62	712.58	-8
HOTrE13	393.54	354.11	315.54	84.99	-25
OxoODE13	488.20	304.85	487.50	49.97	0
DHET1415	137.63	83.70	145.25	34.37	5
HETE15	110.62	92.51	103.88	25.34	-6
DiHDPA1920	846.07	429.10	684.87	219.88	-24
HDHA4	130.74	91.98	118.26	22.75	-11
HETE5	131.45	133.96	114.96	23.10	-14
EpOME910	642.53	706.08	675.52	297.69	5
DiHOME910	2995.49	2856.31	2075.87	1142.79	-44
HODE9	4053.73	3468.59	4720.52	486.53	14
HOTrE9	258.62	268.45	175.43	95.34	-47
OxoODE9	761.85	469.79	986.88	165.52	23
TransEKODE	139.45	107.06	166.50	180.77	16

Table 3.3.1.1: Comparison of pooled plasma samples to cohort plasma
The percentage change was calculated for the mean concentration of each lipid species between the cohort plasma samples and the repeatedly measured, pooled plasma samples. 10 lipid species were found to have a mean greater than 30% different to that of the QC samples (highlighted in yellow). 1016 samples were analysed for CER (pmol/mL of plasma) and NAE species (pg/mL of plasma); 204 samples were analysed for Eico species (pg/ml plasma).

3.2.5 Comparison of cohort plasma concentrations of NAE and CER lipids in range finding study and full cohort analysis

The mean concentrations before adjustment for cofounders and outlier removal were significantly different between the range-finding study (n=204) and the 812 samples analysed afterwards for the full cohort analysis (Table 3.3.1.2). The analysis of a different number of samples, as well as the number of batches (n=4 for range finding study; n=20 for the remaining samples) likely caused this difference. This is highlighted by the strong correlation (R=0.94) between the percentage decrease in lipids found for the cohort samples and the pooled (QC) plasma samples. Batch effects are therefore an important factor to consider in lipidomics analysis, and if not corrected for, may provide incorrect results.

Table 3.3.1.2: Comparison of the mean lipid concentration in the range-finding study and further samples

The mean concentration of lipids for the CER (pmol/ml) and NAE (pg/ml) found for the range-finding study was compared to the mean abundances found for the further 812 samples analysed for the full cohort. Welch Two-Sample T-test was used to assess the significance of the difference of means and the P-value was adjusted for 40 tests using a Bonferroni correction. A strong correlation (R=0.94) was identified between the % decrease found for the cohort samples and the pooled plasma samples (QC), likely highlighting a batch effect.

Lipid	Range finding study; mean n=204)	Further 800 samples; mean n=812)	% decrease from range-finding study in mean lipid concentration	P value (adjusted)	% decrease in QC samples from range-finding study
A22_S18	2.39	1.46	-65	8.80E-15	-46
A24_S18	3.34	2.81	-19	8.80E-15	-7
A26_S18	0.10	0.12	18	NS	11
C18_DS	0.00	0.35	99	8.80E-15	99
C18_S	0.07	2.54	97	8.80E-15	98
C18_S1P	0.29	4.83	94	8.80E-15	98
N16_S18	1.51	1.61	6	NS	13
N20_S18	0.44	0.39	-13	NS	-8
N22_DS18	0.80	0.45	-77	8.80E-15	-41
N22_S18	6.87	5.31	-29	5.82E-07	-15
N22_S19	1.52	0.94	-61	8.80E-15	-66
N23_S18	57.17	36.47	-57	8.80E-15	-44
N23_S20	2.30	1.87	-23	7.42E-12	-6
N24_DS18	12.49	6.69	-87	8.80E-15	-65
N24_DS19	3.52	2.45	-44	2.27E-08	-26
N24_DS20	1.64	1.37	-20	5.69E-03	4
N24_S16	2.42	1.72	-41	5.60E-09	-38
N24_S17	14.85	9.69	-53	8.80E-15	-46

N24_S18	190.40	113.41	-68	8.80E-15	-51
N24_S19	65.79	46.32	-42	8.80E-15	-42
N24_S20	11.71	11.25	-4	NS	10
N24_S22	1.37	1.76	22	5.34E-06	43
N25_DS18	1.34	1.00	-34	1.94E-07	-20
N25_S20	1.07	1.41	24	8.80E-15	32
N26_DS18	0.81	0.76	-6	NS	15
N26_S18	30.71	33.64	9	1.55E-03	20
N26_S19	3.91	4.82	19	5.80E-05	20
N27_S18	1.58	2.30	31	2.17E-12	29
N28_S18	0.61	0.95	35	8.80E-15	40
N29_S18	0.83	1.31	36	7.31E-06	39
AEA	573.18	295.60	-94	8.80E-15	-71
DHEA	491.41	314.64	-56	8.80E-15	-60
DPEA	29.51	19.51	-51	8.80E-15	-74
HEA	26.63	23.28	-14	8.80E-15	-44
LEA	1035.53	514.64	-101	8.80E-15	-115
OEA	1283.39	393.19	-226	8.80E-15	-385
PEA	2888.98	1634.36	-77	8.80E-15	-66
POEA	43.39	42.70	-2	NS	10
PDEA	39.81	33.16	-20	1.33E-03	9
STEA	1039.60	360.23	-189	8.80E-15	-412
VEA	600.42	167.91	-258	8.80E-15	-366

3.3 Calculation of extra lipidomic measures

As many lipidomic species in the same biochemical pathways were measured, summation and ratio traits were created in the aim of identifying further GWAS associations with SNPs influencing the genes of proteins involved in the metabolic pathways of the lipid species measured. The lipid species included in the calculation of each trait are shown in Table 3.4.

3.3.1 Eico species

Summations were created based on PUFA substrate; four independent traits were calculated from the sum of all species measured that are created from the independent PUFAs; LA, AA, ALA, DHA. Summations of the measured lipid species were also calculated based on n-3 (omega-3) and n-6 (omega-6) UFA lipid status. Enzymatic activity-based traits were created from the sum of lipid mediators created via reactions with specific enzymes; 15-LOX, 5-LOX, CYP450, EPHX2, and LOX1, as well as product/precursor ratios to isolate SNPs in the EPHX2 enzyme (DIHAMS) which creates DiHOME species from EpOME species. A further product/precursor ratio was calculated to assess if there were any identifiable associations with SNPs in the conversion of HODE to OxoODE lipid mediators, which is an oxidation reaction.

3.3.2 NAE species

The NAE species are produced and degraded via the same enzymatic pathways, differing only by fatty ethanolamine substrate, therefore the only trait that could be calculated was the total abundance of all measured lipids (sumNAE).

3.3.3 CER species

Many lipid mediators at various steps of the sphingolipid biosynthetic pathway were measured, therefore many traits could be calculated; summations of the species based on class (i.e. CER[NS], CER[NDS], CER[AS], C18 sphingosine species), groups based on the carbon number of the fatty acid or sphingoid chain (e.g. n22sum), individual product/precursor ratios of CER[NS] to CER[NDS], as well as CER[NDS]/C18DS, C18S/CER[NS] and CER[NS]/C18S, C18S1P/C18S, CER[AS]/C18DS, and summations of total abundance (e.g. all sphingolipids, all

CER[NS], all CER[AS], etc.), were calculated. Literature-based calculations without obvious biological meaning were also assessed for an underlying genetic component; those that have been used to identify CER[NS] as biomarkers of disease (Laaksonen *et al.*, 2016), which included ratios of low-carbon fatty acid to higher carbon fatty acid CER species, e.g. N(16)S(18)/N(24)S(18), and a trait for the sum of all high concentration CER species studied as biomarkers (biomcers) was assessed.

Table 3.4: Calculations of further lipid-based traits

Description of the lipid mediators included in the calculations of class-based, enzymatic activity-based, substrate-based abundance summations, product/precursor ratios and further literature biomarker ratios that were assessed for underlying genetic components in this study. Class, lipid class; Trait, trait name; Species, species included in the summed abundance to create the trait (Sum of), or divided in ratios (Ratio of). The figures in Chapter 1 provide a visual aid behind the calculation of the traits.

Class	Trait	Species
Eico	LA	Sum of 13-HODE, 13-OxoODE, 9-HODE, 9-OxoODE, 12,13-EpOME, 1213-DiHOME, 9,10-EpOME, 9,10-DiHOME, TransEKODE
Eico	AA	Sum of 15-HETE, 11,12-DHET, 14,15-DHET, 12-HETE, 5-HETE, 11-HETE
Eico	αLA	Sum of 9-HOTrE, 13-HOTrE
Eico	DHA	Sum of 4-HDHA, 19,20-DIHDPA
Eico	15-LOX	Sum of 13-HOTrE, 15-HETE, 13-HODE, 13-OxoODE
Eico	5-LOX	Sum of 9-HOTrE, 4-HDHA, 5-HETE, 9-HODE, 9-OxoODE
Eico	CYP	Sum of 9,10-EpOME, 9,10-DiHOME, 12,13-EpOME, 12,13-DiHOME, 19,20-DiHDPA, 11,12-DHET, 14,15-DHET
Eico	EPHX2	Sum of 19,20-DiHDPA, 11,12-DHET, 14,15-DHET, 9,10-DiHOME, 12,13-DiHOME
Eico	LOX1	Sum of 9-HODE, 9-OxoODE
Eico	13-OXLAMS	Ratio of 13-OxoODE/13-HODE
Eico	9-OXLAMS	Ratio of 9-OxoODE/9-HODE
Eico	9-DIHAMS	Ratio of 9,10-DiHOME/9,10-EpOME
Eico	13-DIHAMS	Ratio of 12,13-DiHOME/12,13-EpOME
Eico	n-6 PUFA	Sum of 15-HETE, 11,12-DHET, 14,15-DHET, 9,10-EpOME, 12,13-EpOME, 9,10-DiHOME, 12,13-DiHOME, 9-HODE, 9-OxoODE, 13-HODE, 13-OxoODE, TransEKODE, 5-HETE, 12-HETE, 11-HETE
Eico	n-3 PUFA	Sum of 13-HOTrE, 9-HOTrE, 4-HDHA, 19,20-DiHDPA
Eico	sumEicos	Sum of all Eico array species
NAE	sumNAE	Sum of all 11 N-acyl ethanolamine species measured
CER	ratio16to24	Ratio of CER[N(16)S(18)]/ CER[N(24)S(18)]
CER	ratio22to24	Ratio of CER[N(22)S(18)]/ CER[N(24)S(18)]
CER	ratio20to24	Ratio of CER[N(20)S(18)]/ CER[N(24)S(18)]
CER	biomcers	Sum of CER[N(16)S(18)], CER[N(20)S(18)], CER[N(22)S(18)], CER[N(23)S(18)], CER[N(24)S(18)]
CER	ns_sum	Sum of all CER[NS] species
CER	nds_sum	Sum of all CER[NDS] species
CER	s18_sum	Sum of all species with a sphingosine backbone
CER	N_s18sum	Sum of all CER[NS] species with a sphingosine backbone
CER	allxs18	Sum of all species with a 18-carbon backbone (e.g. incl CER[NDS])
CER	s19_sum	Sum of all species with a 19-carbon sphingosine backbone
CER	alls19	Sum of all species with a 19-carbon backbone (e.g. incl CER[NDS])
CER	s20_sum	Sum of all species with a 20-carbon sphingosine backbone

CER	alls20	Sum of all species with a 20-carbon backbone (e.g. incl CER[NDS])
CER	ds18_sum	Sum of all CER[NDS] species with a sphingosine backbone
CER	n22_sum	Sum of all species with a 22-carbon fatty acid
CER	n23_sum	Sum of all species with a 23-carbon fatty acid
CER	n24_sum	Sum of all species with a 24-carbon fatty acid
CER	n25_sum	Sum of all species with a 25-carbon fatty acid
CER	n26_sum	Sum of all species with a 26-carbon fatty acid
CER	n22ratio	Ratio of CER[N(22)S(18)]/CER[N(22)DS(18)]
CER	n24ratio	Ratio of CER[N(24)S(18)]/ CER[N(24)DS(18)]
CER	n24s19ratio	Ratio of CER[N(24)S(19)]/ CER[N(24)DS(19)]
CER	n24s20ratio	Ratio of CER[N(24)S(20)]/ CER[N(24)DS(20)]
CER	n26ratio	Ratio of CER[N(26)S(18)]/ CER[N(26)DS(18)]
CER	c18s1psratio	Ratio of C18S1P/C18S
CER	ndssumc18dsratio	Ratio of CER[NDS]/C18DS
CER	totalsphingo	Sum of all 30 sphingolipid species
CER	assum	Sum of all 3 CER[AS] species
CER	assumc18dsratio	Ratio of CER[AS]/C18DS
CER	nssumndssumratio	Ratio of CER[NS]/CER[NDS]
CER	nssum_c18sratio	Ratio of CER[NS]/C18S
CER	c18s_nssumratio	Ratio of C18S/CER[NS]
CER	ns18sumc18sratio	Ratio of CER[NS] with a sphingosine backbone/C18S
CER	c18sns18sumratio	Ratio of C18S/CER[NS] with a sphingosine backbone
CER	c18sc18s1pratio	Ratio of C18S/C18S1P
CER	sumn22s	Sum of all CER[N(22)S(X)]
CER	sumn24s	Sum of all CER[N(24)S(X)]
CER	sumn24ds	Sum of all CER[N(24)DS(X)]
CER	sumn26s	Sum of all CER[N(26)S(X)]
CER	sumcer	Sum of all CER[NS] and CER[NDS] species
CER	sumc18	Sum of all C18 species

Description of the lipid mediators included in the calculations of class-based, enzymatic activity-based, substrate-based abundance summations, product/precursor ratios and further literature biomarker ratios that were assessed for underlying genetic components in this study. Class, lipid class; Trait, trait name; Species, species included in the summed abundance to create the trait (Sum of), or divided in ratios (Ratio of). The figures in Chapter 1 provide a visual aid behind the calculation of the traits.

3.4 Detection and removal of outliers

In the lipidomic analysis of the full cohort, 999 plasma samples were analysed over two years. Thus, there was a high probability of the presence of a few extreme values that would not represent the main distribution of the cohort. The aim of this section was to identify the most extreme outliers, without substantially losing samples or future statistical significance resulting from the use of too-stringent thresholds (i.e. excluding lipid measurements outside three standard deviations from the mean). A test for outliers of the multiple linear regression model for each lipid species was assessed by testing each observation as a mean-shift outlier based on studentized residuals (Figure 3-4).

Studentized residuals are calculated for an observation by dividing the residual by an estimate of the standard deviation (computed with the observation excluded). Residuals reflect the scale of measurement, which can vary with predictors, therefore studentized residuals are more effective at detecting outliers. Mean-shifting uses the studentized residuals calculated for every value, and shifts the points to the mean studentized residual, replacing the value by the mean; the larger the movement identified, the more likely the value is an outlier. A t test is conducted for each residual and extreme values are identified as outliers. Cook's distance is used to identify influential residuals; how much the model coefficient estimates change if an observation were removed from the data set. An outlier with a large Cook's distance value means the value is influential and the removal of such an outlier would cease its influence on the model. Such high leverage observations are those that have extremely high or low values for the predictor variable relative to the other values [described as in www.jmp.com/].

An example of the impact of the removal of such outliers on genetic studies was assessed using the exemplar lipid species palmitoyl ethanolamide (PEA), a NAE species at high concentration. Six extreme values of PEA were significantly identified via the pipeline described above ($P<0.05$). Assessment of heritability showed that when the outliers were not removed (n=999), the SNP-based heritability of PEA was 49% ($P= 6.11x10^{-16}$). The estimate of heritability increased when the six outliers were removed (n=993); h^2_{SNP} = 53% ($P= 6.11x10^{-16}$). Assessments of preliminary GWAS

using the genotyping data showed the association of the lead genotyped SNP in the NAE degradation enzyme fatty acid amide hydrolase (*FAAH*; rs324420) was not GWAS significant ($P<1x10^{-8}$) when the six outliers remained, however, the SNP reached GWAS significance ($P=5.14x10^{-9}$), alongside another SNP in *FAAH* (rs1571138; $P=1.10x10^{-8}$), when the six outliers were removed. Therefore, for the species with an established genetic association, this provides evidence that including outliers, which likely reflect experimental errors, depowers the study.

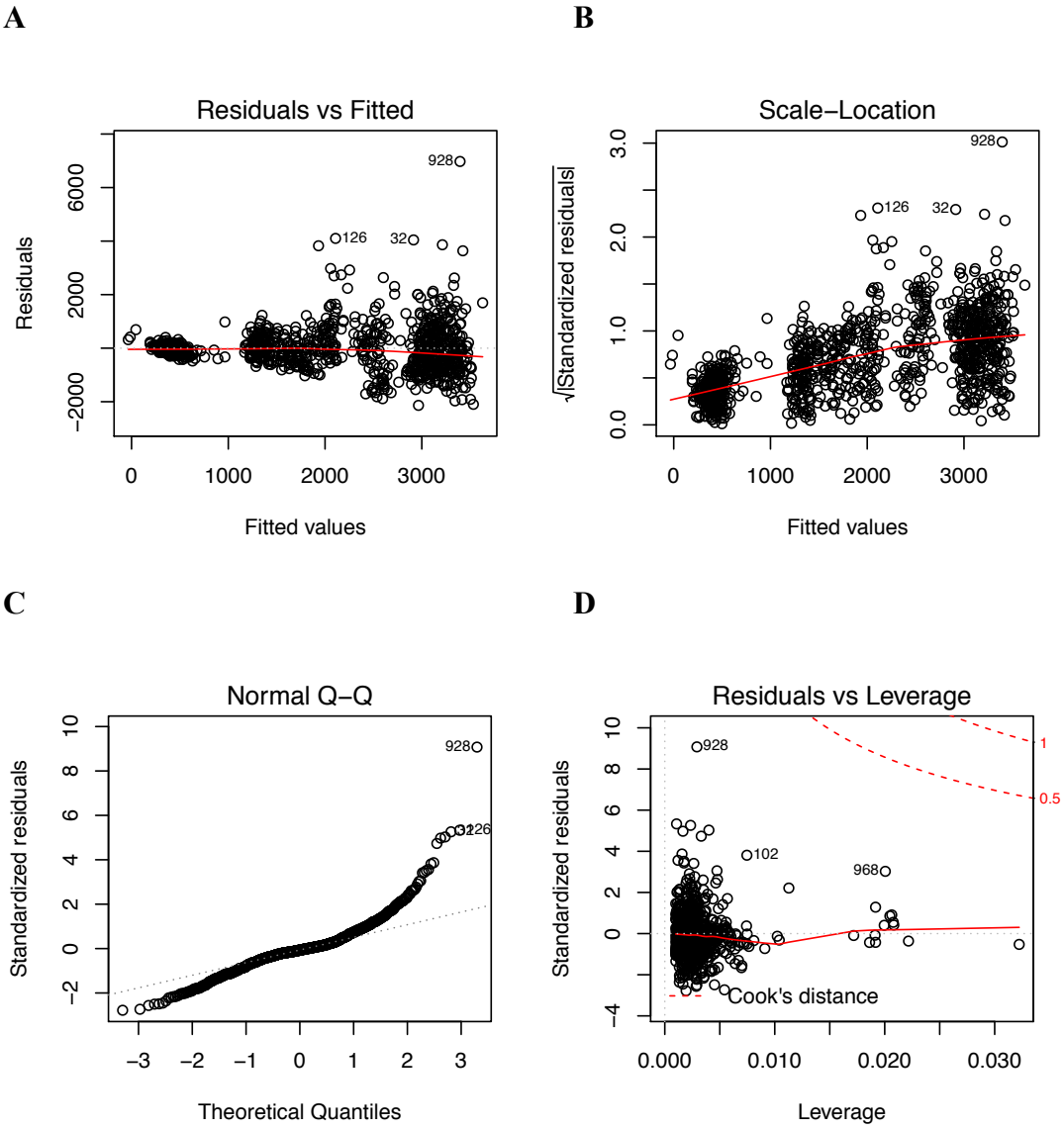


Figure 3-4: Identification of PEA outlier values by testing them as mean-shift outliers.

A) Residuals compared to fitted values. B) The creation of standardised residuals.C) standardised residuals over expected quantiles, identifying outlier values. D) assessment of the leverage of the confirmed outliers; the outliers identified for this lipid are not high leverage observations nor have an estimated large cook's distance, but the large estimated skew of the distribution of PEA values due to the outlier values requires removal of the outliers.

3.5 Assessment of the most appropriate GWAS software for family-based association studies

To identify the most appropriate GWAS software for family-based GWAS studies, a trait for urate levels was used as a technical replicate and positive control throughout the GWAS software assessments. GWAS analyses of urate have been described previously in literature, identifying a region on chromosome 4 at the urate transporter, the *SLC2A9* gene (Döring *et al.*, 2008). The urate phenotype was collected for this cohort (n=1,110), and GWAS analysis using the cohort have also identified the chromosome 4 associations.

Two family-based GWAS software that take into account family structure were assessed; GCTA (Yang *et al.*, 2010; Zaitlen *et al.*, 2013) and FaST-LMM (Lippert *et al.*, 2011), with a comparison software, PLINK1.9 (Purcell *et al.*, 2007; Chang *et al.*, 2015), used to show the effect of non-adjustment for population structure. Test GWAS were undertaken using the urate trait in PLINK, using the --assoc command to perform an asymptotic version of Students t-test to compare two means. FaST-LMM (FLMM) specifying the –ML command for maximum likelihood parameter learning, and GCTA specifying mixed linear model association analyses (--mlma) (Figure 3-5) were undertaken to adjust for the family-based population substructure.

The Manhattan plots resulting from all three analyses show the stack of SNPs expected at chromosome 4. The top SNP for GCTA and FLMM was rs13129697, in the intron of *SLC249,* with P-values of 5.3×10^{-11} and 3.1×10^{-11}, respectively. The SNP associated at GWAS using PLINK too; it was the 3rd most significant SNP and associated at GWAS to a more significant P-value of 3.0×10^{-12}. The SNP is associated with hypouricemia on OMIM (entry # 612076), associated with urate, gout, and uric acid in GWAS Catalog, and gout, joint disorders, and inflammatory polyarthropathies in the UKBiobank PheWAS assessment on Gene Atlas, confirming the SNP's role in influencing urate.

The significance observed for the lead urate SNP is inflated due to the relatedness of the cohort when the analysis was undertaken using PLINK without adjustment for population structure. This is depicted by the respective Quantile-Quantile plot created from the association results and the genomic inflation factor (GIF) of 1.306,

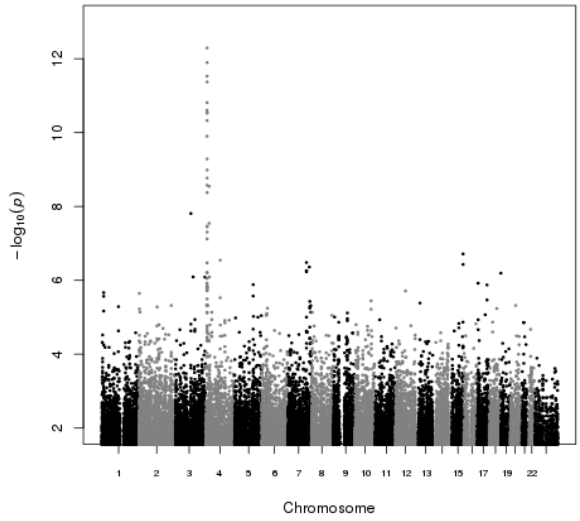
highlighting this inflation (a GIF of 1.000 describes a perfect analysis with no altered population substructure (Devlin *et al.*, 1999)). Therefore, while PLINK provided the most significant result, the results are incorrect and thus inflated, when the family structure of the cohort was unaccounted for.

Both FaST-LMM and GCTA adjusted for the relatedness using linear mixed modelling approaches to deal with population substructure. The resulting GIFs were calculated to 1.004 and 1.001, respectively, which are similar and both acceptable, nearing 1.000 (Figure 3-5). Assessing the speed of each analysis software showed that PLINK ran swiftly to create an association result (<1 minute), GCTA took 3 minutes to analyse the data, and FaST-LMM took 2 hours and required high memory nodes (32GB RAM; University of Manchester computer clusters).

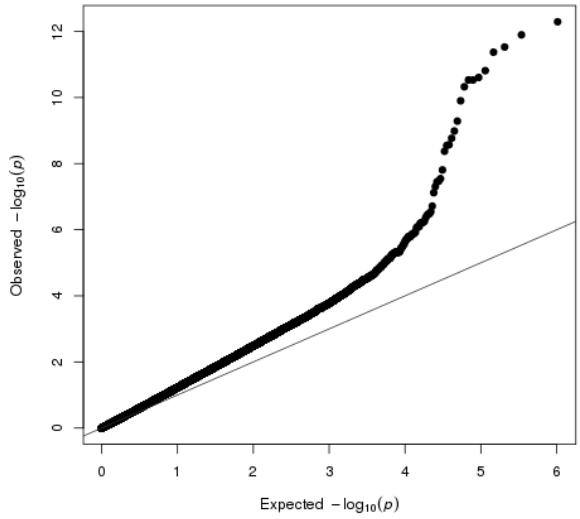
GCTA was chosen as the most appropriate software for lipidomic GWAS analyses due to its accuracy in taking into account family structure, with the added benefit of rapid analyses. It should be noted that GCTA does not analyse sex chromosomes (--autosome), so chromosomes 1-22 were analysed only. Should an X-chromosome or the Y-chromosomes (in the case of studies with a substantial number of males with similar haplotypes) be of interest, a different software would be required.

PLINK software

A



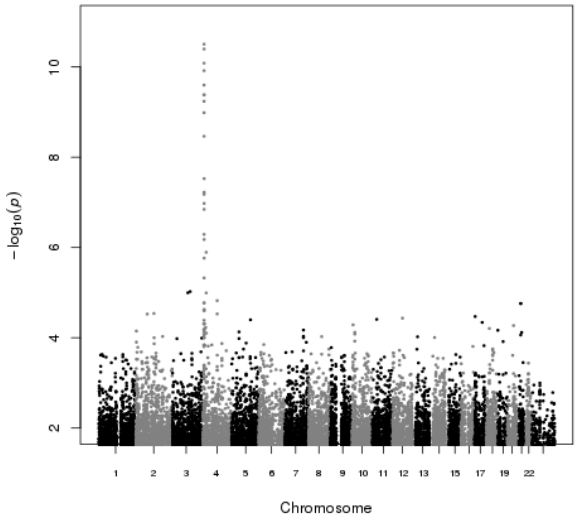
B



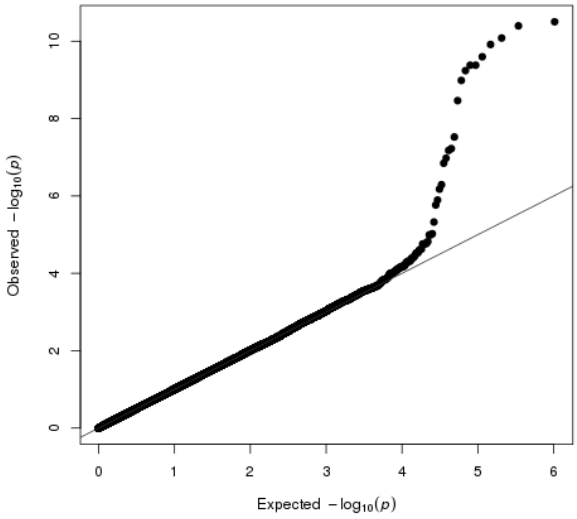
A) Manhattan plot and B) Quantile-Quantile (QQ) plot for urate GWAS using three software; PLINK, FaST-LMM, and GCTA. D) A comparison of the P-values of association (-log10) of GWAS of urate using FaST-LMM (FLMM; Y-axis) and GCTA software (X-axis).

FaST-LMM software

A



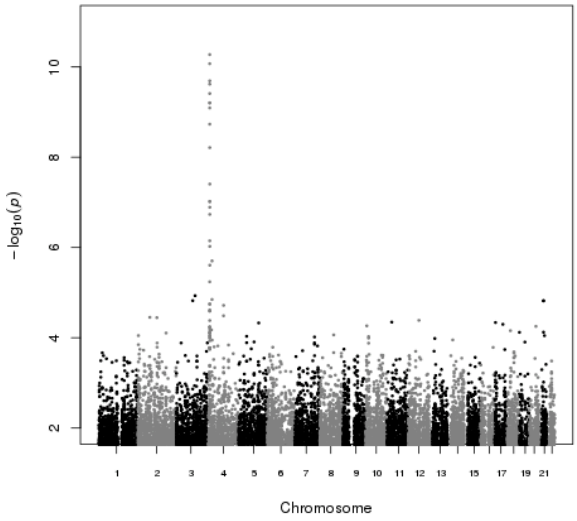
B



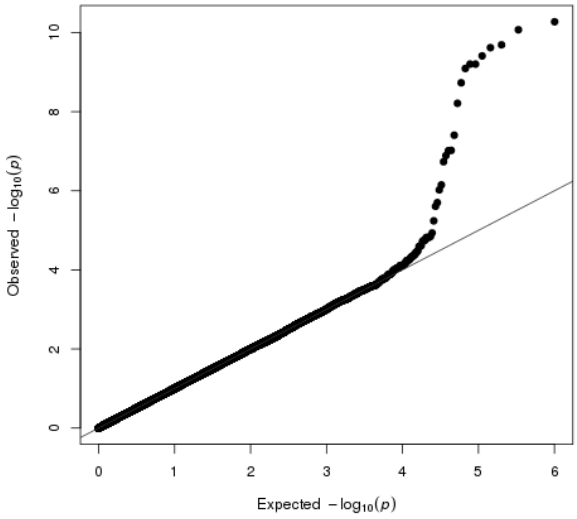
A) Manhattan plot and B) Quantile-Quantile (QQ) plot for urate GWAS using three software; PLINK, FaST-LMM, and GCTA. D) A comparison of the P-values of association (-log10) of GWAS of urate using FaST-LMM (FLMM; Y-axis) and GCTA software (X-axis).

GCTA software

A



B



A) Manhattan plot and B) Quantile-Quantile (QQ) plot for urate GWAS using three software; PLINK, FaST-LMM, and GCTA. D) A comparison of the P-values of association (-log10) of GWAS of urate using FaST-LMM (FLMM; Y-axis) and GCTA software (X-axis).

D

FaST-LMM and GCTA P-value comparison

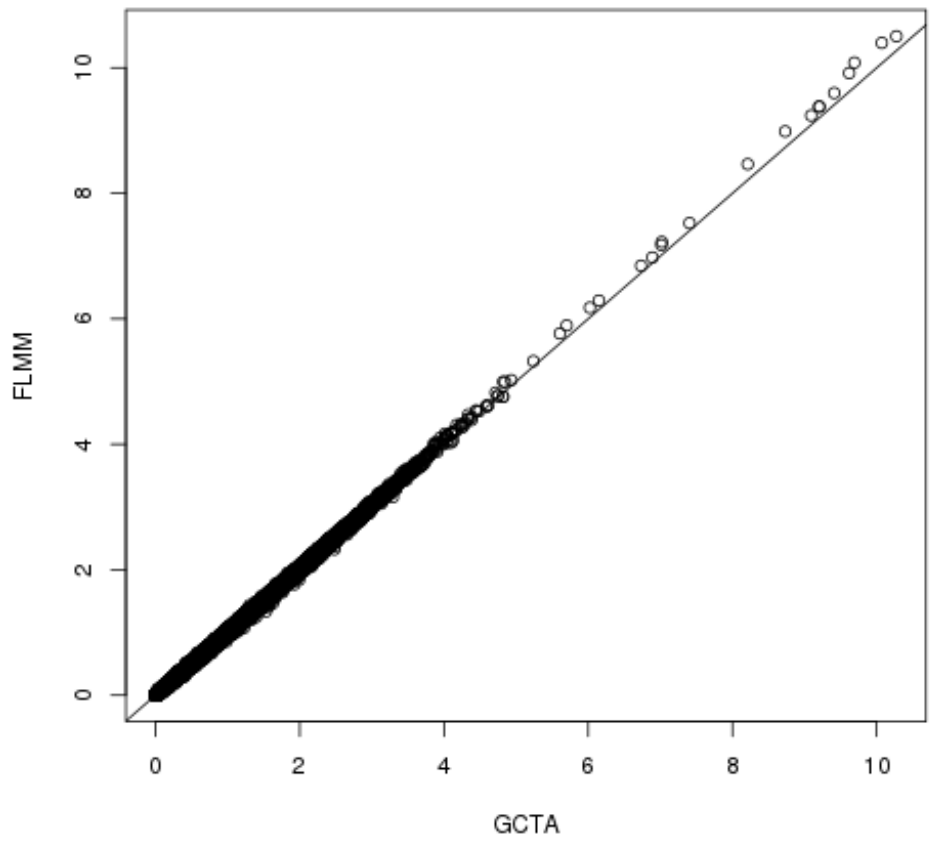


Figure 3-5: Assessment of software for family-based GWAS analyses using the trait urate.

A) Manhattan plot and B) Quantile-Quantile (QQ) plot for urate GWAS using three software; PLINK, FaST-LMM, and GCTA. D) A comparison of the P-values of association (-log10) of GWAS of urate using FaST-LMM (FLMM; Y-axis) and GCTA software (X-axis).

3.6 Assessment of the quality control thresholds for GWAS analyses using imputed data

An assessment of the effect of different quality control thresholds was undertaken to identify the paramount thresholds to use in the analysis of the HRC-imputed genotyping data of this project. A preliminary GWAS was completed for the exemplar lipid vaccenic acid ethanolamide (VEA), a species from the NAE class. This was initially completed using the genotyping data only (0.5M SNPs) that had undergone quality control using a minor allele frequency (MAF) threshold of 0.01, where the minor allele of each SNP must be found in at least 1% of the cohort. SNPs were identified on chromosome 1 at the gene for the NAE degradation enzyme, fatty acid amide hydrolase (*FAAH*; described fully in Chapter 5). The genome inflation factor (GIF) for this GWAS was 0.996. The comparisons described in this section are depicted in Figure 3-6.

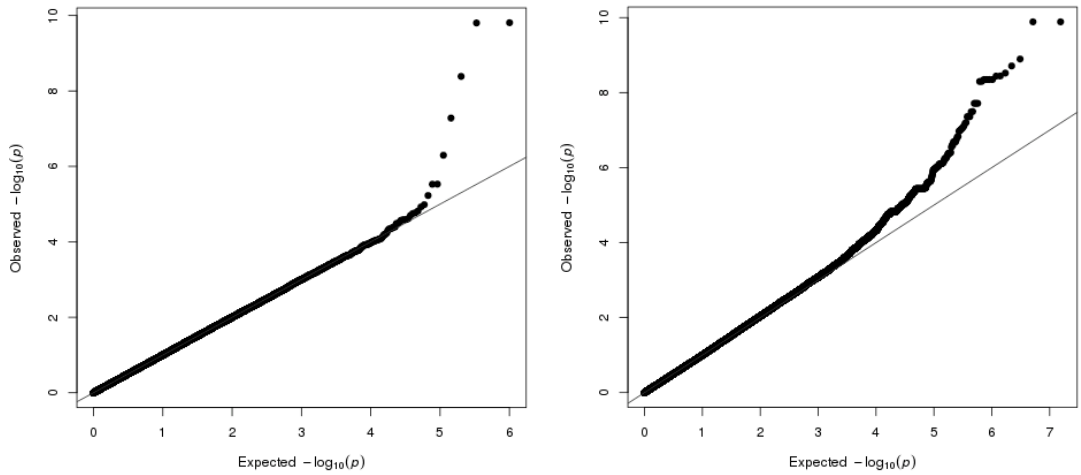
R^2 score, the squared correlation between imputed genotypes and the true, observed, input genotypes, depicts the quality of imputation of each SNP and is a measure of the confidence in imputed dosages (McCarthy *et al.*, 2016). As depicted in Chapter 2, the mean R^2 was 0.33 across the chromosomes of the cohort DNA, which was also the point of the inflexion between the high levels of SNPs with low and more intermediate R^2 scores, which ranged from 0.0 to 0.9. Assessment of the literature provided varying hypothesis about how to choose an R^2 threshold; both the mean score (Gardner *et al.*, 2018) and the inflexion point (Coleman *et al.*, 2016) were suggested thresholds.

The second analysis was completed using the same quality control thresholds as the genotyping data (MAF>0.01), for the imputed data, with the use of the R^2 threshold of 0.30 due to it being the R^2 mean score and inflexion point, and 7.5M SNPs remained. However, this lead to the production of a Quantile-Quantile plot with an increased substructure present due to a number of less reliable imputed SNPs included in the analysis. The GIF for this association study was 0.982.

To avoid false positive results, a more stringent cut off of R^2>0.80 was assessed as a third analysis, and a lower number of SNPs were included in the association study (5M SNPs). This association analysis resulted in a more preferable GIF of 0.985.

However, it is advised to use more stringent MAF thresholds in the analysis of small-moderate cohort association studies (n<10,000), as spurious findings can result with the inclusion of low frequency SNPs in common SNP assessments through GWAS (Marees *et al.*, 2018). Therefore, a more stringent MAF threshold of 0.05 was included to ensure the results were robust. Such threshold requires 50 of the 999 individuals in the cohort to have the minor allele. This resulted in an association study using imputed data that represented that of the genotyping data; a GIF of 0.992, and the creation of a improved Quantile-Quantile plot (Figure 3-6), but allowing for the analysis of further areas of the genome (5M SNPs).

A: genotyped; MAF>0.01 **B: imputed; R^2>0.30; MAF>0.01**



C: imputed; R^2>0.80; MAF>0.01 **D: imputed; R^2>0.80; MAF>0.05**

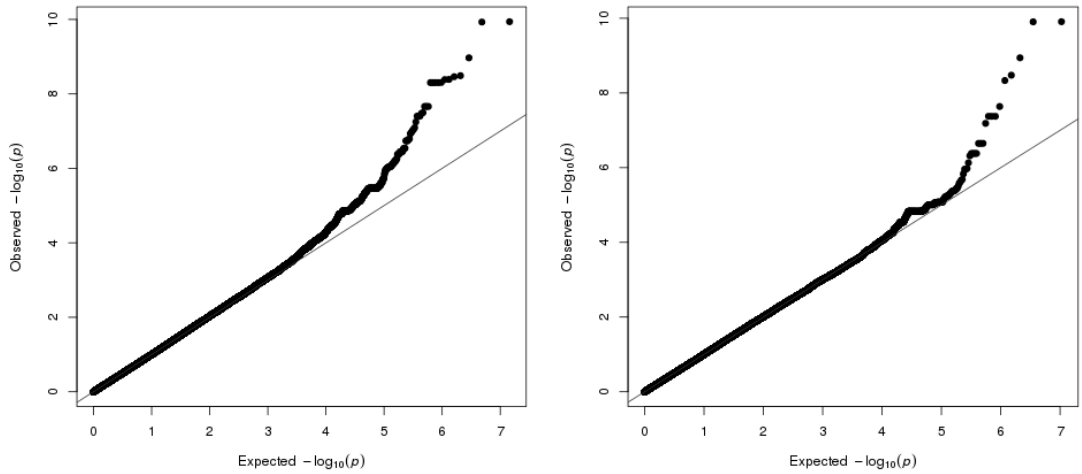


Figure 3-6: Quantile-Quantile plots of the GWAS results showing the effect of minor allele frequency and imputation quality thresholds

Four GWAS analyses were assessed using the NAE lipid, vaccenic acid ethanolamide (VEA); A) genotyping data consisting of 0.5M SNPs with MAF>0.01 threshold B) imputed data consisting of 7.5M SNPs with R^2>0.3 and MAF>0.01 thresholds C) imputed data consisting of 5.0M SNPs with R^2>0.8 and MAF>0.01 D) imputed data consisting of 5.0M SNPs with R^2>0.8 and MAF>0.01. MAF, minor allele frequency; R^2, squared correlation between imputed and true genotypes.

3.7 Discussion

Stringent quality control standards were applied to the detected lipids found in plasma, and a subset of the most robust lipids were identified. Additional traits of interest for inclusion in the genetic analyses were calculated. Extreme outliers were identified and removed. GCTA was identified as the most appropriate family-based GWAS software, and robust quality control thresholds were used to obtain reliable GWAS results using imputed genotyping data.

3.7.1 A greater proportion of detectable NAE and CER species were identified in the cohort plasma compared to the Eico species

Of the species measured in plasma, 23% (19 of 83 species in the assay) of the species of the Eico assay passed quality control assessments (HODE, OxoODE, TransEKODE, EpOME, DiHOME, HETE, DHET, HOTrE, HDHA, and DiHDPA species), 38% of the NAE species passed quality control (11 of 29 species in the assay), and 57% of the CER species passed quality control (CER[NS], CER[NDS], CER[AS], and C18S species; 30 of 53 species in the assay). The NAE and CER classes included more measured lipids species in the cohort plasma samples. This is likely due to the Eico class found at lowest concentration in plasma (Wong *et al.*, 2014), and are therefore likely at the limit of detection of the mass spectrometer.

The Eico species 12-HETE and TXB_2 were found substantially increased in serum compared to plasma, and TXB_2 was identified at negligible levels in the cohort plasma. This stimulated increase in TXB_2 via the blood clotting process would need to be assessed in serum samples to measure and discover the genetic influence over such circulating lipids.

While the commercial standards displayed clearly resolved peaks, any issues, such as lower resolution (peaks joined together), presence of unknown peaks, or small peaks, likely impacted the ability to identify the low concentration lipid species in plasma, due to the tissue being a complex matrix of many endogenous compounds.

Upon completion of the project, class-specific CER deuterated internal standards became available for CER[NS], CER[NDS], CER[AS], CER[AH], and CER[ADS]. The analyses of these deuterated internal standards confirmed the MRM-identified

retention times for the respective species and the quality of the assay for the lipids retained for genetic analysis.

3.7.2 Efficiency of the lipid extraction protocol

Extraction problems can be due to poor solubility of certain compounds in the extraction solvent, solvent saturation effects, analyte co-precipitation with proteins, and poor chromatographic separation due to co-elution with endogenous compounds. However, Solid Phase Extraction (SPE) and Liquid–Liquid Extraction (LLE) used here for the respective lipidomic analyses are considered the most efficient methods for sample preparation (Tsakelidou *et al.*, 2017). The goal of extraction is to obtain quantitative yields of metabolites in the specimen (Lu *et al.*, 2017). The plasma extraction process used here showed a low process efficiency for certain lipids, which is usual for assays of multiple species.

Specifically, the NAE species STEA increased during extraction, which could be due to the addition of artefacts, impurities of solvents, or glassware during extraction (Tsakelidou *et al.*, 2017). For the case of certain lipids, such as the Eico species 9,10-EpOME, 12,13-EpOME, and 9-HOTrE, which showed losses during extraction, any assay variability contributing to their levels in plasma may deflate their estimated heritability (Visscher *et al.*, 2008). Only four Eico species were found to have a recovery greater than 80% in the lipidomics pipeline. These bioactive, signalling lipids are likely not as stable as other more high concentration lipids, such as the analysis of cholesterol lipids. They may be less stable therefore during the lipid extraction process from plasma, or do not extract as well from plasma as the higher concentration plasma lipids. This may have influenced the reported concentrations, but as the lipids were analysed in the same way for all samples, it likely does not add variability to the genetic analyses, which do not depend on quantification.

The addition of known concentrations of the deuterated internal standards, and the CER[N(25)S(18)], C17S, and C17DS internal standards, to each sample before extraction and analysis, allowed for correction of any variation across samples. Furthermore, the results obtained from the simultaneous analysis of lipids from the pooled quality control samples, allowed for specific adjustment across each lipid

measurement, in each batch of samples. This accounted for any losses from extraction, ionization, or processing, in a species-specific manner.

Process efficiency (PE) compared the integral of the analytical peak generated from commercial lipid standards, when lipid solutions are prepared in ethanol, to the integral of the peaks generated when the lipid standards undergo solvent extraction after they have been added to plasma. Process efficiency identifies losses or gains of the amount of each lipid throughout the entire assay, and reflects the impact of matrix (plasma), effect of the extraction of the lipids, chromatography, and detection by the mass spectrometer (Matuszewski et al., 2003). When lipid species are found at a PE of greater than 100%, this suggests that the detected concentrations correspond to the concentration of the synthetic standard and the concentrations of endogenous lipids already present in the biological sample. Estimating PE may point to the quality of the lipid extraction, but is not an accurate estimate, due to the endogenous species in a plasma sample inflating those lipids under investigation, and the quality of the estimate for lipids that are not as well analysed from the samples is deflated due to the addition of such endogenous species.

The process efficiency is a reflection of the extraction recovery. For those species identified with low recovery during extraction it is unlikely that a low extraction recovery influenced the genetic results, as the same analysis was undertaken for all samples, and the genetic analyses do not rely on quantification. Heritability assesses the similarities between the non-shared environment of individuals (e.g. measurement error) and heritability that is estimated as low can be due to measurement error (Abecasis et al., 2000, Visscher et al., 2008). Such error therefore deflates the estimates of heritability. The four heritable species identified from the Eico family include the CYP450-enzyme derived DHET species, which have an expected increased heritability compared to other species that are influenced by diet (e.g. HODE species). However, assessments of heritability for Eico species in only 200 plasma samples requires repetition in a cohort of substantial size.

While efforts have been made in this project to assess the quality of the lipidomics analysis pipeline, species-specific deuterated (or other) internal standards are required to fully assess the pipeline for each unique lipid species, which have unique analysis

patterns. Such standards would aid lipidomics research, if available to scientists at a low cost, as they allow for the analysis of a similarly structured species in the same sample to endogenous species, but can be identified separately by mass spectrometry, allowing for a comparison to the endogenous lipid species.

3.7.3 Plasma causes mass spectrometry matrix effects

Recovery compares known amounts of added lipid standards in plasma before extraction with the same amount added to the plasma sample after extraction. Matrix effect compares known amounts of standards in plasma to those of the same concentration, analysed in an ethanol solution. In this study, a large concentration of standards was added to the pooled plasma samples for comparison (20 pg/ul for NAE and 10 pg/ul for Eico). If there was variability in the concentration of the lipids in the pooled plasma samples, this would have influenced these analyses. While the recovery analysis was analysed using two pooled plasma samples, of which likely contained similar levels of endogenous lipids, the matrix effect analysis may have been influenced by the presence of endogenous lipids in the plasma sample, when comparing to the standards in ethanol. This may mean that the estimates of matrix effects were deflated, as the endogenous species likely added to the signal of the added standards in the pooled plasma sample, although likely only by a small amount as the endogenous lipids are found in plasma at the picogram per millilitre level. Analysis of standard reference materials would aid this study. However, techniques that remove lipids from a plasma sample would likely influence the plasma environment and not reflect the endogenous matrix of a plasma sample.

There is a lack of analyte-free matrix or Certified Reference Materials (CRMs) commercially available. Thus, the assessments of matrix effects and process efficiency here could only be assessed for losses of known amounts of lipid standards. As described, lipids were identified to be affected by ion suppression (i.e. matrix effect <100%), where the presence of high-abundance ions from artefacts found in plasma, suppressed the ionization and signal of co-eluting lipids.

Plasma is a complex matrix due to it's carrier role in circulating many substances throughout the body. It contains water, salts, enzymes, nutrients, hormones, proteins, waste products, antibodies, and clotting factors, which can potentially interfere with

the mass spectrometry analysis of the low concentration lipid metabolites. A high amount of endogenous components are analysed alongside lipids into the LC-MS/MS system when analysing such a complex matrix, and this can result in poor accuracy of measurement if the lipid signal is affected (Tsakelidou *et al.*, 2017).

3.7.3.1 CER[N(18)S(18)] and related 18-carbon non-hydroxy fatty acid-based CER species were not confirmed

CER[NS] species with a 18-carbon non-hydroxy fatty acid that have been analysed in literature in plasma and assessed as CVD biomarker CER species were not confirmed during the quality control analyses. A small peak was identified for the CER[N(18)S(18)] and at MRM analysis, fragments were not identified at the same retention time at a substantial abundance. The blank ethanol sample also carried 45% of the identified peak, therefore this lipid was excluded from further analyses.

The lipid may not be stable in the cohort plasma, may be at low concentration in the plasma samples and at the limit of detection of the mass spectrometer, or may not fragment well during mass spectrometry. The CER assay species that were not confirmed are likely background signals or impurities, and this emphasises the importance of confirmation of the peaks of each lipid species, as in many cases large, clear peaks at high concentration were present in the selected reaction transition used for the assay, but did not respond to all specific MRM transitions. The future availability of commercial standards for each lipid species in the CER assay would aid the accuracy of the measurement of these sphingolipids.

3.7.4 Extra lipidomic traits were calculated for genetic assessment

Hypothesis-free testing of ratios between all possible metabolite pairs in GWAS is an innovative approach in the discovery of new biologically meaningful associations (Petersen *et al.*, 2012). Ratios between metabolite concentrations have been introduced previously as biomarkers (e.g. urinary hydroxyproline to creatinine as an indicator of nitrogen dioxide exposure) and used in many biomedical applications (blood phenylalanine to tyrosine concentrations to identify carriers of phenylketonuria (PKU) risk alleles) (Petersen *et al.*, 2012). Such ratios are calculated and analysed regularly in studies assessing CER species as biomarkers (Laaksonen *et al.*, 2016),

and in GWAS of PUFAs (Suhre *et al.*, 2011) and other metabolites (Illig *et al.*, 2010). As an example of the use of such metabolic ratios, Illig *et al.* (2010) identified the correct biology behind multiple GWAS with metabolite ratios, identifying the correct genomic loci of relevant enzymes or transporter genes. Therefore, many biochemical product-precursor ratios and further calculated traits were analysed.

3.7.5 Inclusion of outliers alters genetic results

Participant outlier removal is well described for GWAS analyses based on relationship, ancestry, and ethnicity (Marees *et al.*, 2018), however, it is poorly described based on metabolite measurements, and thus there are no standards. Here, the substantial effect of the removal of a few extreme values after adjustments for covariates, was shown to increase the GWAS power in the analyses of metabolites.

3.7.6 GCTA is the most appropriate software for family-based GWAS analyses of multiple lipidomic traits

Of the family-based GWAS software recommended by the Wellcome Trust Advanced Course, GCTA was identified as most appropriate. With the added benefit of speed, the software was capable of running the analyses to high accuracy over a short time period. Other family-based association software have also been described (Ott *et al.*, 2011), however FaST-LMM and GCTA were recently created and have adaptations for use by high performance computer clusters, which allows for simultaneous analyses of multiple lipidomic traits.

3.7.7 The use of stringent imputation quality control thresholds provides more consistent GWAS results

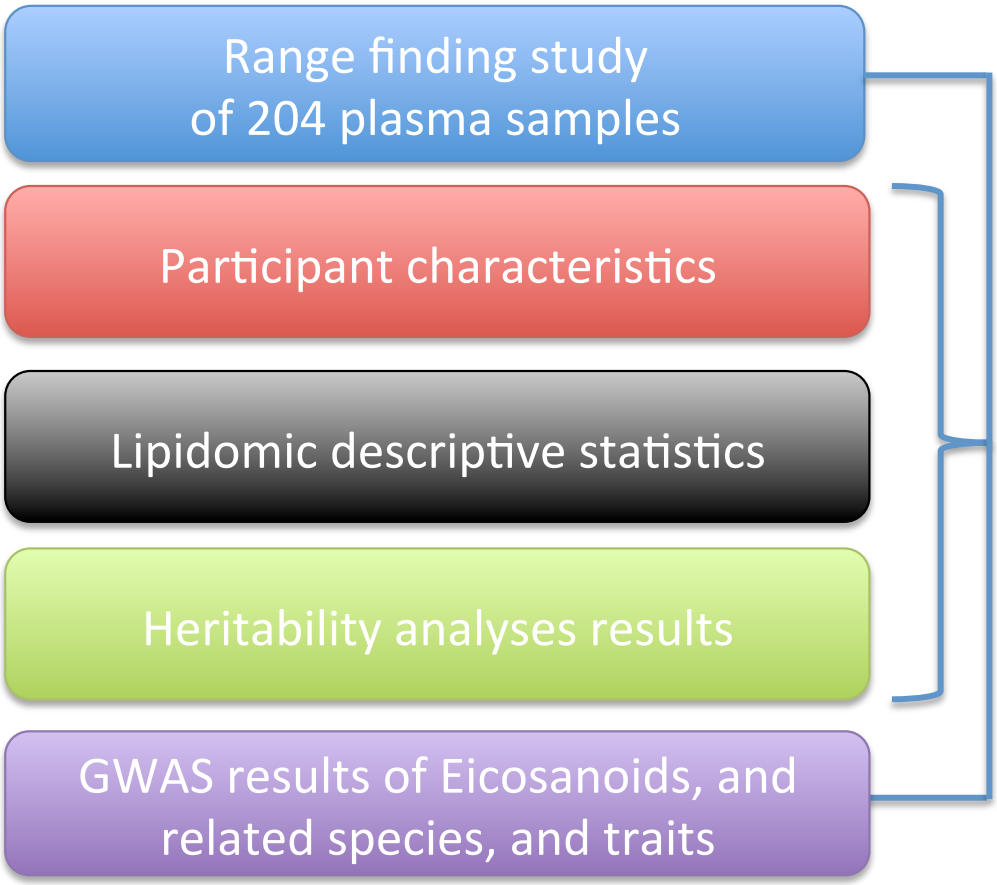
While the assessment of a basic GWAS and preparation of data for imputation is well explained (Marees *et al.*, 2018), the standard quality control thresholds for post-imputation analysis are not well described. Here, it is shown that stringent quality control thresholds (R^2>0.8; MAF>0.05) were required for robust Quantile-Quantile plots from GWAS results. It is likely that sequencing of the genome would provide further accuracy and confidence in the assessments of lower frequency variants.

3.8 Conclusion

To ensure the final results of the genetic analyses were robust, quality control assessments were undertaken at all stages of the project. This allowed for identification of the most abundant and consistently measured lipid species to support high quality genetic analyses. Lipid metabolites were detected in plasma comprised of 19 Eico species, 11 NAE species, and 30 CER species. Extra lipidomic traits based on abundance, metabolite ratios, and literature biomarker estimates, were calculated and included in analyses. The most extreme outlier values were identified and removed. GCTA was identified as the most appropriate software for family-based GWAS analyses with the benefit of high-speed analyses. The imputation threshold of R^2>0.8 and the threshold of MAF>0.05 resulted in the most robust GWAS analyses.

Chapter 4

Range finding study



4.1. Introduction, aim and objectives

A range finding study was completed to undertake a genetic assessment of 60 lipids from the NAE, CER, and Eico classes to identify those lipids under particular genetic influence for full cohort analyses (999 participants; 196 families). This study included 204 participants from 31 families. Of the 204 samples, 196 had genotyping data available.

Heritability estimates of each class of lipids resulted in 21% of detected Eico species, 33% of detected CER species, and 82% of detected NAE species estimated as significantly heritable. As the Eico class resulted in fewer heritable species than the NAE and CER classes (4, 9, and 10 lipid species estimated as significantly heritable in 196 samples, respectively), they were not taken forward for full cohort analyses (Chapter 5). In addition, the GWAS of 196 participants resulted in three CER traits that associated to GWAS significance with variants influencing the gene of the rate limiting step of the CER biosynthetic pathway (*SPTLC3*), confirming the major genetic influence over plasma CER lipid species.

Of the Eico species, LA-derived 13-HODE was most abundant in plasma. Correlation analyses identified strong correlations between PUFAs LA and ALA-derived Eico species, and independently derivatives of AA. Four species of Eico were estimated as significantly heritable, including CYP450-derived DHET isomers. The heritable species did not significantly associate with genomic loci at GWAS. Four other Eico traits associated with SNPs to GWAS significance (namely 5-HETE, 15-HETE, sum of AA-derived lipids, and the ratio of 9-OxoODE/9-HODE), but their identified associations were not directly linked to the genes of their respective biosynthetic pathways.

The objectives of the study were as follows:

1. Analyse 200 plasma samples for NAE, CER, and Eico lipid species.
2. Assess the heritability of all 60 detected plasma lipids from the three classes of bioactive lipids to identify the lipid classes under substantial genetic influence.
3. Select the lipid classes of interest for further targeted lipidomics analyses of the full cohort.

4. Undertake and assess GWAS on the class of lipids not taken forward for full cohort analyses.

4.2. Population characteristics of the range finding study

Plasma samples of 204 participants from 31 British Caucasian families were analysed by lipidomics techniques, of which 196 had genotyping data available and were included in the range finding genetic analyses (Figure 4-1). The families consisted of 1-24 members (median of 5 members). Participant descriptions are listed in Table 4.1.

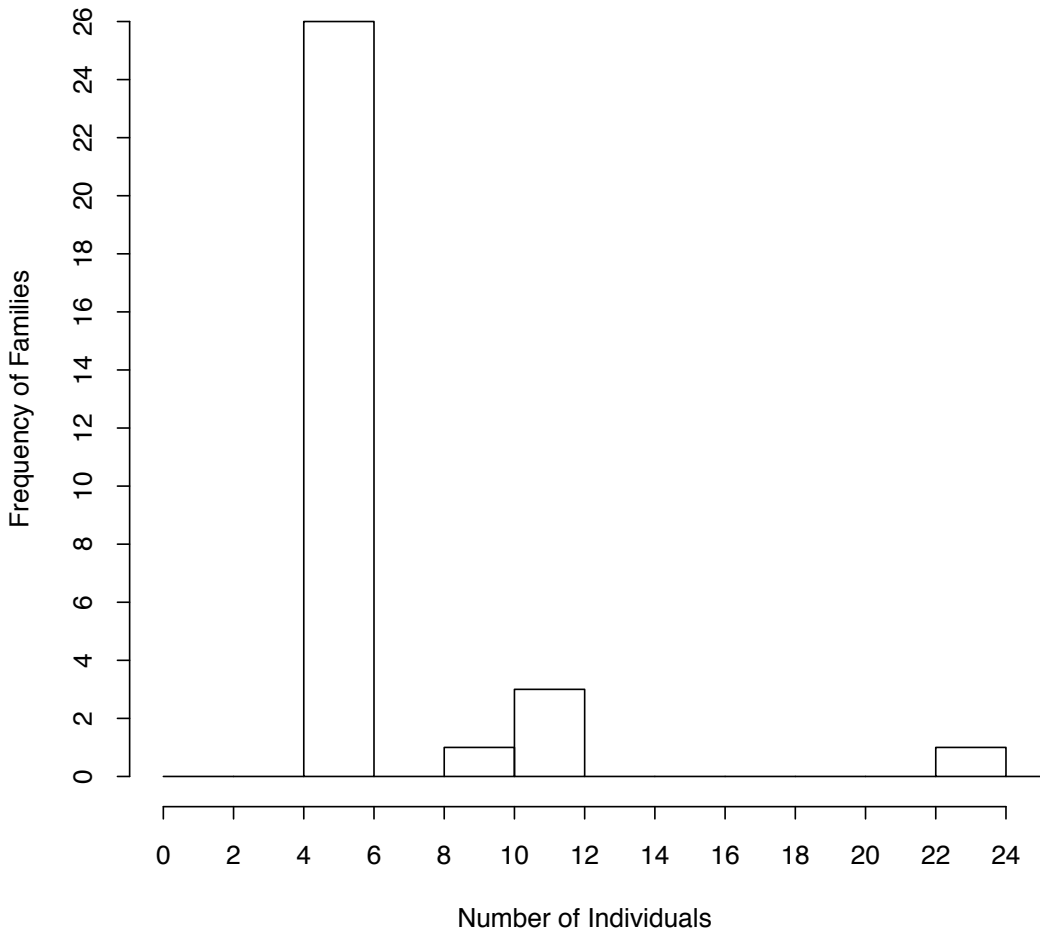


Figure 4-1: Distribution of the participant families included in the genetic analyses of the range finding study.

The histogram depicts the 31 families and the number of individuals in each family. The mean and median number of individuals in each family was 5.

Table 4.1: Summary statistics of the range finding study participants.

Data is shown as mean and standard deviation (SD) or percentage (%); BMI, body mass index; WHR, waist-hip ratio; Mean blood pressure, the mean of three readings taken in the clinic.

Trait	Mean (SD)
Gender	52% Male
Hypertensive	34%
Mean blood pressure	133/80 mmHg
Age (years)	46 (16)
BMI	25.97 (4.61)
WHR	0.86 (0.09)
Cholesterol (mmol/L)	5.46 (1.10)

4.3. Range finding study lipidomic descriptive statistics

The joint NAE and CER lipidomics assay was run over 19 extraction batches and six mass spectrometry batches. The Eico assay was run over 19 extraction batches and five mass spectrometry batches. 15% of the samples contained white blood cells or red blood cells, determined visually.

Summary statistics of the lipid species are in Appendix Table 0.3. Of the 30 plasma CER species identified, CER[N(24)S(18)] was most abundant (190.40 ± 68.53 pmol/ml). Of the 11 plasma NAE species identified, STEA was most abundant (1.04 ± 0.46 ng/ml). Of the 19 plasma Eico species identified, 13-HODE was most abundant (7.32 ± 6.63 ng/ml). The concentrations of the measured Eico species are depicted in Figure 4-2 (the abundance of the CER and NAE classes are described in Chapter 5).

The predictors included for each lipid species in the multiple linear regression analysis are shown in Appendix Table 0.4, which varied per lipid species. NAE species showed a positive association with total cholesterol. Pooled quality control samples and mass spectrometry batch were the most significant predictors for all classes of lipids.

As the lipid mediators studied can exert individual bioactivities (Quehenberger *et al.*, 2010), all lipid species were treated uniquely for all analyses, and intra-class correlation analyses are depicted in Figure 4-3. Eico species separated into two groups, one consisting of PUFAs LA- and ALA-derived species; the DiHOME, OxoODE, HODE, EpOME and HOTrE isomers, and another of the AA-derived DHET and HETE species. CER[NS] and CER[NDS] correlated strongly with other lipid mediators of their respective classes. Sphingosine base species (C18S-X) formed a group while CER[AS] species were found amongst the CER[NS]. All NAE lipids were strongly correlated, with OEA, PEA, and VEA species showing a particularly strong relationship ($R>0.7$).

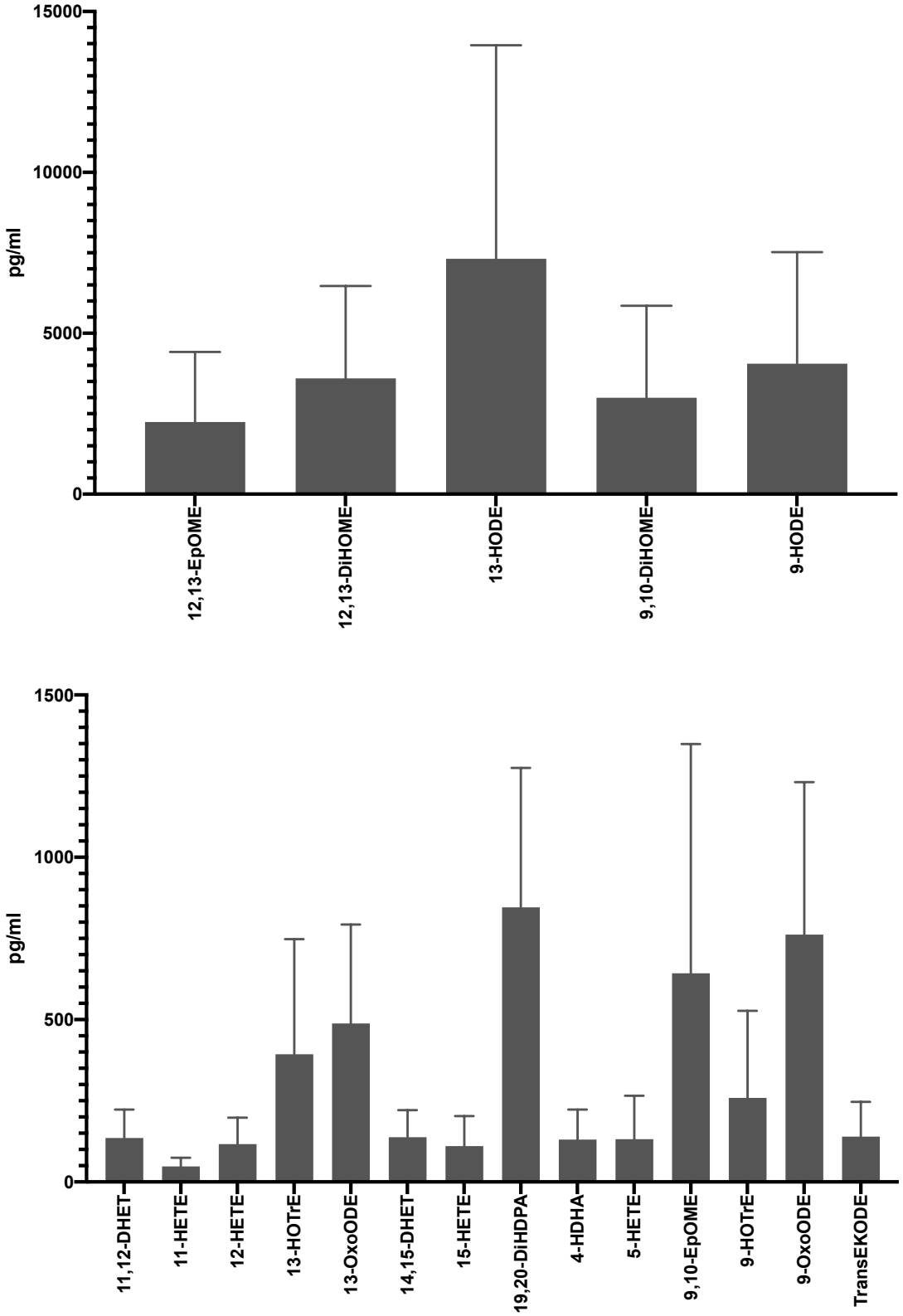


Figure 4-2: Plasma concentrations of the Eico species detected

Data shown as mean (pg/ml) ± SD of n=204 plasma samples.

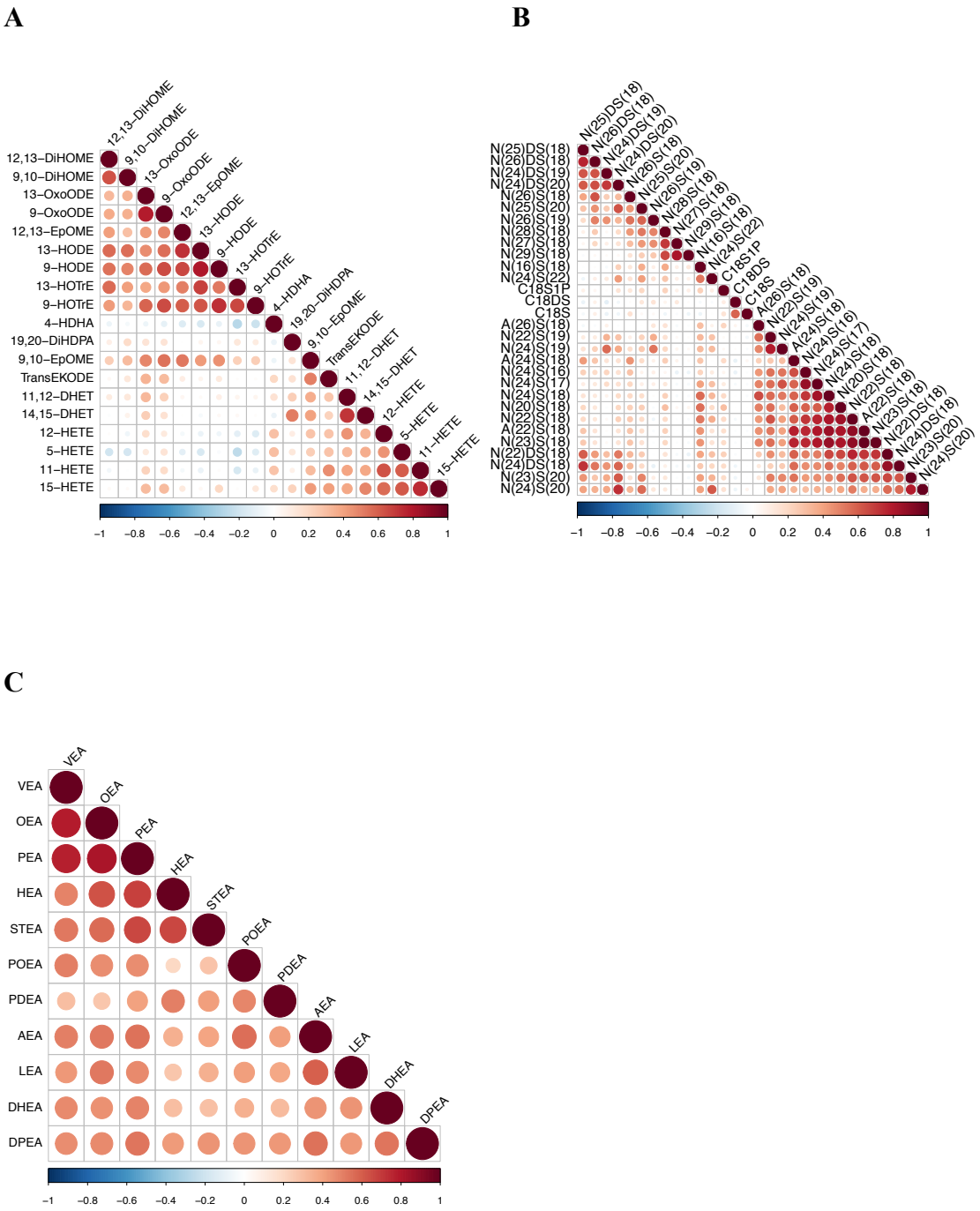


Figure 4-3: Intra-class lipid correlations of the range finding study

Assessment of the relatedness within the classes of lipids (A: Eico, B: CER, C: NAE) was explored using rquery.cormat in R. This takes into account strength of relationship (correlation coefficient; depicted as a scale of colours) and P-value (size of circle produced). The correlation was completed on the covariate-adjusted, standardised residuals with outliers removed; the data used for genetic analyses.

4.4. Heritability results for the range finding study

The Eico lipid species were less heritable than NAE and CER species (Figure 4-4, Table 4.2). Of the nineteen Eico species measured, only four species had significant estimated heritability ($P_{adj}<0.05$), namely PUFA AA and CYP450 enzyme-derived 11,12-DHET, 14,15-DHET, as well as DHA-derived 4-HDHA, and LA-derived TransEKODE. The significant heritability estimated for the four species ranged from 33%-59% of the variance in the concentration of the plasma lipids.

A larger subset of NAE and CER species were estimated as significantly heritable; nine of the eleven NAE lipid species detected were estimated as significantly heritable (28%-53%) and ten of the thirty CER species detected were significantly heritable (27%-68%). The most heritable lipids identified from the range finding study for each class include the α-hydroxy fatty acid-based CER species, CER[A(24)S(18)] (h^2_{QTDT} = 65%, P=6.00x10^{-15}; h^2_{SNP} = 68%, P=1.83x10^{-14}), the NAE species pentadecanoyl ethanolamide (PDEA; h^2_{QTDT} = 53%, P=2.20x10^{-7}; h^2_{SNP} = 51%, P=1.80x10^{-7}), and the Eico species 4-hydroxy-docosahexaenoic acid (4-HDHA; h^2_{QTDT} = 59%, P=3.80x10^{-9}; h^2_{SNP} = 59%, P=1.89x10^{-9}). The particular roles and relative importance of the CER and NAE lipids is unknown, while 4-HDHA has been found to regulate endothelial cell proliferation and angiogenesis in a mouse model of oxygen-induced retinopathy (Sapieha *et al.*, 2011).

Table 4.2: Heritability estimates

Heritability was estimated by pedigree-based QTDT software and SNP-based GCTA software. The P-values were adjusted for class-specific multiple testing (30 tests for CER, 19 tests for Eico, and 11 tests for NAE). Class, lipid class; h^2, heritability estimate; ChiSq, chi-squared values from QTDT; SE, standard error from GCTA. A P-value of >0.05 denotes a non-significant P-value not provided by QTDT software. Significant P-values were considered P_{adj}<0.05 (marked in red).

		QTDT				GCTA			
Class	**Lipid**	**h^2**	**ChiSq**	**P_{adj}-value**	**n**	**h^2**	**SE**	**P_{adj}-value**	**n**
CER	CER[A(22)S(18)]	0.26	5.29	6.45E-01	193	0.25	0.128407	3.77E-01	193
CER	CER[A(24)S(18)]	0.65	67.99	6.00E-15	191	0.68	0.092871	1.83E-14	191
CER	CER[A(26)S(18)]	0.29	7.35	2.01E-01	190	0.29	0.12828	8.23E-02	190
CER	CER[C18DS]	0.42	21.3	1.20E-04	195	0.44	0.126033	3.21E-05	195
CER	CER[C18S1P]	0.35	9.12	7.50E-02	175	0.34	0.140735	3.95E-02	175
CER	CER[C18S]	0.23	7.15	2.25E-01	187	0.24	0.11883	9.11E-02	187
CER	CER[N(16)S(18)]	0.13	1.86	>0.05	196	0.14	0.112678	2.12E+00	196
CER	CER[N(20)S(18)]	0.11	0.56	>0.05	192	0.13	0.13803	5.51E+00	192
CER	CER[N(22)DS(18)]	0.34	10.75	3.00E-02	193	0.37	0.12585	8.85E-03	193
CER	CER[N(22)S(18)]	0.31	10.27	3.90E-02	193	0.33	0.120692	1.27E-02	193
CER	CER[N(22)S(19)]	0.34	13.02	9.00E-03	194	0.35	0.120324	4.74E-03	194
CER	CER[N(23)S(18)]	0.29	7.3	2.07E-01	192	0.31	0.128771	7.46E-02	192
CER	CER[N(23)S(20)]	0.29	7.45	1.89E-01	194	0.31	0.129841	7.49E-02	194
CER	CER[N(24)DS(18)]	0.31	6.61	3.03E-01	192	0.3	0.135087	1.55E-01	192
CER	CER[N(24)DS(19)]	0.22	5.22	6.69E-01	192	0.23	0.130188	3.17E-01	192
CER	CER[N(24)DS(20)]	0.27	5.84	4.71E-01	194	0.26	0.132403	3.23E-01	194
CER	CER[N(24)S(16)]	0.25	5.48	5.76E-01	192	0.28	0.13168	2.02E-01	192
CER	CER[N(24)S(17)]	0.21	4.23	1.19E+00	192	0.23	0.122326	4.15E-01	192
CER	CER[N(24)S(18)]	0.32	9.79	5.40E-02	193	0.34	0.124615	1.78E-02	193
CER	CER[N(24)S(19)]	0.33	8.59	1.02E-01	193	0.33	0.133158	4.65E-02	193
CER	CER[N(24)S(20)]	0.25	4.57	9.75E-01	194	0.26	0.131082	4.22E-01	194
CER	CER[N(24)S(22)]	0.15	2.17	>0.05	194	0.15	0.118638	2.08E+00	194
CER	CER[N(25)DS(18)]	0.38	13.27	9.00E-03	192	0.37	0.13112	4.45E-03	192
CER	CER[N(25)S(20)]	0.16	1.67	>0.05	196	0.15	0.128159	3.18E+00	196
CER	CER[N(26)DS(18)]	0.28	8.32	1.17E-01	194	0.28	0.124153	6.22E-02	194
CER	CER[N(26)S(18)]	0.27	10.17	4.20E-02	195	0.27	0.11725	1.90E-02	195
CER	CER[N(26)S(19)]	0.39	8.16	1.29E-01	195	0.4	0.144391	5.18E-02	195
CER	CER[N(27)S(18)]	0.07	0.54	>0.05	192	0.08	0.112521	6.01E+00	192
CER	CER[N(28)S(18)]	0.14	1.86	>0.05	195	0.15	0.118916	2.13E+00	195

CER	CER[N(29)S(18)]	0.16	2.13	>0.05	193	0.15	0.120365	2.13E+00	193
Eico	11,12-DHET	0.33	8.8	5.70E-02	193	0.34	0.132561	2.77E-02	193
Eico	14,15-DHET	0.34	11.69	1.14E-02	195	0.37	0.126756	3.32E-03	195
Eico	19,20-DiHDPA	0.02	0.04	>0.05	194	0.04	0.106168	6.56E+00	194
Eico	12,13-DiHOME	0	0.04	>0.05	194	0	0.108123	9.50E+00	195
Eico	9,10-DiHOME	0.12	0	>0.05	195	0.12	0.11538	2.24E+00	192
Eico	12,13-EpOME	0.2	3.7	1.04E+00	174	0.19	0.129397	6.09E-01	174
Eico	9,10-EpOME	0.25	4.08	8.25E-01	173	0.26	0.144317	3.43E-01	173
Eico	4-HDHA	0.59	40.24	3.80E-09	190	0.59	0.110816	1.89E-09	190
Eico	11-HETE	0.16	2.8	1.80E+00	194	0.16	0.114373	8.45E-01	194
Eico	12-HETE	0.13	1.7	>0.05	194	0.17	0.122841	1.12E+00	194
Eico	15-HETE	0.17	2.08	>0.05	183	0.19	0.132105	1.06E+00	183
Eico	5-HETE	0.18	2.6	>0.05	179	0.17	0.124675	1.09E+00	179
Eico	13-HODE	0	0	>0.05	194	0.01	0.105175	9.09E+00	194
Eico	9-HODE	0	0	>0.05	194	0	0.100506	9.50E+00	194
Eico	13-HOTrE	0.07	0.42	>0.05	177	0.09	0.117605	3.89E+00	177
Eico	9-HOTrE	0	0	>0.05	167	0	0.123155	9.50E+00	167
Eico	13-OxoODE	0.05	0.22	>0.05	196	0.06	0.107425	5.54E+00	196
Eico	9-OxoODE	0.13	1.69	>0.05	194	0.14	0.114543	1.48E+00	194
Eico	TransEKODE	0.33	9.67	3.61E-02	193	0.34	0.133059	1.41E-02	193
NAE	AEA	0.17	2.85	1.01E+00	196	0.18	0.117665	4.51E-01	196
NAE	DHEA	0.11	1.09	>0.05	195	0.12	0.118545	1.51E+00	195
NAE	DPEA	0.38	8.84	3.19E-02	194	0.38	0.137039	1.53E-02	194
NAE	HEA	0.48	19.47	1.10E-04	194	0.46	0.128517	7.68E-05	194
NAE	LEA	0.28	6.55	1.16E-01	195	0.28	0.125221	4.59E-02	195
NAE	OEA	0.43	16.02	6.60E-04	194	0.42	0.13428	4.25E-04	194
NAE	POEA	0.45	13.29	3.30E-03	194	0.46	0.136435	9.91E-04	194
NAE	PEA	0.43	17.67	3.30E-04	195	0.41	0.128612	2.09E-04	195
NAE	PDEA	0.53	31.52	2.20E-07	196	0.51	0.115949	1.80E-07	196
NAE	STEA	0.33	8.89	3.19E-02	195	0.31	0.128156	2.54E-02	195
NAE	VEA	0.32	6.89	9.57E-02	194	0.33	0.138769	4.13E-02	194

Heritability was estimated by pedigree-based QTDT software and SNP-based GCTA software. The P-values were adjusted for class-specific multiple testing (30 tests for CER, 19 tests for Eico, and 11 tests for NAE). Class, lipid class; h^2, heritability estimate; ChiSq, chi-squared values from QTDT; SE, standard error from GCTA. A P-value of >0.05 denotes a non-significant P-value not provided by QTDT software. Significant P-values were considered $P_{adj}<0.05$ (marked in red).

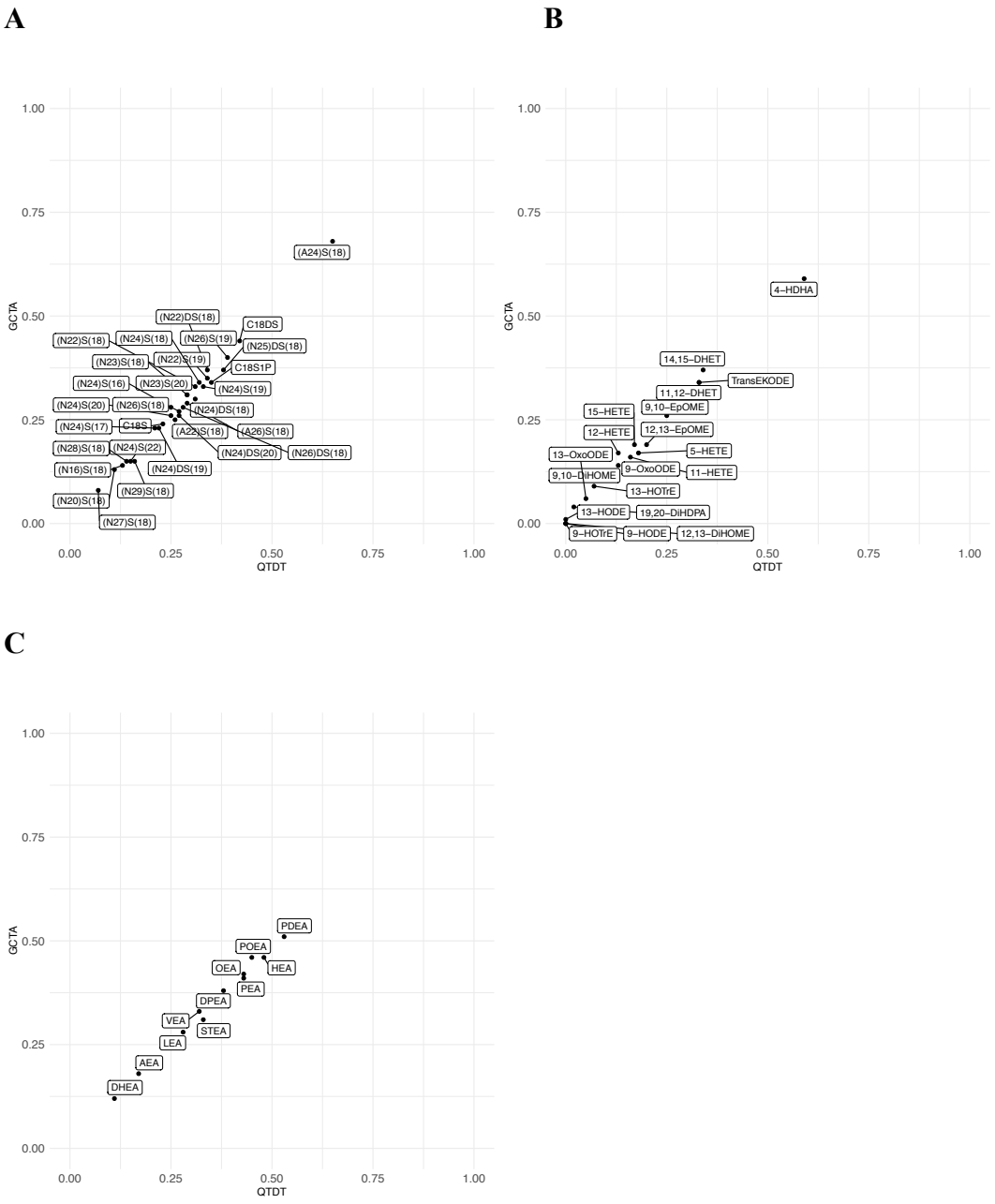


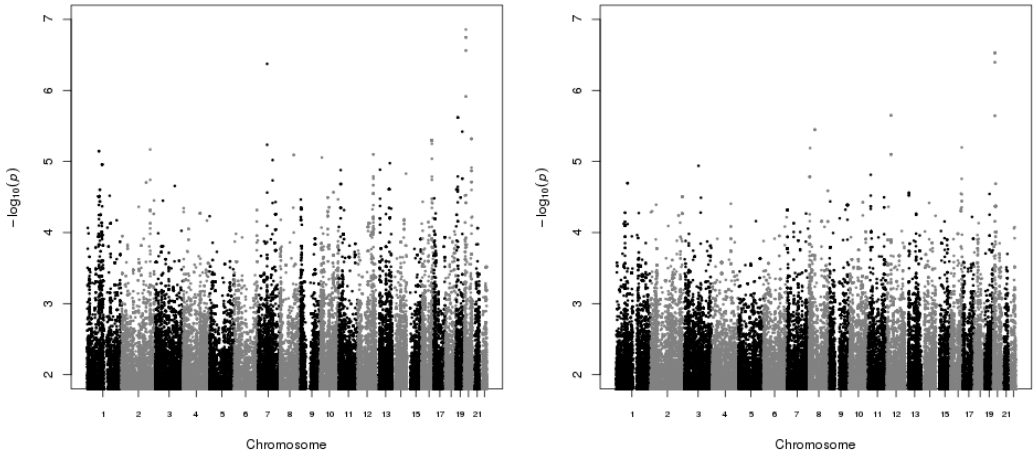
Figure 4-4: Heritability estimates

The panels depict the heritability of A) CER B) Eico C) NAE lipids. Heritability was estimated using SNP-based GCTA software (y-axis) and reported pedigree-based QTDT software (x-axis). Non-significant heritability estimates are depicted for reference (described in Table 4.2).

4.5. Range finding study GWAS results

CER and NAE species were taken forward for full cohort analyses with results presented in the following Chapter. The remaining focus of this Chapter is on the results of the Eico class. However, of note, the range finding study GWAS analyses of 196 samples for CERs and related mediators highlighted top SNPs for the lipid CER[N(24)S(19)], its precursor CER[N(24)DS(19)], as well as the trait for the sum of all 19-C sphingosine CER species (which includes CER[N(22)S(19)], CER[N(24)DS(19)], CER[N(24)S(19)], and CER[N(26)S(19)]), in the region of the serine palmitoyltransferase gene (*SPTLC3*) locus (e.g. rs680379, $P_{S19sum}= 3.17x10^{-8}$, Figure 4-5). The SNPs are GTEx confirmed eQTLs of liver *SPTLC3*. The genomic inflation factors for each of the analyses; CER[N(24)S(19)], CER[N(24)DS(19)], and the trait for the sum of all 19-C sphingosine CER species, were 1.017, 1.008, and 1.019, with lead SNP associations at the *SPTLC3* locus as follows; $P_{rs438568}= 1.39x10^{-7}$, $P_{rs1321940}=2.94x10^{-7}$, and $P_{rs438568}=1.56x10^{-8}$, respectively. The SNPs rs438568, rs1321940, rs364585, rs168622, rs680379, rs686548 were GWAS significant ($P<5x10^{-8}$) for association with the 19-C sphingosine CER species trait.

CER[N(24)S(19)] **CER[N(24)DS(19)]**



Sum of all 19-C sphingosine CER species

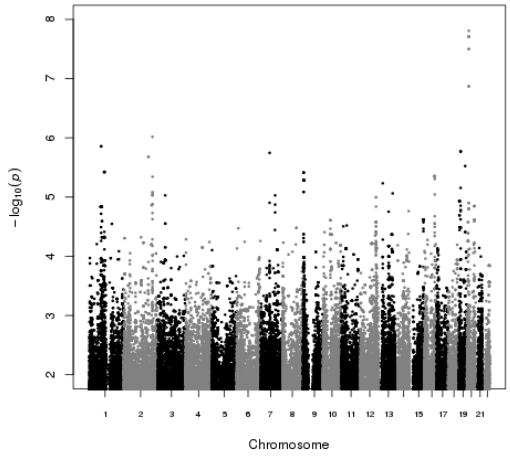


Figure 4-5: Manhattan plots of the GWAS results for CER species showing top associations with *SPTLC3* on chromosome 20 (n=196).

Each SNP measured (point) are depicted on chromosomes 1-22 (x-axis). The P-values are expressed as -log10(P-values), the most significant are those SNPs at the top of the plot; where a y-axis value of 8 denotes a P-value of $1.0x10^{-8}$. A stack of SNPs are shown at chromosome 20 in the region of *SPTLC3* for all three traits.

4.5.1. GWAS results of Eico class

Unless otherwise stated in this section, the SNPs described were not identified as eQTLs or in previously published GWAS (assessment via GWAS catalog, GTEx, and the UK Biobank Gene Atlas browser). The top 20 SNP associations resulting from the GWAS of each lipid species of the Eico group are summarised in Appendix Table 0.5. The genomic inflation factors from the GWAS analyses of each trait are presented in Table 4.3, with Quantile-Quantile plots presented as panels of Appendix Figure 0-1. Future work in large cohort studies will be required to understand fully the influence of genetics on the plasma levels of the Eico lipids.

Table 4.3: Genomic inflation factors for GWAS of Eicos and related traits

The table depicts the genomic inflation factors (GIF) for each GWAS analysis. X, more than half of the variance components were constrained for the trait.

Lipid	GIF	Lipid	GIF
AA	1.00	9-HOTrE	0.95
DHA	0.99	LA	0.94
11,12-DHET	1.00	13-OxoODE	1.02
14,15-DHET	1.01	9-OxoODE	1.02
19,20-DiHDPA	1.01	TransEKODE	1.01
12,13-DiHOME	0.92	aLA	1.00
9,10-DiHOME	1.01	Lox1	0.91
12,13-EpOME	1.01	cyp450	0.96
9,10-EpOME	1.01	epdi9	113.18
4-HDHA	1.01	epdi1213	X
11-HETE	1.01	lox15	1.00
12-HETE	1.01	ALOX5	0.95
15-HETE	1.02	omega3	0.96
5-HETE	1.01	omega6	0.94
13-HODE	1.01	oxho13	1.00
9-HODE	0.96	oxho9	1.00
13-HOTrE	1.02	EPXH2	0.89
SumEicos	0.90		

4.5.1.1. Significant GWAS associations

Of the 19 Eico and further traits included in the range finding study analyses of 196 samples with lipidomics and genotyping data available for analysis, 4 traits associated with SNPs to GWAS significance ($P<5x10^{-8}$). The Manhattan plots for the traits described in this section are depicted in Figure 4-6.

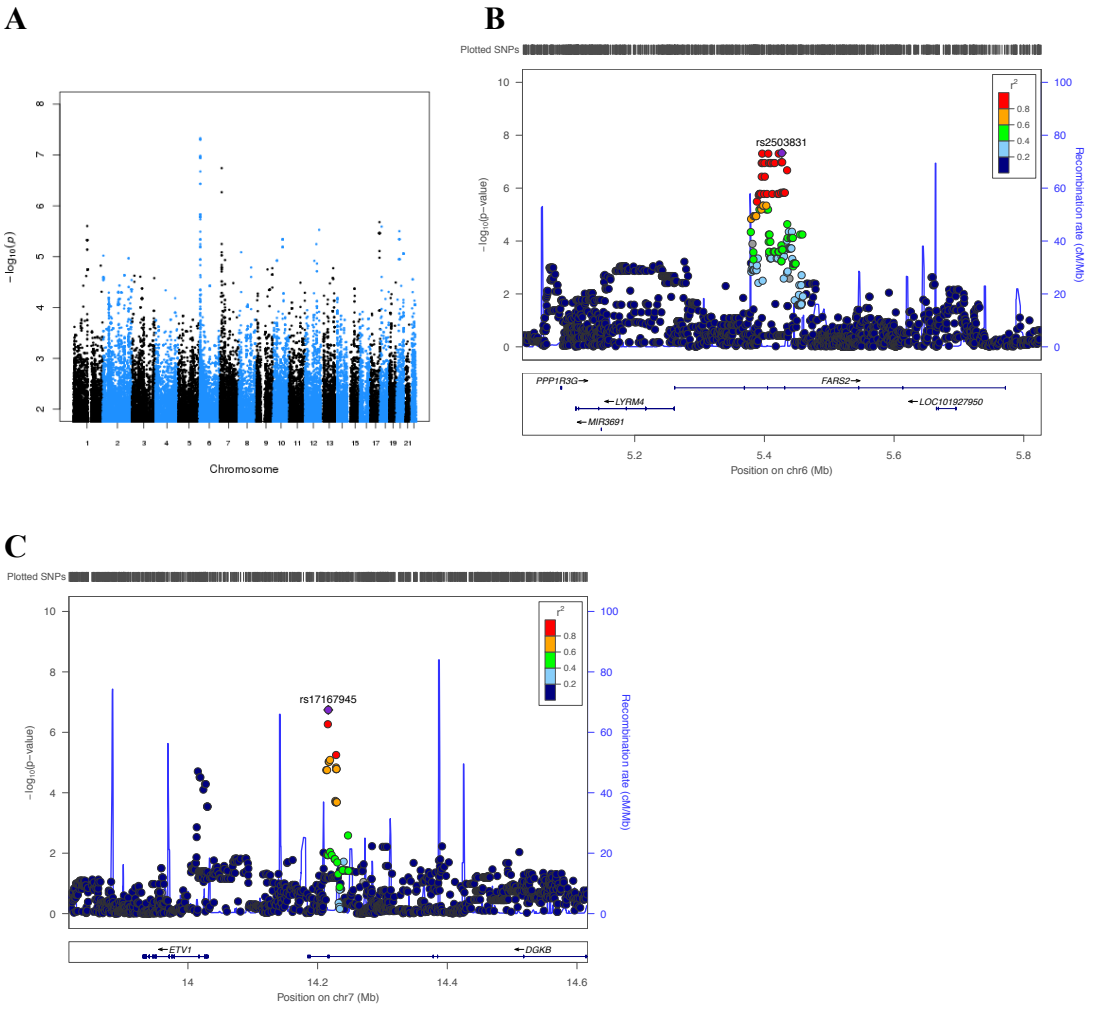
A trait describing the sum of all lipid derivatives of the PUFA AA, significantly associated with a stack of SNPs on chromosome 6, supported by three genotyped SNPs, in the intron of the gene phenylalanyl-tRNA synthetase 2, mitochondrial (*FARS2*; e.g. rs2503831 $P=4.67x10^{-8}$). These SNPs have a currently unknown relationship with AA. The trait also associated with a stack of non-significant SNPs on chromosome 7, including support from a genotyped SNP, in the intron of diacylglycerol kinase beta (*DGKB*).

15-HETE and 5-HETE species associated significantly at GWAS with intergenic SNPs. 15-HETE associated at GWAS ($P=2.92x10^{-8}$) to a single imputed SNP in an intergenic region on chromosome 16, upstream to the gene glutamate ionotropic receptor NMDA type subunit 2A (*GRIN2A*). 5-HETE associated at GWAS significance with two independent intergenic SNPs; rs77345935 on chromosome 6 ($P=4.32x10^{-8}$) and rs73106770 on chromosome 20 ($P=4.46x10^{-8}$). The chromosome 6 association was supported by a genotyped SNP and found in the locus downstream to defensin beta 110 (*DEFB110*), a family of antimicrobial and cytotoxic peptides produced by neutrophils, and upstream to transcription factor AP-2 delta (*TFAP2D)*. The top SNPs on chromosome 20 were all imputed SNPs and found downstream to synapse differentiation inducing 1 (*SYNDIG1*), interferon-induced transmembrane family of proteins, and upstream to cystatin F (*CST7*). The mechanism behind the association with these AA-derived species is currently unknown and where support for the association is not confirmed by genotyped SNPs (i.e. the 15-HETE association), these associations may not be replicable in a larger cohort.

The ratio of 9-OxoODE to its precursor 9-HODE (9-OXHO) was a trait created to identify SNPs in the genes of any proteins that could be involved in the oxidation of HODE lipid species to OxoODE products. Currently no enzyme is known to be involved in this reaction. At GWAS, this trait associated with a single imputed SNP in

chromosome 8 (rs34637388, $P=6.81x10^{-9}$). The SNP is found in the intron of cysteine rich secretory protein LCCL domain containing 1 (*CRISPLD1*). The relationship with Eico species is unknown and it is likely that this association would not be repeated in the analysis of a cohort of substantial size.

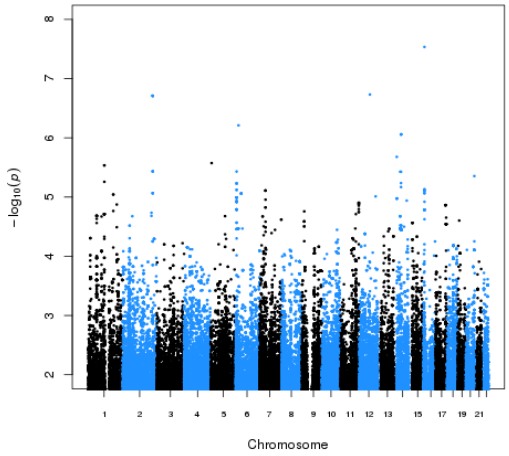
Trait: AA



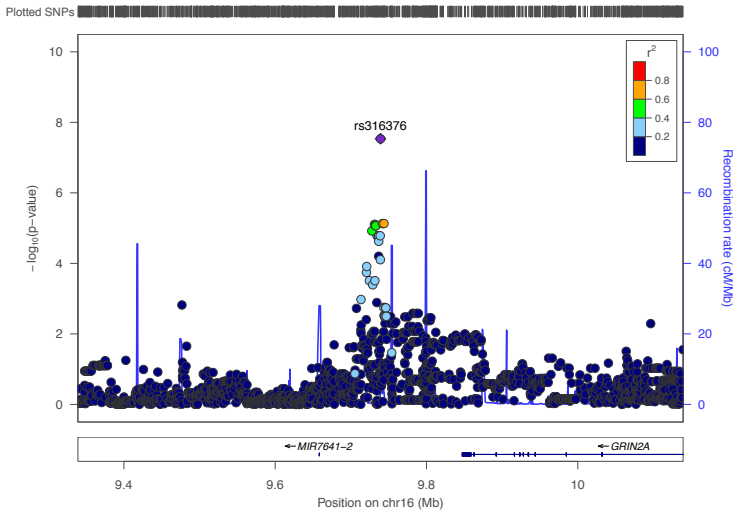
A) Manhattan plot of the GWAS. The plot depicts each SNP measured (point) on each chromosome from 1-22 (x-axis; autosomes only, alternating black and blue in colour). The P-values are expressed as -log10(P-values), the most significant and therefore lowest P-values are those SNPs at the top of the plot; where a y-axis value of 8 denotes a P-value of 1.0×10^{-8}. B-C) LocusZoom plots showing 500,000 base pairs either side of the lead SNP of an association. The linkage disequilibrium between SNPs is depicted, and each SNP measured is shown as a point. The P-value of association is shown on the left Y-axis, and the recombination rate of the locus is shown on the right Y-axis. Nearby genes are depicted underneath the plot.

Trait: 15-HETE

A

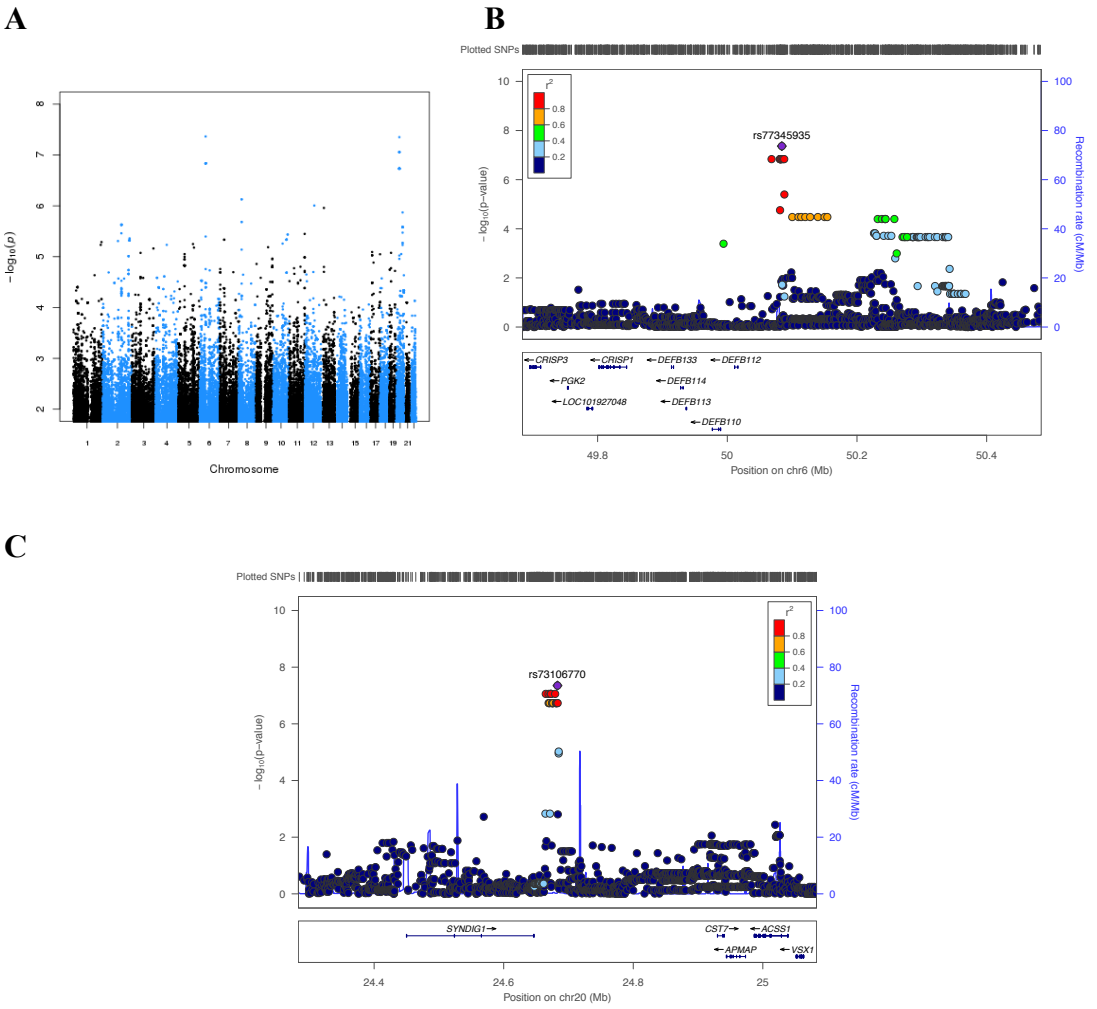


B



A) Manhattan plot of the GWAS. The plot depicts each SNP measured (point) on each chromosome from 1-22 (x-axis; autosomes only, alternating black and blue in colour). The P-values are expressed as -log10(P-values), the most significant and therefore lowest P-values are those SNPs at the top of the plot; where a y-axis value of 8 denotes a P-value of 1.0×10^{-8}. B-C) LocusZoom plots showing 500,000 base pairs either side of the lead SNP of an association. The linkage disequilibrium between SNPs is depicted, and each SNP measured is shown as a point. The P-value of association is shown on the left Y-axis, and the recombination rate of the locus is shown on the right Y-axis. Nearby genes are depicted underneath the plot.

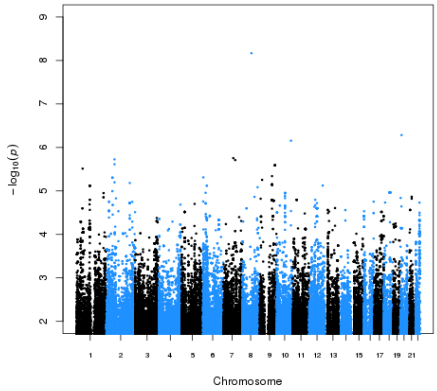
Trait: 5-HETE



A) Manhattan plot of the GWAS. The plot depicts each SNP measured (point) on each chromosome from 1-22 (x-axis; autosomes only, alternating black and blue in colour). The P-values are expressed as -log10(P-values), the most significant and therefore lowest P-values are those SNPs at the top of the plot; where a y-axis value of 8 denotes a P-value of 1.0×10^{-8}. B-C) LocusZoom plots showing 500,000 base pairs either side of the lead SNP of an association. The linkage disequilibrium between SNPs is depicted, and each SNP measured is shown as a point. The P-value of association is shown on the left Y-axis, and the recombination rate of the locus is shown on the right Y-axis. Nearby genes are depicted underneath the plot.

Trait: 9-OXHO

A



B

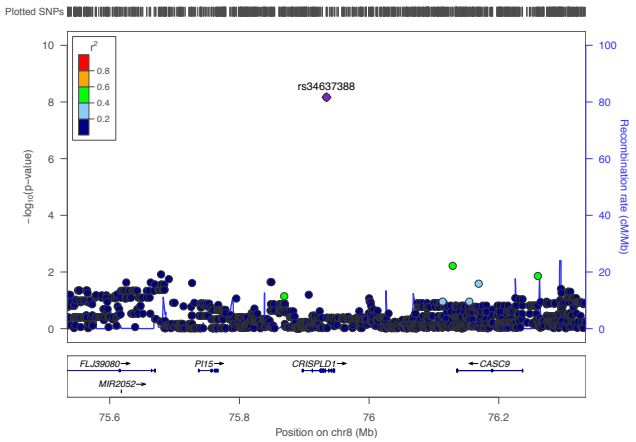


Figure 4-6: Manhattan and LocusZoom plots for AA, 15-HETE, 5-HETE, 9-OXHO traits

A) Manhattan plot of the GWAS. The plot depicts each SNP measured (point) on each chromosome from 1-22 (x-axis; autosomes only, alternating black and blue in colour). The P-values are expressed as -log10(P-values), the most significant and therefore lowest P-values are those SNPs at the top of the plot; where a y-axis value of 8 denotes a P-value of 1.0×10^{-8}. B-C) LocusZoom plots showing 500,000 base pairs either side of the lead SNP of an association. The linkage disequilibrium between SNPs is depicted, and each SNP measured is shown as a point. The P-value of association is shown on the left Y-axis, and the recombination rate of the locus is shown on the right Y-axis. Nearby genes are depicted underneath the plot.

4.5.1.2. Significantly heritable species

The four Eico lipids that were estimated as significantly heritable were not found to have GWAS significant associations ($P<5x10^{-8}$). These GWAS results would likely change if the analysis was repeated with a larger number of samples and the relationship between the SNPs identified with all species described here is currently unknown.

The species 11,12-DHET had a top association on chromosome 4 (rs6551813, $P=6.38x10^{-8}$) with a variant in the intron of the gene trans-2,3-enoyl-CoA reductase like (*TECRL*). The protein belongs to the steroid 5-α reductase family. It's isomer 14,15-DHET, associated with a SNP on chromosome 10 (rs113862732, $P=1.19x10^{-6}$), the intergenic SNP is found upstream to the gene zinc finger MIZ-type containing 1 (*ZMIZ1*).

4-HDHA associated at GWAS with a stack of non-significant SNPs on chromosome 12 (lead SNP rs11108140, $P= 3.08x10^{-7}$). The SNPs are downstream to ubiquitin specific peptidase 44 (*USP44*) and upstream to phosphoglycerate mutase 1 pseudogene 5 (*PGAM1P5*) and netrin 4 (*NTN4*). TransEKODE associated with a SNP on chromosome 4 (lead SNP rs75592902, $P= 5.64x10^{-8}$) in the intron of Rap1 GTPase-GDP dissociation stimulator 1 (*RAP1GDS1*).

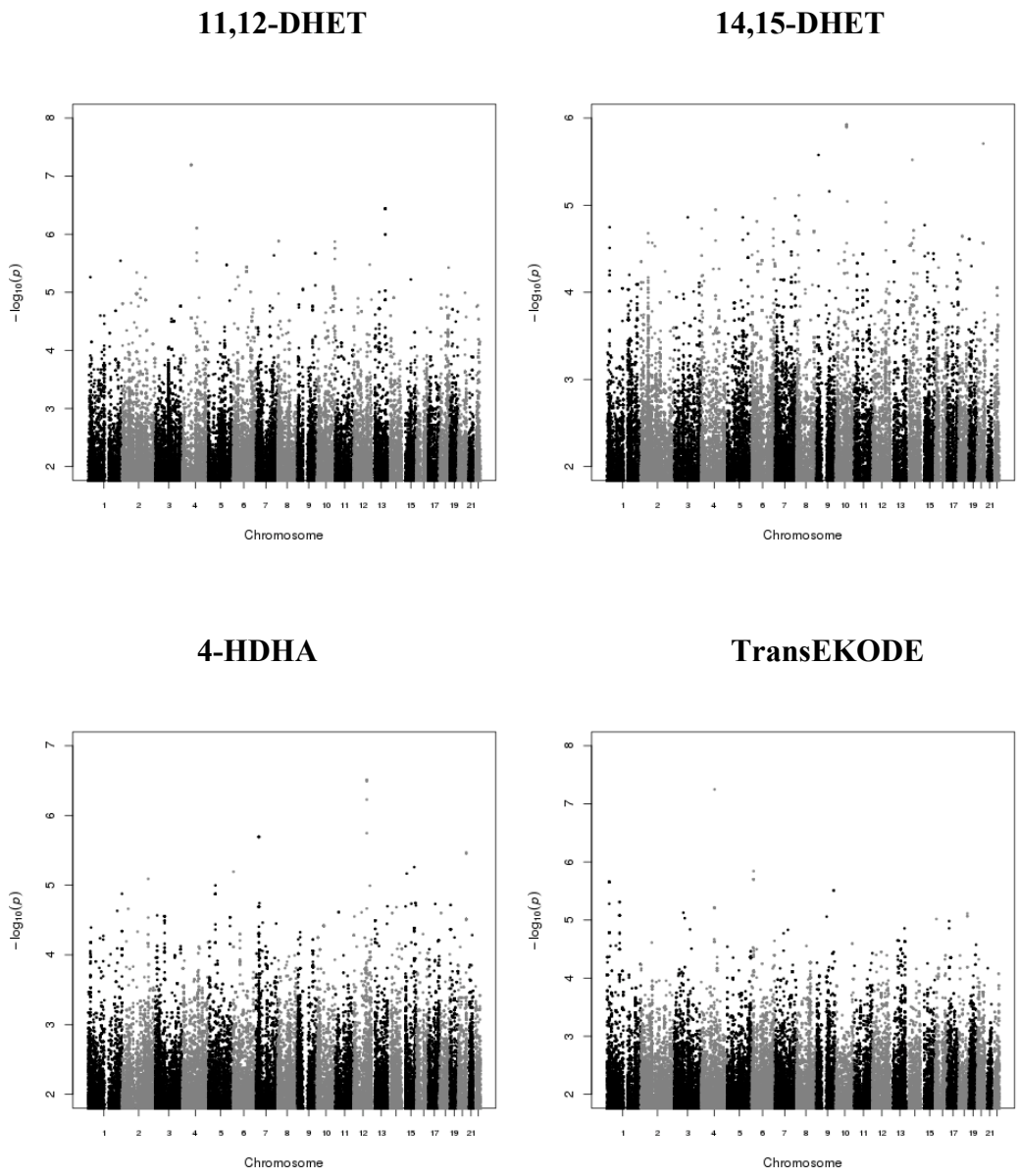


Figure 4-7: Manhattan plots of the GWAS results for the heritable Eico species

The plot depicts each SNP measured (point) on each chromosome from 1-22 (x-axis). The P-values are expressed as -log10(P-values), the most significant and therefore lowest P-values are those SNPs at the top of the plot; where a y-axis value of 8 denotes a P-value of 1.0×10^{-8}.

4.5.1.3. Other findings

Non-significant findings that seemed to point to a specific locus, via stacks of SNPs nearing association, were found for the traits described in this section (Figure 4-8). The trait "omega-3", comprised of summed n-3 lipids, associated at GWAS with a stack of non-significant SNPs on chromosome 1 (e.g. rs7526572, P= 3.61×10^{-7}). The SNPs are intronic variants of the pseudogene *NBPF13P* (neuroblastoma breakpoint family member 13) and GTEx confirmed eQTLs of the gene in the lung, as well as the gene *RP11-325P15.1* in subcutaneous adipose tissue and the lung. The variants are upstream to protein kinase AMP-activated non-catalytic subunit beta 2 (*PRKAB2*). The association with n-3 lipids is currently unknown.

The trait for a summation of the lipid species formed through reactions involving soluble epoxide hydrolase (*EPHX2;*19,20-DiHDPA, 11,12-DHET, 14,15-DHET, 9,10-DiHOME, and 12,13-DiHOME) associated at GWAS with stacks of SNPs at chromosomes 4 and 12. The chromosome 12 SNPs (e.g. rs10902512, P= 2.18×10^{-6}) are intronic variants of polypeptide N-acetylgalactosaminyltransferase 9 (*GALNT9*). The SNP is a GTEx confirmed eQTL of *RP13-977J11.2* in multiple tissues including whole blood, and the gene NANONGNB pseudogene 2 (*NANOGNBP2*) in the cerebellum. The SNPs at chromosome 4 (e.g. rs10020476, P=9.30×10^{-6}) are confirmed eQTLs of *SLC7A11-AS1* in the aorta and found upstream to nocturnin (*CCRN4L)* thought to be involved in circadian rhythms. The involvement of these loci with *EPHX2* or Eico species is unknown, and the *EPHX2* locus on chromosome 8 was not associated with this trait.

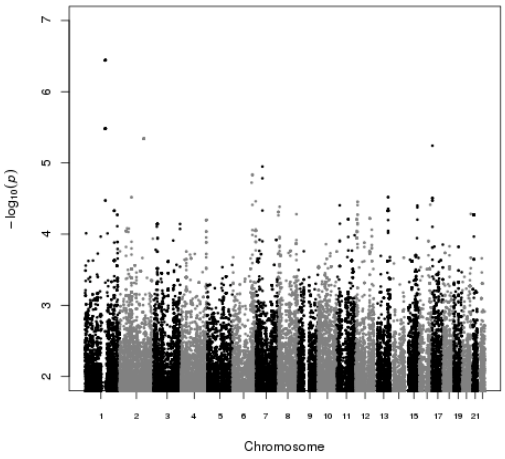
Two further traits were created to identify SNPs in the gene on chromosome 8 of the enzyme soluble epoxide hydrolase (*EPHX2*), involved in the hydrolysis of fatty acid epoxides to product lipid species; the trait of the ratio of 9,10-DiHOME to its precursor 9,10-EpOME (epdi910), and the trait of the ratio of 12,13-DiHOME to its precursor 9,10-EpOME (epdi1213). These traits were not analysed as epdi910 showed extreme inflation with a genome inflation factor of 114, and the results are therefore unlikely to be true (Quantile-Quantile plot depicted in Figure 4-9). The trait epdi1213 had more than half of the variance components constrained, so this trait was also excluded from GWAS analyses. While ratios are informative in GWAS (Kalsbeek *et*

al., 2018) and are regularly used in the metabolomics field for biomarker discovery (Laaksonen *et al.*, 2016), in this instance this ratio was uninformative, likely due to the sample size.

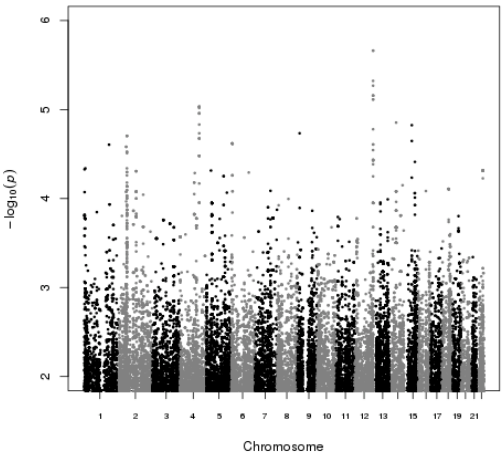
12,13-EpOME associated with a stack of SNPs on chromosome 9 scattered across the chromosome. The lead SNP, rs2795361 (P= 8.14×10^{-8}), is a GTEx confirmed eQTL of the pseudogenes growth arrest specific 2 like 1 pseudogene 2 (*GAS2L1P2*), vomeronasal 1 receptor 51 pseudogene (*VN1R51P*), ankyrin repeat domain 18C, pseudogene (*ANKRD18CP*), and *RP11-498P14.5* in multiple non-haematological tissues (brain, esophagus, thyroid, testis, colon). The lead SNP is found downstream to the gene coiled-coil domain containing 180 (*CCDC180*), while other top SNPs on chromosome 9 were found in the intron of tudor domain containing 9 (*TDRD7*; e.g. rs4567164, P=1.84×10^{-7}). The isomer, 9,10-EpOME, associated with a top SNP (rs1014053, P= 6.99×10^{-6}) in an intergenic region of chromosome 13 and no other information was found on this association.

11-HETE associated with a stack of SNPs on chromosome 6 (e.g. rs65972 29, P= 2.74×10^{-7}) upstream to basic transcription factor 3 pseudogene 7 (*BTF3P7*), and a confirmed eQTL of the gene in the mucosa of the esophagus. 9-OxoODE associated with top SNPs in the intron of interaction protein for cytohesin exchange factors 1 (*IPCEF1*) on chromosome 6 (e.g. rs12209958, P= 9.60×10^{-8}). The isomer of 9-OxoODE, 13-OxoODE, associated with a SNP on chromosome 4 (rs58081193, P= 1.35×10^{-6}) in the intron of coiled-coil serine rich protein 1 (*CCSER1*). The mechanisms behind these associations are unknown.

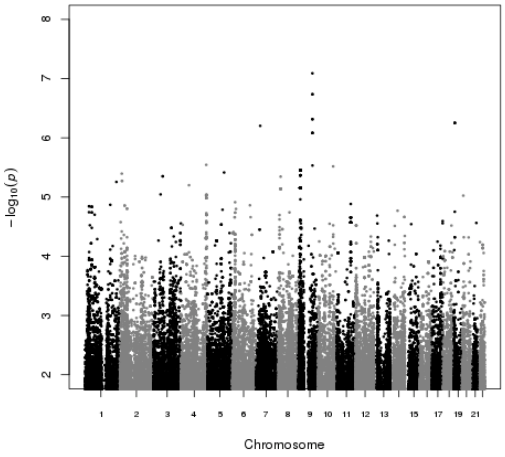
Omega-3



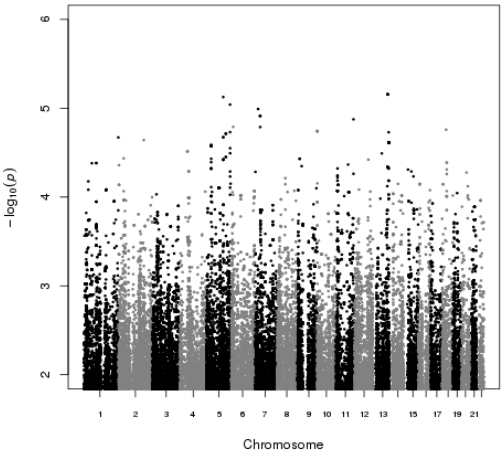
EPHX2



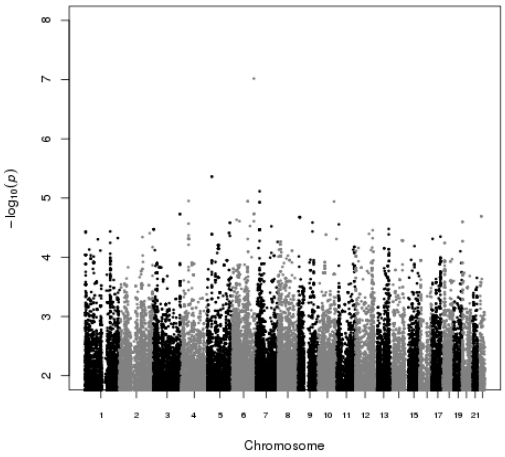
12,13-EpOME



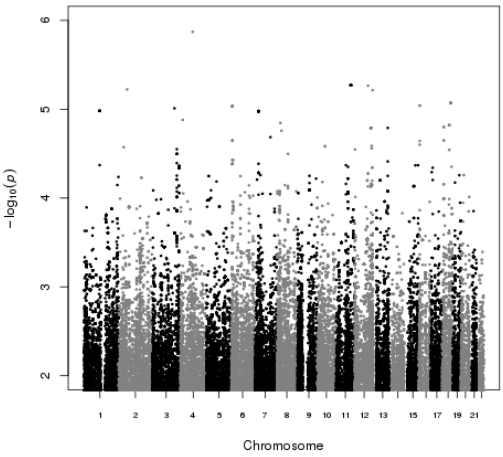
9,10-EpOME



9-OxoODE



13-OxoODE



11-HETE

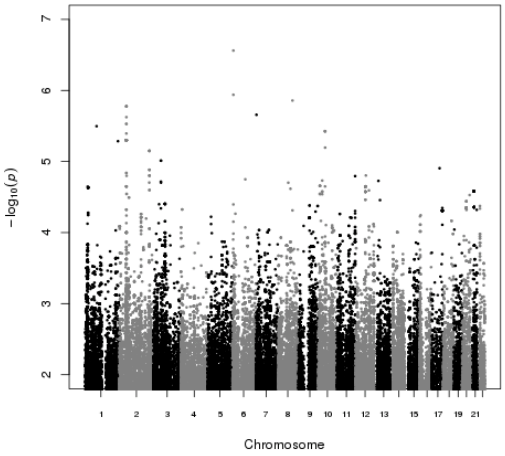


Figure 4-8: Manhattan plots depicting the GWAS results of suggestive non-significant associations

Manhattan plots depicting the results of the genome-wide association studies. The plot depicts each SNP measured (point) on each chromosome from 1-22 (x-axis; autosomes only, alternating black and blue in colour). The P-values are expressed as -log10(P-values), the most significant and therefore lowest P-values are those SNPs at the top of the plot; where a y-axis value of 8 denotes a P-value of $1.0x10^{-8}$.

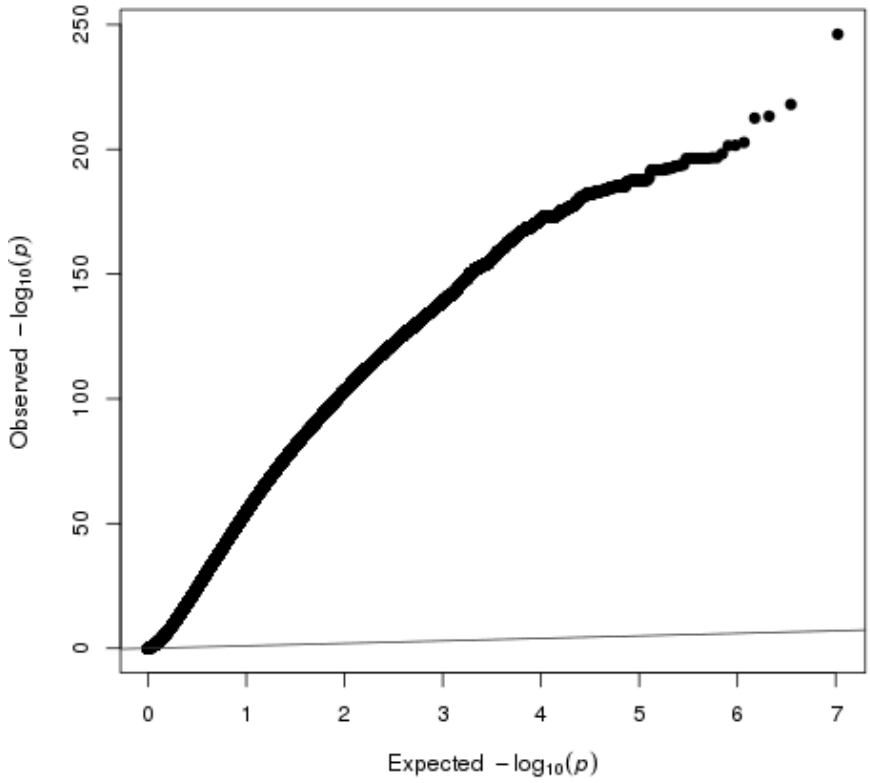


Figure 4-9: Quantile-Quantile plot of the GWAS for the trait epdi910

The plot depicts the expected versus observed P-values of association achieved at GWAS for the trait epdi910, which as shown, was highly inflated.

4.6. Discussion

This study identified species of the CER, NAE, and Eico classes under significant genetic influence. 10 CER species, 9 NAE species, and 4 Eico species were found significantly heritable in the 196 samples tested. The gene encoding a subunit of the protein responsible for the rate limiting step of the sphingolipid pathway (Perry *et al.*, 2000), the enzyme serine palmitoyltransferase, was identified in the 196 samples to associate at GWAS with two CER species; a CER[NS] and a precursor CER[NDS], highlighting a genomic loci of great genetic effect on CER lipid species. CER and NAE lipids were taken forward for full cohort analyses because of the particular genetic influence over them in comparison to the Eico class.

4.6.1. Participant characteristics

The range finding study conducted was a good representation of the full cohort (Chapter 5), as the participant characteristics are very similar. Healthy blood pressure is considered between 90/60 mmHg and 120/80 mmHg; healthy BMI is considered between 18.5-24.9; healthy WHR is considered as 0.9 or less (different for gender); and healthy total blood cholesterol levels is below 5.00 mmol/L [NHS website: https://www.nhs.uk/, date assessed 30/09/19].

As the mean blood pressure was 133/80 mmHg, the mean systolic blood pressure measured was raised. This is as anticipated, due to the ascertainment strategy enriching for hypertension; 34% of participants were classified as hypertensive. The gender of the range finding study participants was balanced (52% male) and the mean age of the cohort was 46 years old.

The participants' mean BMI was raised, with a reading of 25.97, the mean WHR was at the upper limits of the healthy boundary (0.86), and the mean total blood cholesterol measured for the participants was raised also at 5.46 mmol/L. While cholesterol is part of a separate lipid biosynthetic pathway to that of the bioactive lipids studied here, it was included as a potential covariate here, as it has been previously to explore lipid species such as CER (Demirkan *et al.*, 2012; Laaksonen *et al.*, 2016), and was used to adjust for a dietary influence on the lipid species studied.

4.6.2. Adjustments for batch effects

Pooled quality control samples and mass spectrometry batch were the most significant predictors for all classes of lipids, with the relationship depending on lipid species. This highlights the importance of adjustments of batch effects in bioanalysis.

4.6.3. LA-derived 13-HODE is the most abundant plasma Eico species

The most abundant of the 19 plasma Eico species identified, 13-HODE, is a LA-derived lipid generated through the activity of 15-lipoxygenase (*ALOX15*). The lipid has been shown to be changed in concentration postprandial, which may be why it was at high concentration in this cohort of non-fasting plasma samples, and why it was not estimated as significantly heritable (Gouveia-Figueira *et al.*, 2015).

4.6.4. Intra-class correlations

Correlation analyses highlighted strong correlations between lipids of the Eico class. The class separated into two groups, one consisting of the LA- and ALA-derived DiHOME, OxoODE, HODE, EpOME and HOTrE isomers, and another of the AA-derived DHET and HETE species. Therefore, the levels of lipid mediators in plasma were related depending on the prevalence of the parent PUFA, which may be a reason why many of the lipids were not significantly heritable; they are potentially more influenced by non-genetic factors such as PUFA substrate abundance. However, further studies are required to assess this hypothesis.

The CER species, CER[NS] and precursor CER[NDS], were correlated. CER[NS)] species with an 18-C sphingosine backbone, formed a group through correlation, as did the species with varying non-hydroxy fatty acid carbon length (16-20 C atoms). The CER[AS] correlated strongly with CER[NS], which could be due to the same enzymes involved in their biosynthesis (Figure 1-2). Sphingosine base species, C18S, C18S1P and C18DS, did not correlate with either the CER[NS] or CER[NDS]. This might be due to the cell survival roles of C18S and C18S1P species (Cuvillier *et al.*, 1996) as potential regulators of CER[NS]-induced apoptosis, and the important genetic influence of the enzymes in catalysing the *de novo* production of CER[NS] and CER[NDS] species, as opposed to the level of the C18DS substrate available.

NAE species were all strongly correlated and although they derive from different fatty acid substrates, they are produced similarly through the same enzymatic reactions and degraded by one enzyme only, the fatty acid amide hydrolase protein (*FAAH*) (Figure 1-1). Therefore, it is likely that the enzymes have a greater role in plasma levels of NAE species, than the fatty acids they are created from.

The Eico species correlated strongly with other mediators derived from the same fatty acid substrate. This suggests that levels of plasma Eico species could be more determined by substrate, and not enzymatic reactions, something that was not observed for the more heritable CER and NAE.

4.6.5. The Eico class contains fewer heritable species than NAE and CER

Heritability analysis of the range finding study suggested that CER and NAE classes contained more genetically influenced plasma lipid species than the class of Eico and related species, in this small sample size. Of the 19 Eico species analysed, four were estimated as significantly heritable, while 9 of 11 NAE and 10 of 30 CER had substantial significant heritability (Table 4.2). The Eico species 9,10-EpOME, 12,13-EpOME, and 9-HOTrE may have had deflated heritability estimates due to the extraction losses occurring during the measurement of these three species.

CYP450 enzymes, which are well studied for their genetic influence over drug metabolism (Zanger *et al.*, 2013), are involved in the production of DHET derivatives of PUFA LA that were estimated as substantially heritable (Figure 1-4). The particular role and importance of the most heritable species, 4-HDHA, a DHA-derivative, is unknown, but it has been shown to inhibit endothelial cell proliferation and sprouting angiogenesis in proliferative retinopathy models (Sapieha *et al.*, 2011). This may provide important information for future drug targets of hereditary retinopathies (Humphries *et al.*, 1992), but it will first need to be confirmed in large cohort studies.

4.6.6. GWAS results of the Eico class suggest genetic signals would be identified in a future large cohort study

While the lipid species estimated as significantly heritable did not associate with genome loci at GWAS significance, it is likely that associations would be identified with repetition of this cohort in a substantial size.

The AA-derived metabolite 15-HETE has been shown to confer concentration-dependent protection to glutamate toxicity in cultured cortical neurons (Hampson *et al.*, 2002). The association of 15-HETE to the region on chromosome 16 upstream to the gene glutamate ionotropic receptor NMDA type subunit 2A (*GRIN2A*) may be promising as a target for follow up in another cohort of samples.

Another finding of interest is the association of AA-derived metabolite 5-HETE with the genomic locus at transcription factor AP-2 delta (*TFAP2D*). The protein encoded by this gene regulates phospholipase A2 group IVA (*PLA2G4A)* involved in Eico biosynthesis, during development and aging (Ryan *et al.*, 2014). The enzyme catalyses the hydrolysis of membrane phospholipids in the release of arachidonic acid, which is subsequently metabolized into Eico species including 5-HETE [PubMed Gene ID: 5321]. 5-HETE also associated at GWAS to the genomic locus of the gene for defensin protein. In insects, RNAi treatment of PLA_2 has been shown to reduced the expression of inflammatory markers including defensin (Hwang *et al.*, 2013), thought to be mediated through Eico species (Stanley *et al.*, 2019).

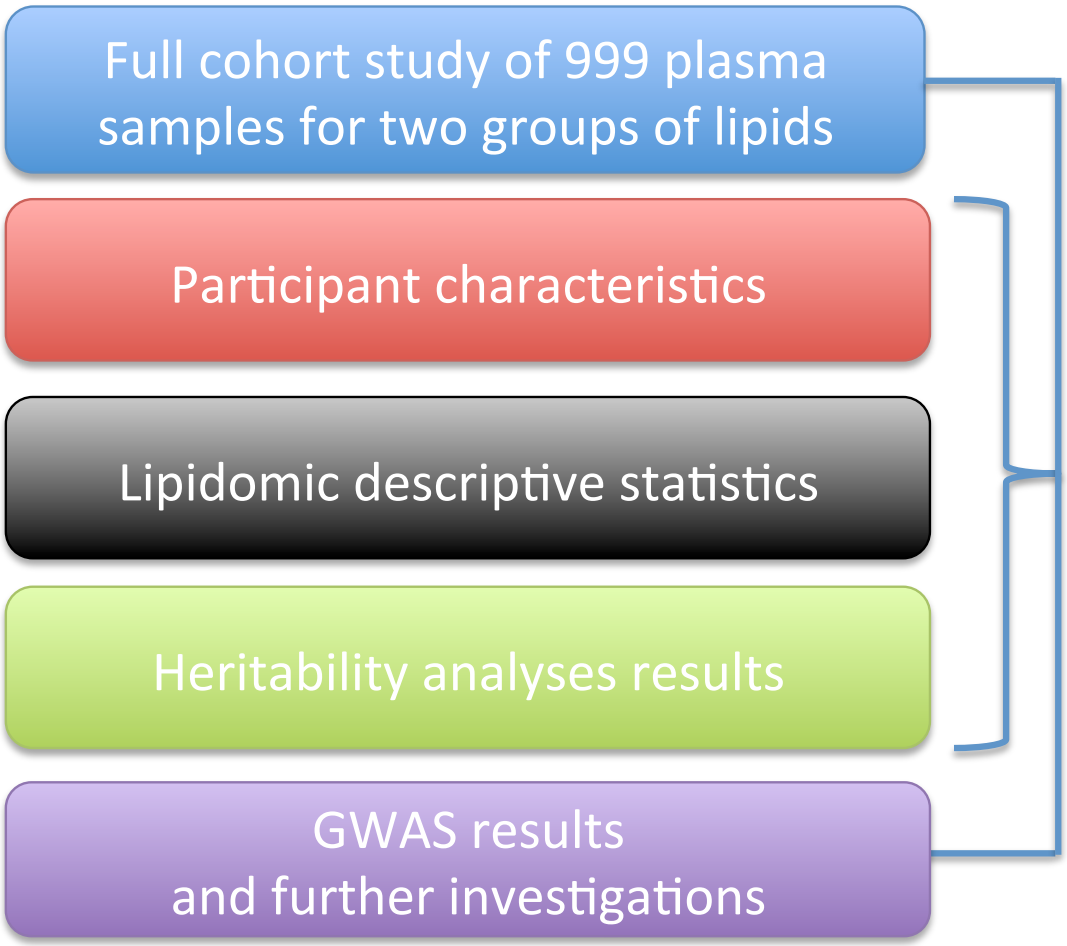
The omega-3 fatty acid ALA has been shown to be dependent on AMP-activated protein kinases to improve adipose tissue function (Zhou *et al.*, 2015), and such omega-3 PUFAs have been shown to antagonise inflammation via activation of AMP kinases in macrophage (Xue *et al.*, 2012). The association of the omega-3 trait with variants on chromosome 1 upstream to protein kinase AMP-activated non-catalytic subunit beta 2 (*PRKAB2*) may be of interest in future studies.

4.7. Conclusion

The aim of this study was to complete a range finding study of three classes of bioactive lipids measured in plasma to identify those lipids under particular genetic influence for full cohort analyses. The plasma samples with genotyping data available were used to analyse the heritability and initial GWAS of 60 bioactive lipid species and further calculated traits. NAE and CER species were identified as substantially heritable, with top SNP associations identified for two CER species with the *SPTLC3* locus. Conversely, analysis of the Eico and related species only identified four species that were significantly heritable and therefore the class was not taken forward for full cohort analysis. A GWAS was completed on the group of Eico lipids not taken forward for full cohort analyses, which suggested significant associations may be identified in future large cohort analyses.

Chapter 5

Heritability and family-based GWAS analyses of the *N*-acyl ethanolamine and ceramide lipidome



5.1. Introduction, aim and objectives

The range finding study (Chapter 4) identified that NAE and CER lipids were more heritable than the Eico lipids species. To identify the common DNA variants that influence plasma NAE and CER species, the lipids were analysed by GWAS of the full cohort including 999 participants from 196 families.

Results: NAE species

PEA, biosynthesised from the fatty acid palmitic acid, was the NAE species at highest concentration in the cohort. A positive association was identified with previously measured total cholesterol and NAE levels. All of the 11 NAE species were estimated as significantly heritable (45-82%) in this cohort of British Caucasians. PDEA had the highest estimated heritability but the relative importance and role of this lipid is unknown. Four NAEs (DHEA, LEA, PEA, and VEA), as well as the sum of all NAEs, associated at GWAS with SNPs in the gene encoding fatty acid amide hydrolase (*FAAH*), which catalyses the degradation of NAEs. As the intra-class correlation analysis showed that NAE species were all positively correlated, it is likely that in a larger sample size, all species would associate with this genomic locus.

Results: CER species

CER[N(24)S(18)], created using the fatty acid lignoceric acid, was the most abundant CER species in plasma. CER species were correlated based on the CER biosynthetic pathway; CER[NS] and CER[NDS] correlated and CER[AS] species were found amongst the CER[NS] species, which are biosynthesised in a similar manner. All of the CER species were estimated as significantly heritable (18-62%). At GWAS, seven CER[NS] and two CER[NDS] species significantly associated with SNPs in an intergenic region on chromosome 20, which are confirmed liver eQTLs of the gene encoding the third subunit of serine palmitoyltransferase (*SPTLC3*). A novel association was identified for CER[N(26)S(19)] at a locus on chromosome 6, upstream to the gene for protein CD83, an immunoglobulin membrane receptors found on blood cells.

The ratio of CER[NS]/CER[NDS], is indicative of delta 4-desaturase, sphingolipid 1 (DEGS1) activity. A set of SNPs in the upstream region of the *DEGS1* gene on

chromosome 1 associated with this ratio for CER[N(24)S(19)]/CER[N(24)DS(19)], and all significant SNPs were confirmed eQTLs of *DEGS1* in whole blood. This locus associated via 2SMR with numerous blood cell phenotypes in the UKBiobank data. A further set of SNPs upstream of the gene encoding sphingosine-1 phosphate phosphatase (*SGPP1*) associated with CER[N(24)S(16)]. This enzyme is involved in the recycling of CER[NS] species from sphingosine and sphingosine-1-phosphate (C18S1P) All significant SNPs at this locus were also associated via 2SMR with blood cell phenotypes. The association with the variants influencing CER traits and blood cell counts may provide initial evidence of interplay between CER and haematological phenotypes. The variants identified for NAE and CER species did not confirm the role of these lipids in CVD via 2SMR.

The objectives were as follows:

1. Analyse 800 plasma samples for the NAE and CER lipid species
2. Assess the heritability of 11 NAE and 30 CER species in this larger number of samples
3. Complete a genome-wide association study on the lipids and further calculated traits
4. Identify the main genomic loci influencing the plasma lipid species
5. Assess the significant associations via Ensembl and UCSC genome browser to identify the nearest gene, and GTEx to identify if the SNPs are eQTLs of genes in tissues related to plasma lipid metabolism
6. Assess the significant associations using two-sample Mendelian randomisation techniques, GWAS Catalog, and UK Biobank PheWAS analyses to identify causation in literature phenotypes and disease

5.2. Population Characteristics

Plasma samples of 999 participants from 196 British Caucasian families were included in the genetic analyses, a further 800 samples to the range finding study. The families consisted of 1-24 members (mean of 5 members) with plasma available for lipidomics analyses and genotyping data available (Figure 5-1). Participant descriptions are listed in Table 5.1.

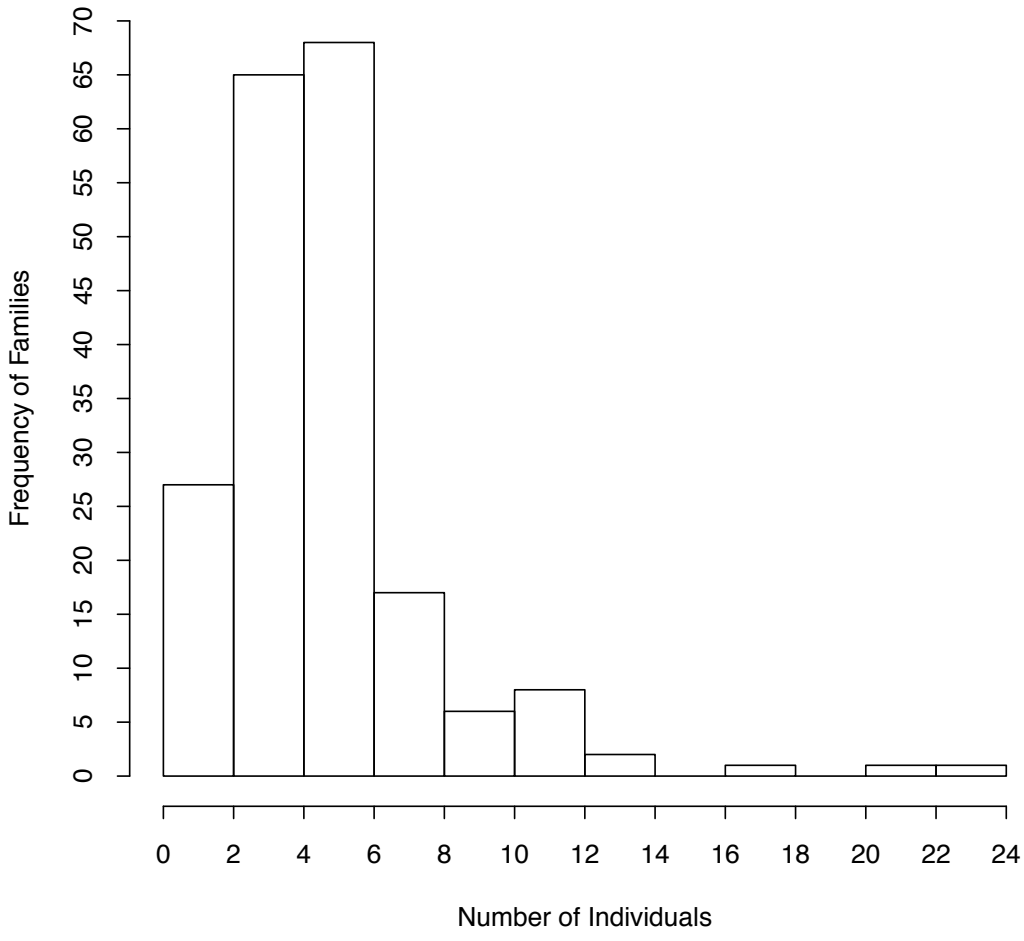


Figure 5-1: The frequency of individuals in each collected family

The graph depicts the range of individuals (1-24) with plasma available for lipidomics analyses and genotyping data available. 999 individuals were assessed from 196 families. The mean number of individuals from each family was 5.

Table 5.1: Summary statistics for the study participants.

Data shown as mean and standard deviation (SD) unless otherwise indicated; BMI, body mass index; WHR, waist-hip ratio; Mean Blood Pressure, mean of three readings taken in the clinic.

Trait	Mean (SD)
Gender	47% Male
Hypertensive	33%
Mean Blood Pressure	138/83 mmHg
Age (years)	49 (15)
BMI	26.04 (4.33)
WHR	0.86 (0.09)
Cholesterol (mmol/L)	5.61 (1.20)

5.3. Lipidomics descriptive statistics

The assay was run over 88 extraction batches and 24 mass spectrometry batches. 16% of the samples were noted as visually showing evidence of haemolysis or containing white blood cells. Of the 11 NAE species identified in plasma, PEA was at highest abundance (1.89 ± 1.36 ng/ml [mean ± SD]) (Figure 5-2). Of the 30 plasma CER species, CER[N(24)S(18)] was most abundant (128 ± 61 pmol/ml, Figure 5-3). Summary statistics of each lipid species can be found in Appendix Table 0.6. NAE species showed a positive association with total cholesterol, confirming the observation made in the smaller range finding study (Chapter 4), depicted in the Appendix (Table 0.7). Pooled quality control samples and analytical batch were identified as the most significant predictor covariates for both classes of lipids.

As the lipid mediators studied can exert individual bioactivities, all lipid species were treated uniquely for all analyses; the intra-class correlation analyses are summarised in Figure 5-4. CER[NS] and CER[NDS] correlated strongly with other lipid mediators of their respective classes, with CER[NS] more similar depending on the carbon length of the fatty acyl chain. The sphingosine base (C18S) and sphinganine (C18DS) grouped together. Sphingosine 1-phosphate (C18S1P) was separated out from the rest of the lipids, as did CER[A(24)S(18)], with the other CER[AS] species found

amongst the CER[NS]. NAE lipids showed a positive correlation with all other species of the same class.

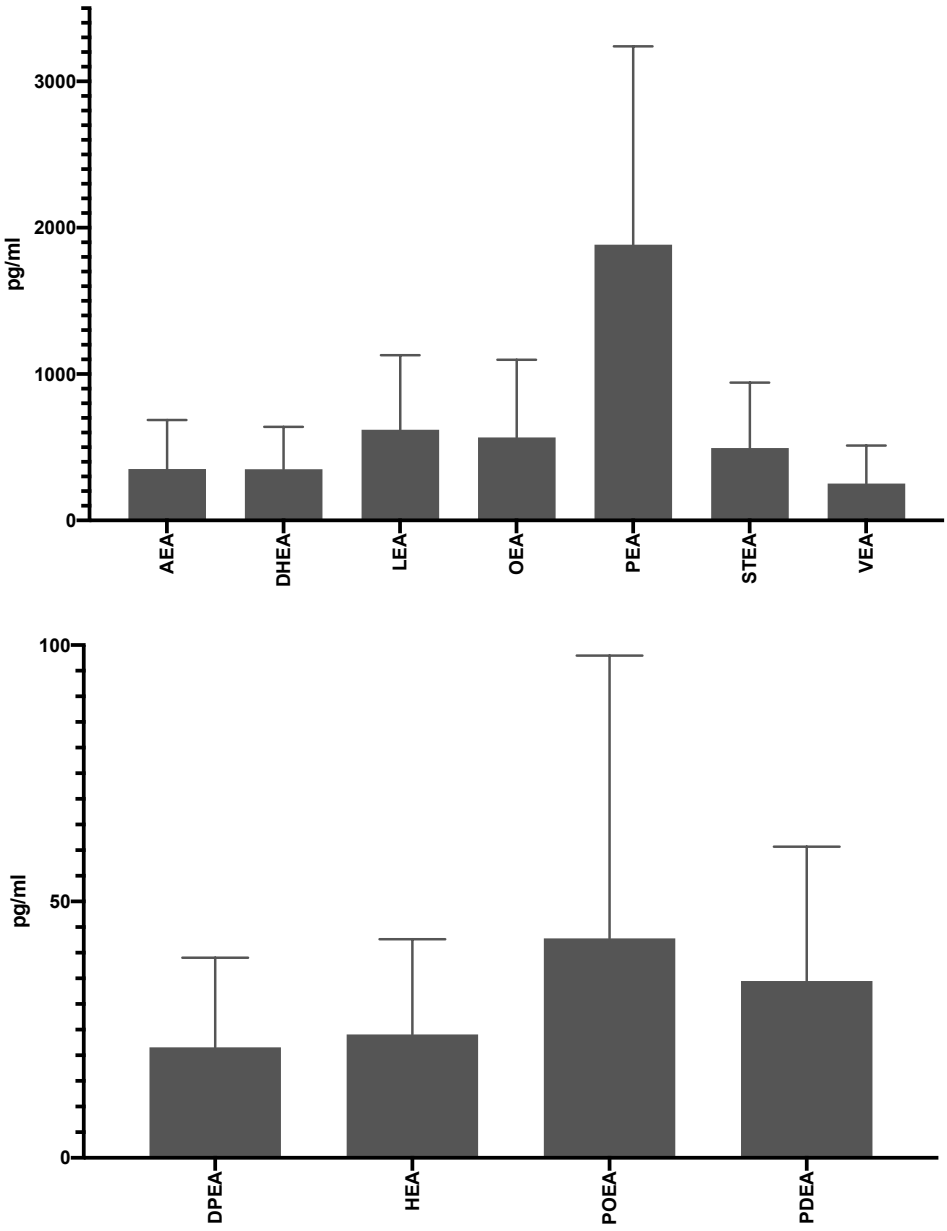
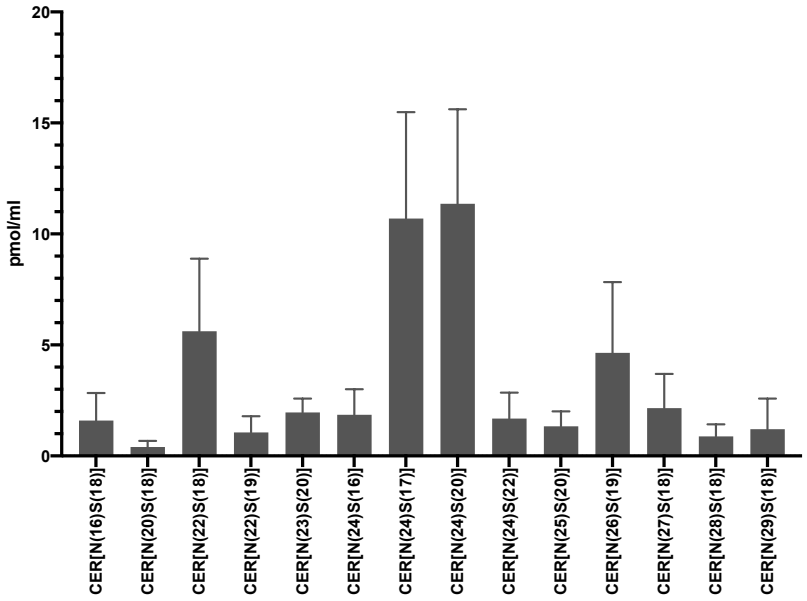


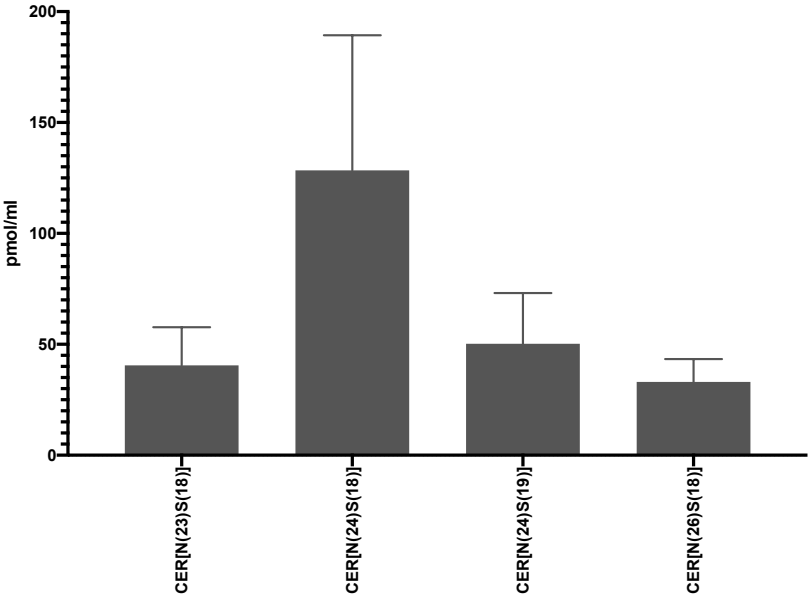
Figure 5-2: Plasma levels of NAE species

Data shown as mean (pg/ml) ± SD of 999 unadjusted, raw NAE plasma concentrations.

A **CER[NS]**

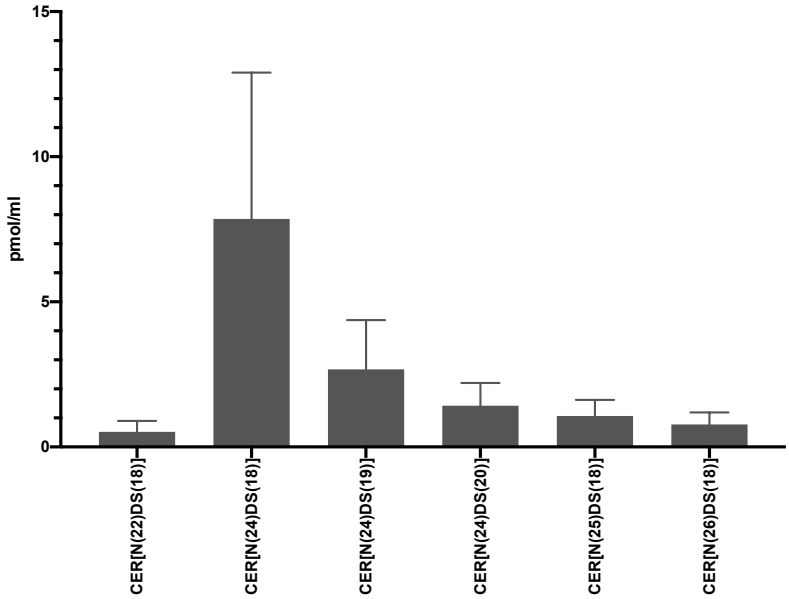


B **CER[NS] cont.**

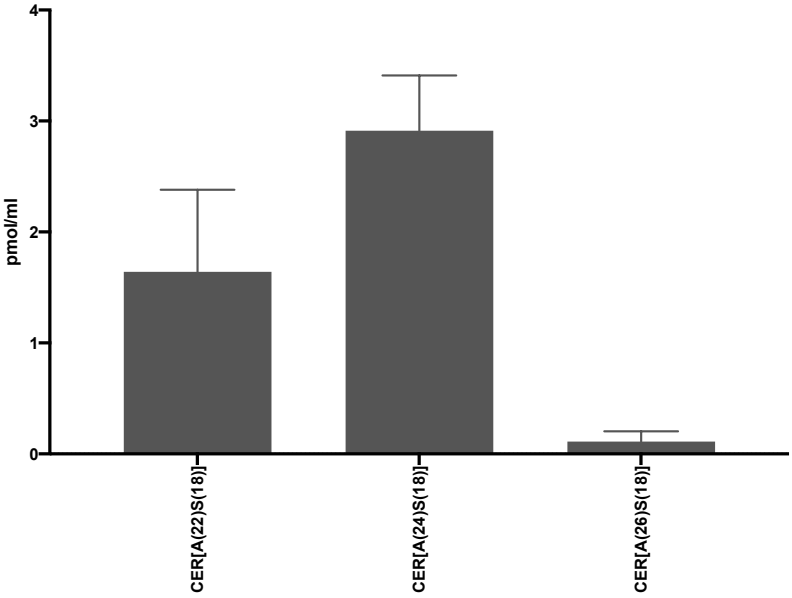


Data shown as mean (pmol/ml) ± SD of 999 unadjusted, raw CER plasma abundance.

C **CER[NDS]**



D **CER[AS]**



Data shown as mean (pmol/ml) ± SD of 999 unadjusted, raw CER plasma abundance.

E **Sphingoid base species (C18S)**

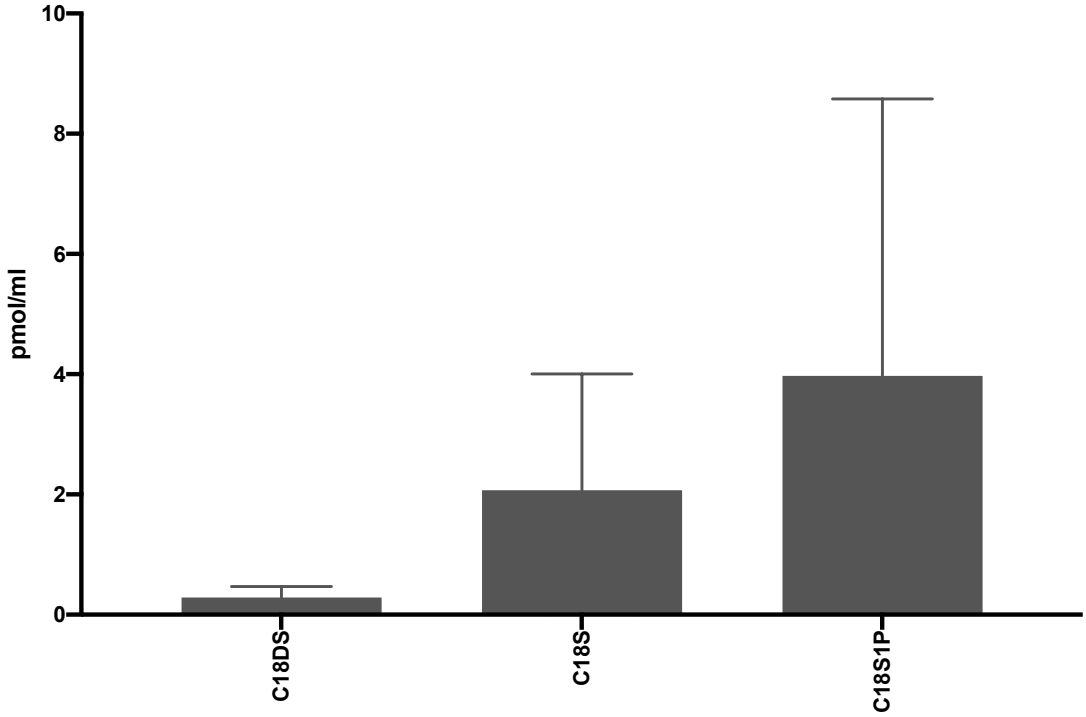
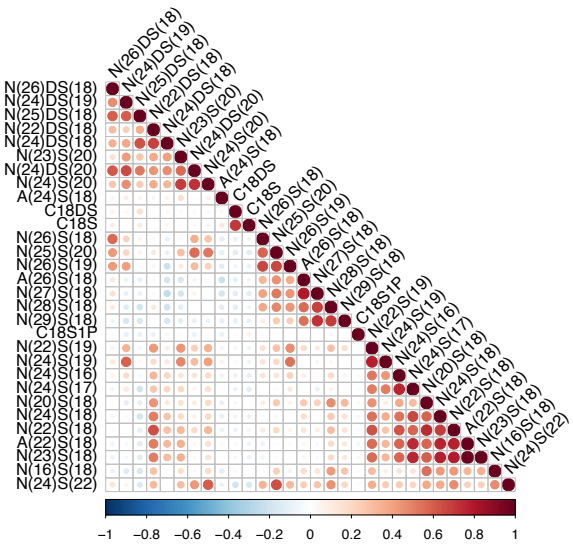


Figure 5-3: Plasma levels of A-B) CER[NS], C) CER[NDS], D) CER[AS], E) C18S species

Data shown as mean (pmol/ml) ± SD of 999 unadjusted, raw CER plasma abundance.

A



B

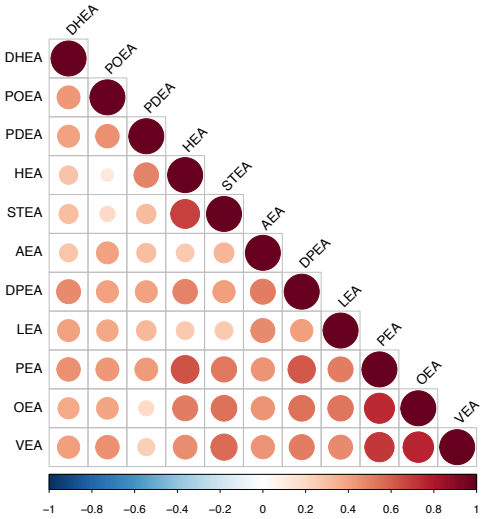


Figure 5-4: Intra-correlation analysis of ceramides, sphingoid bases, and related sphingolipids, and *N*-acyl ethanolamines

The figure depicts the assessment of the relatedness within the two classes of lipids (A: CER, B: NAE). The tool rquery.cormat was used in R, which takes into account strength of relationship (correlation coefficient; depicted as a scale of colours) and P-value (size of circle produced). The correlation was completed on the covariate-adjusted, standardised residuals used for genetic analyses.

5.4. NAE and CER lipid species are highly heritable

The NAE species had estimated heritabilities ranging from 45% to 82% ($P_{adj}<6.72x10^{-15}$), with pentadecanoyl ethanolamide (PDEA) having the highest estimated heritability. CER species showed a wide range in estimated heritability. Of the CER classes examined, CER[NS] species had heritabilities estimated between 18% - 62% ($P_{adj}<4.50x10^{-2}$), CER[NDS] species had estimated heritability of 32% - 52%, ($P_{adj}<3.00x10^{-11}$), while CER[AS], sphingosine-1-phosphate (C18S1P), sphingosine (C18S) and dihydrosphingosine (C18DS) were also significantly heritable ($P_{adj}<0.05$). The correlation between the estimates calculated via pedigree-based or SNP-based software was R=0.99, revealing a strong relationship between heritability estimates created via differing software and their respective methods (Figure 5-5 and Table 5.2).

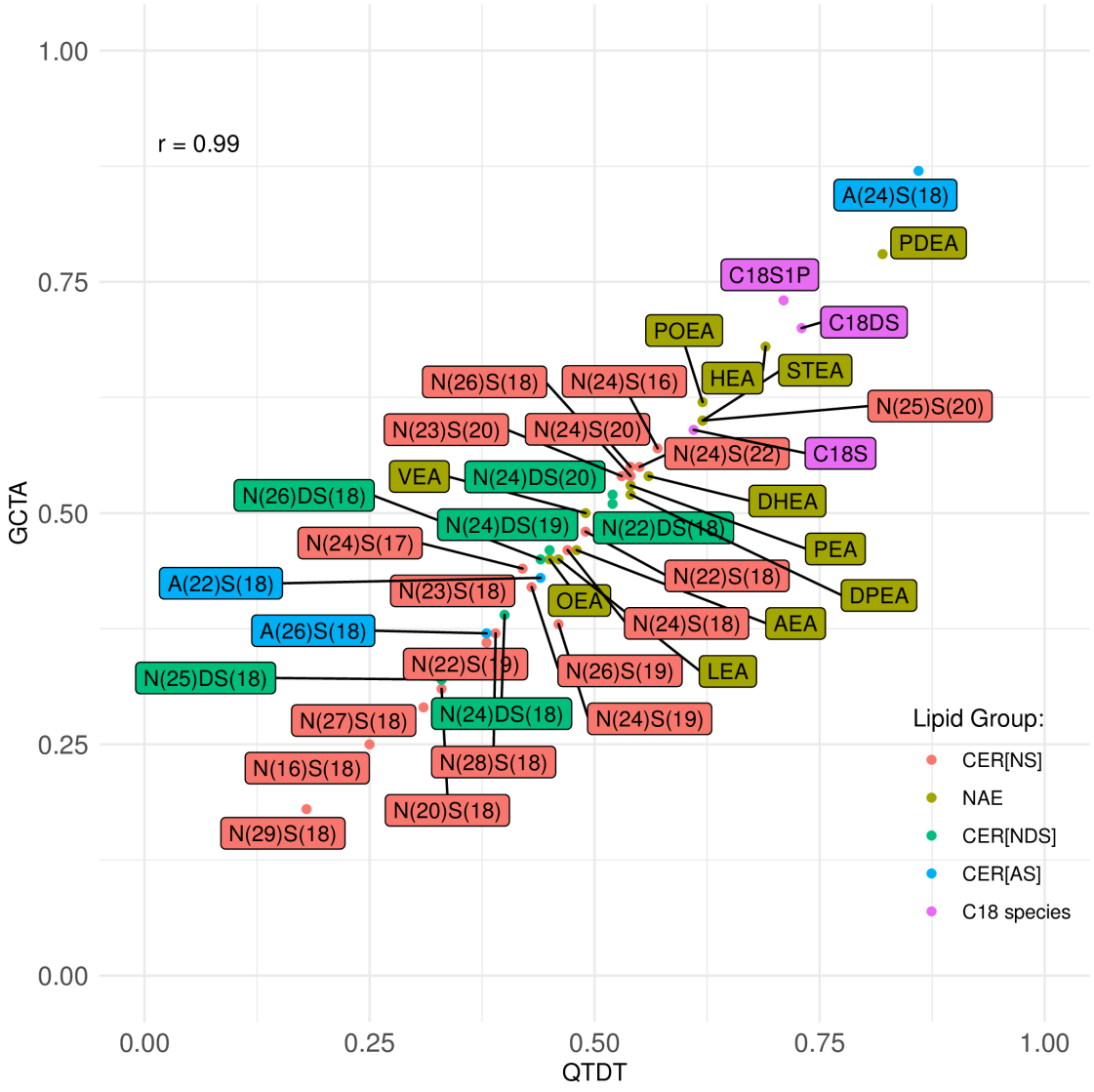


Figure 5-5: Heritability estimates of NAE and CER found in human plasma.

Heritability was estimated for each lipid species using SNP-based GCTA software (y-axis) and reported pedigree-based QTDT software (x-axis). The species are color-coded by group; CER[NS], non-hydroxy CER species; CER[NDS], non-hydroxy dihydroceramide species; CER[AS], alpha-hydroxy CER species; C18 species, sphingosine bases; NAE, *N*-acyl ethanolamine species. Values for this graph are described in Table 5.2.

Table 5.2: Heritability estimates of plasma CER and NAE lipid species

Heritability was estimated by GCTA SNP-based software using the genotyping data and a genetic relationship matrix, and QTDT pedigree-based software using kinship coefficients based on the reported relationships for A) CER and B) NAE lipid species. h^2, estimated additive heritability; SE, standard error; ChiSq, chi-squared; n, number of individuals included in analysis after outlier removal; P-adj, P-value adjusted for multiple testing by Bonferroni methods (30 for CER measured, 11 for NAE measures).

A

CER	QTDT				GCTA			
Lipid	h^2	ChiSq	n	P_{adj}	h^2	SE	n	P_{adj}
CER[A(22)S(18)]	0.44	84.18	993	1.50E-18	0.43	0.056928	993	1.83E-14
CER[A(24)S(18)]	0.86	404.46	994	1.80E-88	0.87	0.035209	994	1.83E-14
CER[A(26)S(18)]	0.38	48.94	868	9.00E-11	0.37	0.063439	868	5.86E-11
CER[C18DS]	0.73	211.87	992	1.50E-46	0.70	0.051171	992	1.83E-14
CER[C18S]	0.61	124.09	990	2.40E-27	0.59	0.059826	990	1.83E-14
CER[C18S1P]	0.71	177.98	970	3.00E-39	0.73	0.054038	970	1.83E-14
CER[N(16)S(18)]	0.25	22.1	993	9.00E-05	0.25	0.058959	993	3.43E-05
CER[N(20)S(18)]	0.33	49.56	991	6.00E-11	0.31	5.74E-02	991	1.01E-10
CER[N(22)DS(18)]	0.52	103.99	989	6.00E-23	0.51	0.058143	989	1.83E-14
CER[N(22)S(18)]	0.49	97.58	993	1.50E-21	0.48	0.058559	993	1.83E-14
CER[N(22)S(19)]	0.38	64.1	992	3.00E-14	0.36	0.057068	992	1.83E-13
CER[N(23)S(18)]	0.40	67.73	992	6.00E-15	0.39	0.057396	992	3.33E-15
CER[N(23)S(20)]	0.53	91.26	994	3.00E-20	0.54	0.061473	994	1.83E-14
CER[N(24)DS(18)]	0.40	69.41	992	2.40E-15	0.39	0.059638	992	1.83E-14
CER[N(24)DS(19)]	0.45	72.76	994	3.00E-16	0.46	0.062131	994	1.83E-14
CER[N(24)DS(20)]	0.52	106.58	993	1.80E-23	0.52	0.059327	993	1.83E-14
CER[N(24)S(16)]	0.57	150.02	992	6.00E-33	0.57	0.055013	992	1.83E-14
CER[N(24)S(17)]	0.42	80.42	994	9.00E-18	0.44	0.057249	994	1.83E-14
CER[N(24)S(18)]	0.47	100.76	991	3.00E-22	0.46	0.055695	991	1.83E-14
CER[N(24)S(19)]	0.46	72.38	991	6.00E-16	0.38	0.061448	993	2.98E-13
CER[N(24)S(20)]	0.54	102.13	996	1.50E-22	0.55	0.059431	996	1.83E-14
CER[N(24)S(22)]	0.55	92.78	992	1.80E-20	0.55	0.061416	992	1.83E-14
CER[N(25)DS(18)]	0.33	50.59	993	3.00E-11	0.32	0.058387	993	1.69E-11
CER[N(25)S(20)]	0.62	126.14	994	9.00E-28	0.60	0.05895	994	1.83E-14
CER[N(26)DS(18)]	0.44	88.36	993	1.50E-19	0.45	0.059183	993	1.83E-14
CER[N(26)S(18)]	0.54	115.12	995	2.10E-25	0.54	0.059431	995	1.83E-14
CER[N(26)S(19)]	0.43	57.35	991	1.20E-12	0.42	0.063209	991	9.73E-13
CER[N(27)S(18)]	0.31	32.14	993	3.00E-07	0.29	0.060025	993	8.93E-08
CER[N(28)S(18)]	0.39	48.34	994	1.20E-10	0.37	0.062254	994	1.19E-10
CER[N(29)S(18)]	0.18	10.13	990	4.50E-02	0.18	0.059202	990	1.71E-02

B

NAE	QTDT				GCTA			
Lipid	h^2	ChiSq	n	P_{adj}	h^2	SE	n	P_{adj}
AEA	0.48	86.58	994	1.10E-19	0.46	0.05962	994	6.72E-15
DHEA	0.56	131.88	990	2.20E-29	0.54	0.056986	990	6.72E-15
DPEA	0.54	99.41	986	2.20E-22	0.52	0.060463	986	6.72E-15
HEA	0.69	231.86	964	2.20E-51	0.68	0.04969	964	6.72E-15
LEA	0.46	116.79	994	3.30E-26	0.45	0.053217	994	6.72E-15
OEA	0.45	84.31	994	4.40E-19	0.45	0.057784	994	6.72E-15
POEA	0.62	121.33	961	3.30E-27	0.62	5.94E-02	961	6.72E-15
PEA	0.54	140.13	993	2.20E-31	0.53	0.05411	993	6.72E-15
PDEA	0.82	339.83	954	7.70E-75	0.78	0.044354	954	6.72E-15
STEA	0.62	209.21	994	2.20E-46	0.60	0.049033	994	6.72E-15
VEA	0.49	105.02	994	1.10E-23	0.50	0.057676	994	6.72E-15

Heritability was estimated by GCTA SNP-based software using the genotyping data and a genetic relationship matrix, and QTDT pedigree-based software using kinship coefficients based on the reported relationships for A) CER and B) NAE lipid species. h^2, estimated additive heritability; SE, standard error; ChiSq, chi-squared; n, number of individuals included in analysis after outlier removal; P-adj, P-value adjusted for multiple testing by Bonferroni methods (30 for CER measured, 11 for NAE measures).

5.5. Genome-wide association study of *N*-acyl ethanolamines

There were conventionally GWAS significant ($P<5x10^{-8}$) associations between four NAEs (N-docosahexaenoyl ethanolamide, DHEA; N-linoleoyl ethanolamide, LEA; N-palmitoyl ethanolamide PEA; vaccinoyl ethanolamide, VEA), as well as the sum of all NAEs (sumEA), with SNPs in the gene encoding fatty acid amide hydrolase (*FAAH*; Figure 5-6), which catalyses the degradation of NAEs (Chapter 1, Section 1.2.1.). The leading SNP is a missense variant (rs324420; C385A; P129T) and eQTL of *FAAH* in multiple tissues including whole blood (Appendix Table 0.9). Presence of the missense variant causes FAAH to display normal catalytic properties but decreased cellular stability (Chiang *et al.*, 2004) and enhanced sensitivity of the enzyme to proteolytic degradation (Sipe *et al.*, 2002). The magnitude of the genetic effect was considerable; the A allele of the lead SNP rs324420 increased plasma NAE species (e.g. PEA; beta=0.30, SE = 0.05). A LocusZoom plot and Manhattan plot of the association with the lipid PEA is depicted in Figure 5-7, showing the strong

association of this genomic locus with the lipid. Genomic inflation factors for the GWAS of both NAE and CER are presented in Appendix Table 0.8, with Quantile-Quantile plots, Manhattan plots, and trait distributions presented in Appendix Figures 0-2 - Figure 0-4.

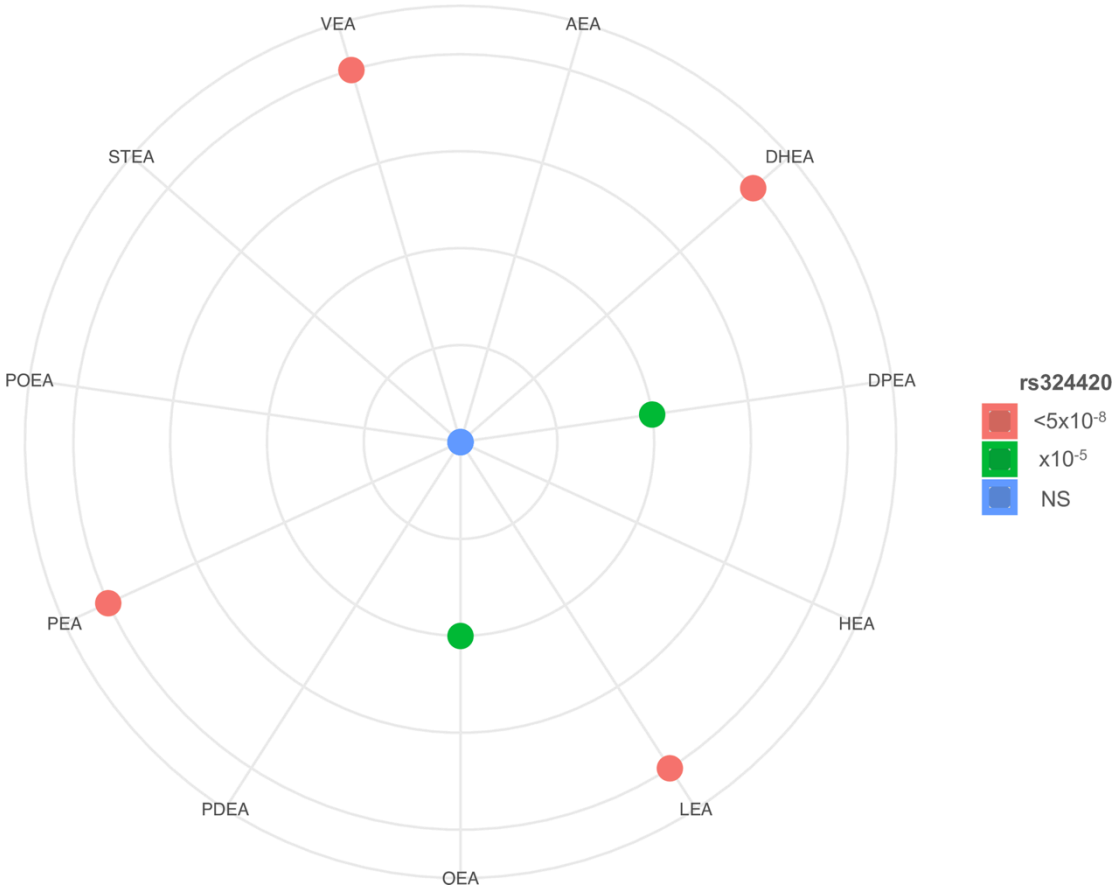
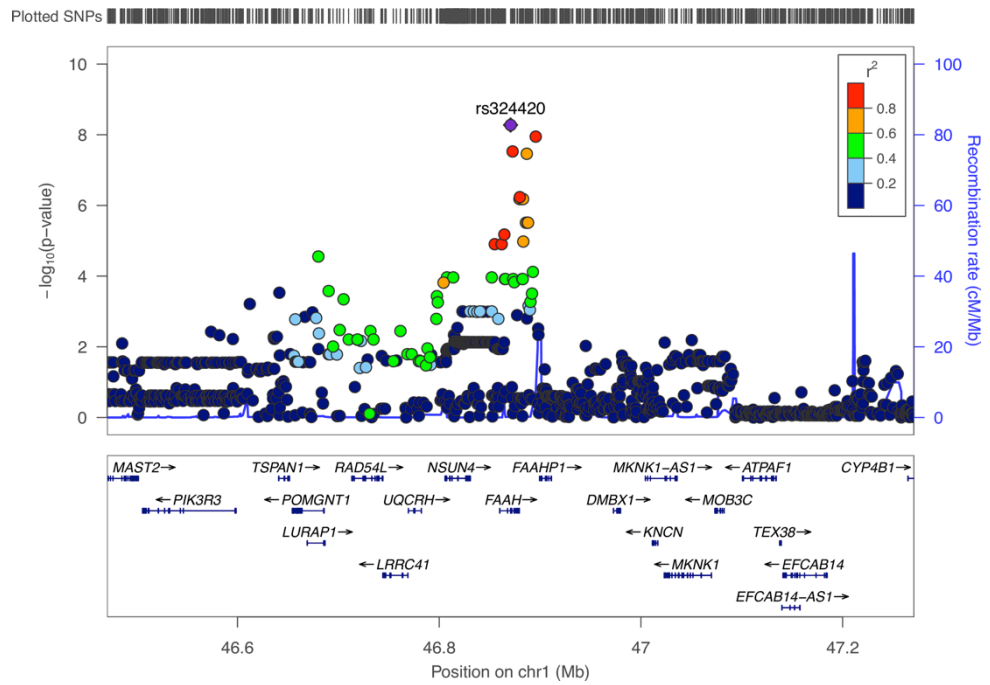


Figure 5-6: Family-based GWAS results for *N*-acyl ethanolamines and the lead SNP rs680379 in fatty acid amide hydrolase (*FAAH*).

The radar plot depicts the P-value for association between the lead SNP and eQTL of *FAAH* (rs324420) and each NAE species. The P-values were grouped into "$<5x10^{-8}$" ($P<5x10^{-8}$, outermost ring), "$x10^{-6}$" ($P=5.0x10^{-8}$ - $9.9x10^{-6}$ [of which there are no NAE species, included for reference]), "$x10^{-5}$" ($1.0x10^{-5}$ - $9.9x10^{-5}$), and "NS" (not significant) at the center of the radar.

A



B

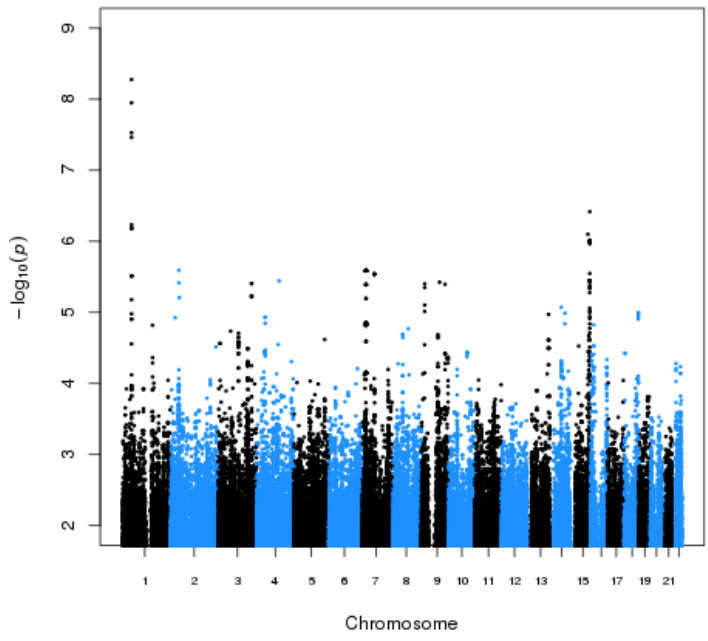


Figure 5-7: The association of PEA with FAAH SNP rs324420 on chromosome 1

A) LocusZoom plot of the *FAAH* loci association B) Manhattan plot of the 22-chromosome GWAS results for PEA highlighting a stack of significant SNPs on chromosome 1 at *FAAH*.

5.6. Genome-wide association study of ceramides and related sphingolipids

Seven CER[NS] and two CER[NDS] species were significantly associated with SNPs in an intergenic region on chromosome 20 (Figure 5-8). An example of the identified association between the locus and CER[N(24)S(19)] is shown in Figure 5-9, with further details of all GWAS results in Appendix Table 0.10. Assessing the SNPs using GTEx confirmed them as liver eQTLs (Appendix Table 0.11) found 20,000 bases downstream of the gene encoding the third subunit of serine palmitoyltransferase (*SPTLC3*; Figure 5-9), which catalyses the rate-limiting step of CER biosynthesis (Perry *et al.*, 2000). Therefore, the SNPs are associated with differences in the expression of the *SPTLC3* gene in the liver, which contributes to plasma levels of CER species. The SNPs had considerable phenotypic effects, for example the A allele of the SNP rs680379, identified previously to be associated with blood CER species in literature, was associated here with a per-allele increase in CER (e.g. CER[N(24)S(19)]; beta = 0.46, SE = 0.04). Furthermore, the summed total of all CER species with 24-carbon non-hydroxy fatty acids, and, independently, those with 19- and 20-carbon sphingosine bases, were found associated with the same SNPs at the *SPTLC3* locus (Appendix Table 0.12).

A novel association was identified for CER[N(26)S(19)] at a locus on chromosome 6, upstream to the gene for protein CD83, an immunoglobulin membrane receptor found on blood cells (e.g. rs6940658, $P=2.07 \times 10^{-8}$; depicted in Figure 5-10, with further details in Appendix Table 0.10) (Ju *et al.*, 2016).

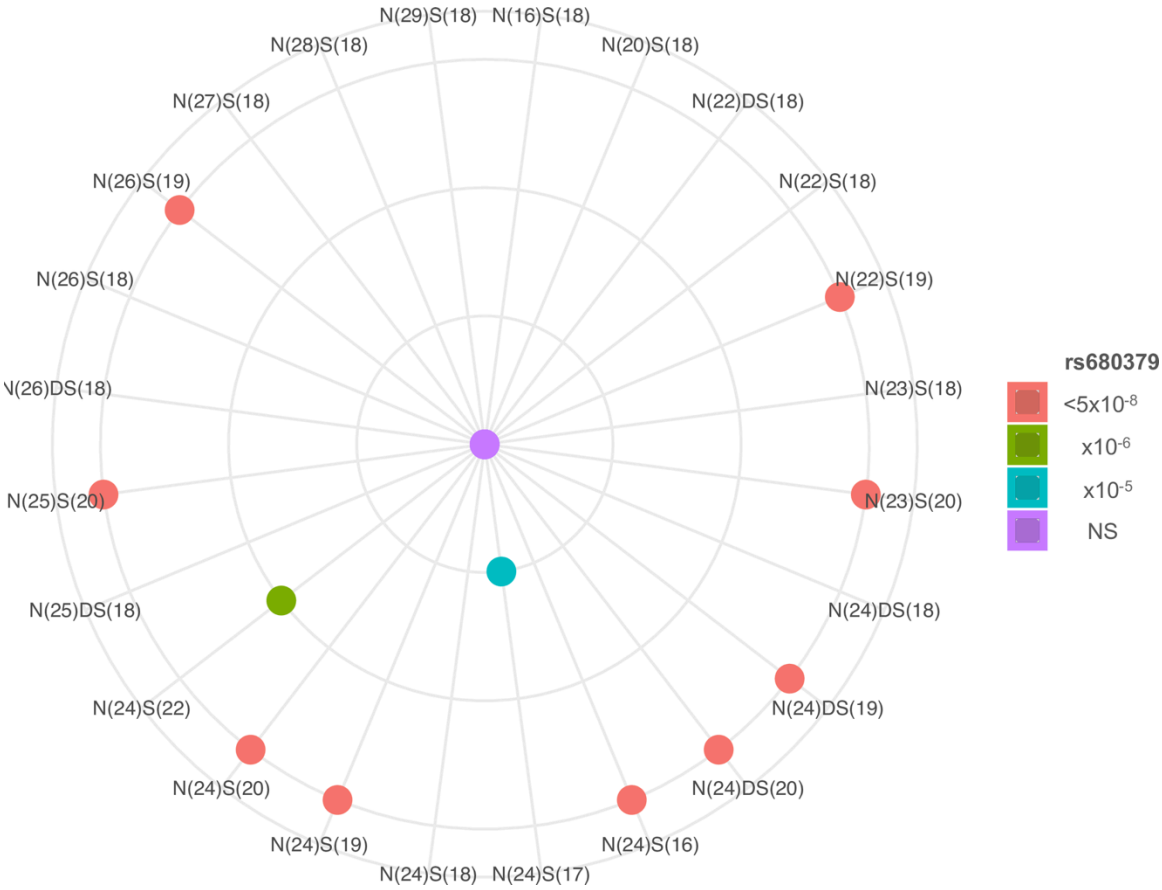
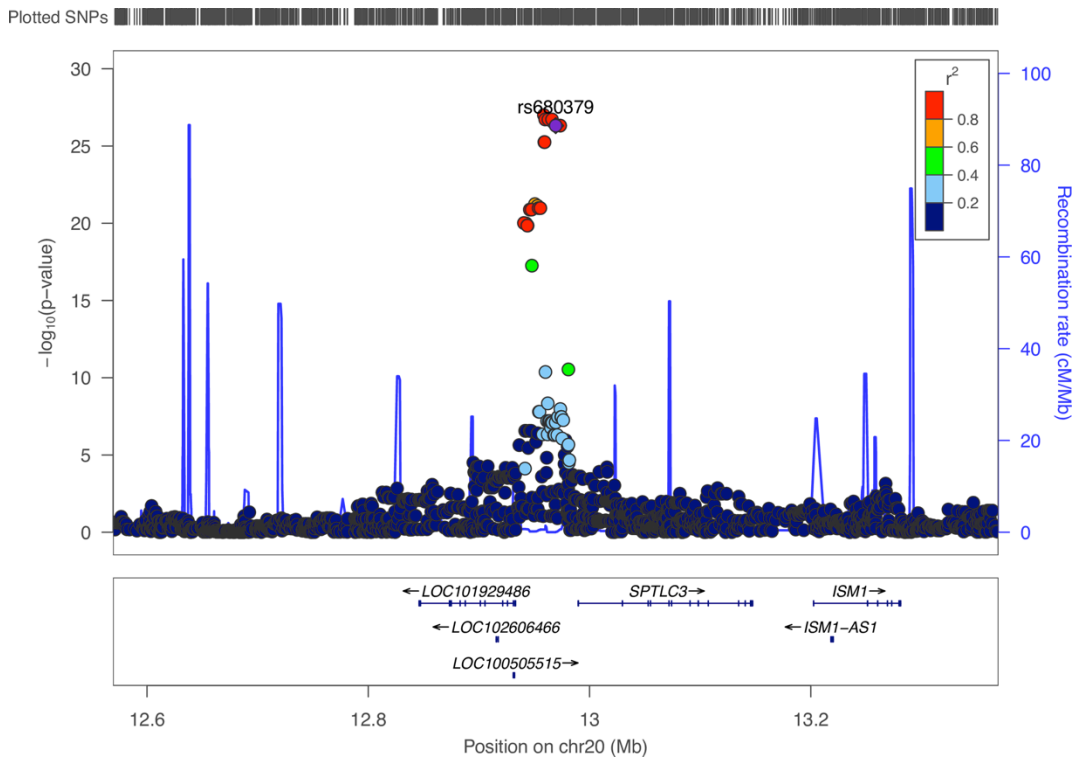


Figure 5-8: Family-based GWAS results for CER[NS] and precursor CER[NDS] with an exemplar SNP rs680379 in serine palmitoyltransferase (*SPTLC3*).

The radar plot depicts the P-value for association between the lead SNP and liver eQTL of *SPTLC3* (rs680379) with CER species. The P-values were grouped into "$<5x10^{-8}$" ($P<5x10^{-8}$, outermost ring), "$x10^{-6}$" ($P=5.0x10^{-8}$ - $9.9x10^{-6}$, "$x10^{-5}$" ($1.0x10^{-5}$ - $9.9x10^{-5}$), and "NS" (not significant) at the center of the radar.

A



B

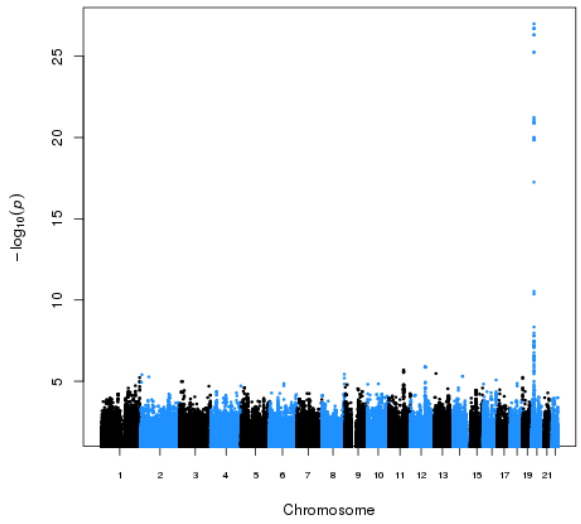
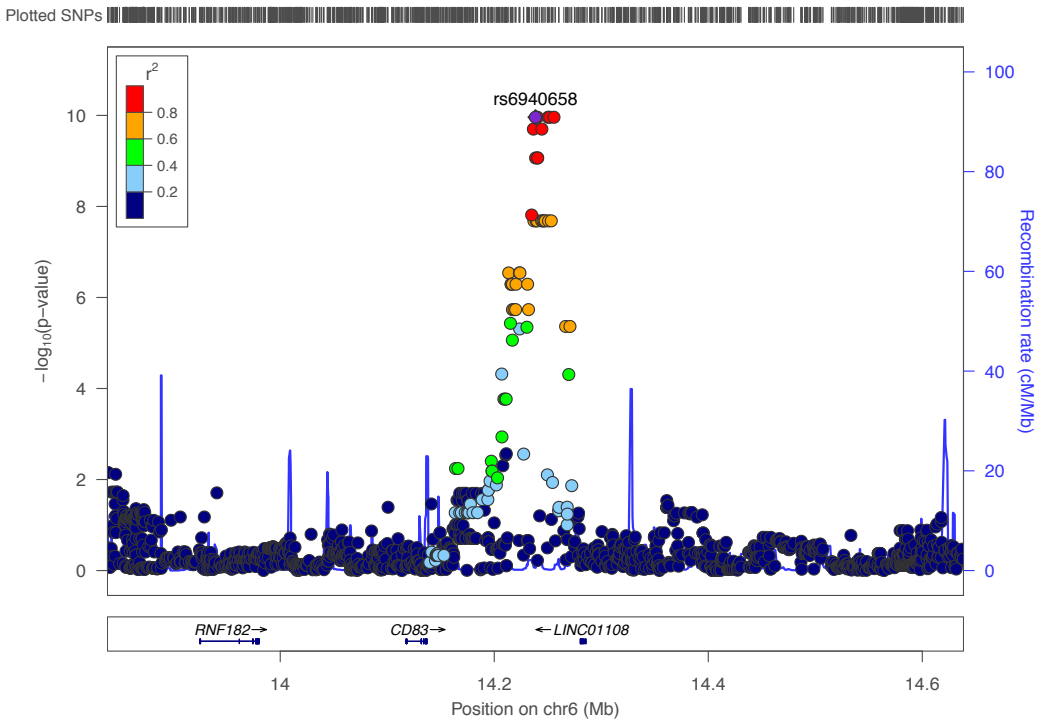


Figure 5-9: The association of CER[N(24)S(19)] with *SPTLC3* SNP rs680379

A) LocusZoom plot of the *SPTLC3* association B) Manhattan plot of the 22-chromosome GWAS results for the CER highlighting a stack of significant SNPs on chromosome 20 at *SPTLC3*.

A



B

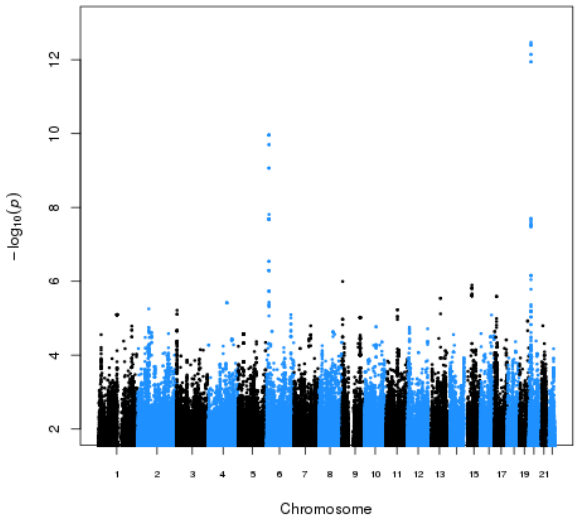


Figure 5-10: The association of CER[N(26)S(19)] to the CD83 locus.

A) LocusZoom plot of the *CD83* association B) Manhattan plot of the 22-chromosome GWAS results for the CER highlighting a stack of significant SNPs on chromosome 20 at *SPTLC3* and chromosome 6 at *CD83*. This finding was supported by three significant genotyped SNPs.

5.7. The association of CER and related traits with hematological phenotypes

The ratio of CER[NS]/CER[NDS] is indicative of delta 4-desaturase, sphingolipid 1 (*DEGS1*) activity which adds a double bond to the structure of CER[NDS] species to create CER[NS] species (Chapter 1, Section 1.2.2.). A set of SNPs in the upstream region of the *DEGS1* gene on chromosome 1 associated with this ratio, e.g. the product-precursor ratio of CER[N(24)S(19)]/CER[N(24)DS(19)] ($P=4.34x10^{-8}$) (Figure 5-11), with further details in Appendix Table 0.10 - Table 0.13. All significant SNPs were confirmed eQTLs of *DEGS1* in whole blood.

The Gene Atlas Browser of PheWAS in the UK Biobank study (geneatlas.roslin.ac.uk) was used to assess the association of significant SNPs identified here with the extensive number of phenotypes measured in the UK Biobank cohort. The GWAS Catalog, a database of previously published GWAS, was also used to assess the signficiant GWAS findings (https://www.ebi.ac.uk/gwas/). The SNPs identified at the *DEGS1* locus associated with numerous blood cell phenotypes measured in the UKBiobank data and found on GWAS Catalog, for example platelet and white blood cell traits, implicating the creation of CER[NS] species from CER[NDS] with alterations in blood cells (Table 5.3 and Appendix Table 0.13).

A further set of SNPs upstream of the gene encoding sphingosine-1 phosphate phosphatase (*SGPP1*) were associated with CER[N(24)S(16)] (e.g. rs7160525, $P=5.67x10^{-10}$; Figure 5-12). This enzyme is involved in the creation of CER[NS] species from sphingosine and sphingosine-1-phosphate (C18S1P) (Chapter 1, Section 1.2.2.). All significant SNPs at this locus were also associated with blood cell phenotypes identified from the UK Biobank data and found on the GWAS Catalog (Table 5.3). This association of SNPs influencing the creation of CER[NS] species via the de novo biosynthetic pathway (via *DEGS1*)and recycling from other sphingolipid mediators (via *SGPP1*), implicates CER[NS] species in the alteration of multiple blood cell phenotypes.

Of note, the identified SNPs for the major genomic loci influencing the NAE and CER pathways (*FAAH* and *SPTLC3*, respectively) were not found to have associated previously with CVD traits nor other previously studied phenotypes in the UK Biobank data nor the GWAS Catalog database.

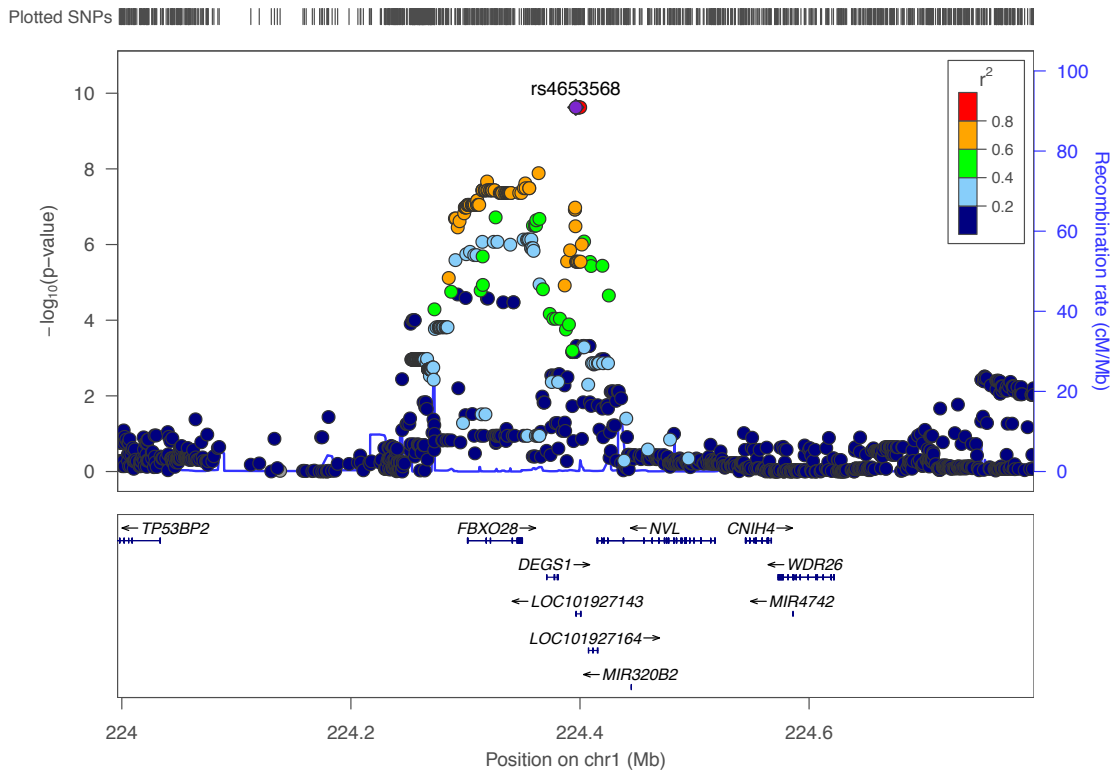


Figure 5-11: The association of the ratio of CER[N(24)S(19)]/CER[N(24)DS(19)] with the locus of delta 4-desaturase, sphingolipid 1 (*DEGS1*) on chromosome 1.

LocusZoom plot of the *DEGS1* association. This finding was supported by two significant genotyped SNPs.

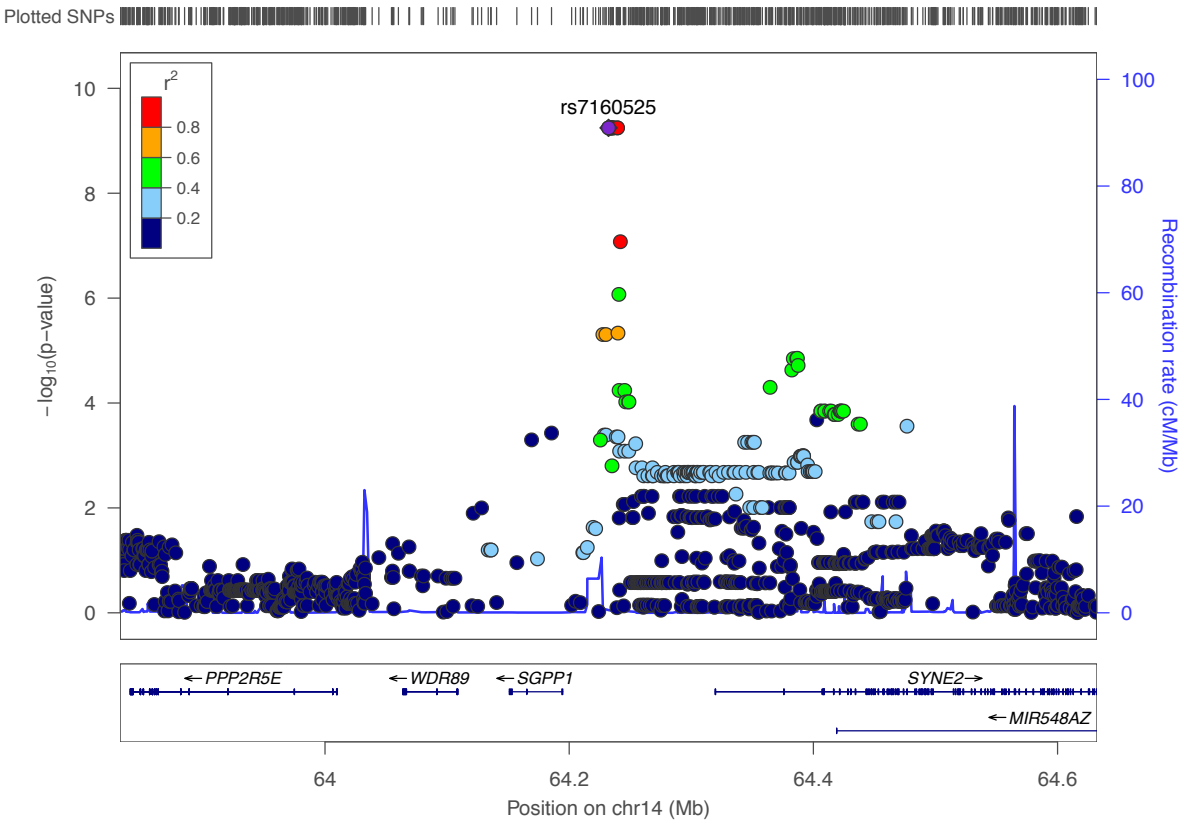


Figure 5-12: The association of CER[N(24)S(16)] to the locus at sphingosine-1-phosphate phosphatase 1 (*SGPP1*) on chromosome 14.

LocusZoom plot of the *SGPP1* association. This finding was supported by rs7157785, a GWAS significant genotyped SNP.

Trait	Effect allele	beta	P_{adj}	MAF	Study
DEGS1; rs12038372					
N24S19ratio	T	0.26	1.30E-08	0.41	HTO
Mean platelet (thrombocyte) volume	T	0.01	3.39E-18	0.42	UKB
Neutrophil count	T	-0.02	1.14E-09	0.42	UKB
White blood cell (leukocyte) count	T	-0.02	3.19E-09	0.42	UKB
SGPP1; rs7160525					
CER[N(24)S(16)]	A	0.36	5.67E-10	0.14	HTO
Mean platelet (thrombocyte) volume	A	-0.02	3.28E-29	0.16	UKB
Red blood cell (erythrocyte) distribution width	A	0.02	5.96E-14	0.16	UKB
Platelet count	A	0.89	5.56E-13	0.16	UKB
High light scatter reticulocyte percentage	A	-0.003	1.27E-12	0.16	UKB
High light scatter reticulocyte count	A	-0.0001	6.69E-11	0.16	UKB
Immature reticulocyte fraction	A	-0.0009	1.06E-10	0.16	UKB
Reticulocyte percentage	A	-0.007	1.91E-08	0.16	UKB

Table 5.3: UK Biobank PheWAS results for the significant SNPs associated with the ratio CER[N(24)S(19)]/CER[N(24)DS(19)] and CER[N(24)S(16)].

The lipid traits are in bold and the blood cell traits are in plain text. The table depicts the results from this lipid GWAS (HTO) and the GWAS results of blood cell traits from the UK Biobank Gene Atlas database (UKB).

5.8. Confirmation of the interplay between CER and hematological phenotypes, but not cardiovascular disease, through two-sample Mendelian randomisation

The significant SNPs identified at *SGPP1*, *DEGS1,* and *SPTLC3* were used as instruments for two-sample Mendelian randomisation analyses (2SMR). The GWAS results from a published blood cell count GWAS was used as the outcome variables (Astle *et al.*, 2016), to assess the causality of the CER lipids on multiple blood cell traits via the SNPs as instruments in the analysis. The SNPs at *SGPP1* and *DEGS1* were identified as significant ($P<0.05$) in influencing platelet, red blood cell, and white blood cell traits before adjustment for multiple testing, however after adjustment for 71 test performed (Appendix Table 0.13), only associations with CER[N(24)S(16)] and blood cell traits remained significant ($P<0.05$; summarised in Table 5.4).

The variants identified in *SGPP1* which significantly associated with CER[N(24)S(16)] at GWAS, showed a negative relationship between the CER and mean platelet volume, reticulocyte count and percentage/fraction of red blood cells, and a positive relationship with platelet and lymphocyte counts, and red cell distribution width. This may suggest that this species of CER alters the level of blood cell counts, or the blood cells increase the levels of this CER in circulation, when recycled from other lipid mediators via *SGPP1*.

As the lipids have been implicated in both CVD and type-2 diabetes, 2SMR assessment of coronary heart disease using a previously published large-scale GWAS study of the disease (Nikpay *et al.*, 2015) was undertaken for the significant SNPs identified for *SPTLC3*, *FAAH*, *SGPP1*, and *DEGS1*, and the results were not significant for a causal role of the lipid species measured here in CVD. The association of CER via *SPTLC3* was also not significant for type-2 diabetes (Mahajan *et al.*, 2014).

ID	Outcome	Exposure	beta	SE	Padj
1247	Mean platelet volume	CER[N(24)S(16)]	-0.08	0.014	8.13E-07
1251	Platelet count	CER[N(24)S(16)]	0.06	0.014	3.35E-03
1259	High light scatter reticulocyte count	CER[N(24)S(16)]	-0.06	0.014	8.68E-04
1260	High light scatter reticulocyte percentage of red cells	CER[N(24)S(16)]	-0.07	0.014	7.22E-05
1267	Reticulocyte fraction of red cells	CER[N(24)S(16)]	-0.06	0.014	4.40E-04
1269	Red cell distribution width	CER[N(24)S(16)]	0.08	0.013	1.62E-08
1270	Reticulocyte count	CER[N(24)S(16)]	-0.05	0.014	1.80E-02
1275	Lymphocyte counts	CER[N(24)S(16)]	0.05	0.014	2.54E-02
1276	Immature fraction of reticulocytes	CER[N(24)S(16)]	-0.05	0.013	6.94E-03

Table 5.4: Significant results from the two-sample Mendelian randomisation analysis.

The table depicts the significant association of CER[N(24)S(16)] after adjustment for 71 tests as exposures to multiple blood cell count outcomes via 2SMR analysis. More information on the study (via id) and outcome can be found in Appendix Table 0.13. The beta and standard error (SE) values describe the relationship between the CER traits and blood cell phenotypes.

5.9. Discussion

5.9.1. Participants are at increased cardiovascular risk

The full cohort of 999 participants was balanced by gender (47% male) and a mean age of 49 years old. Healthy blood pressure is considered less than 120/80 mmHg and due to the ascertainment strategy enriching for hypertension, the mean blood pressure reading of the cohort was raised at 138/83 mmHg. Other cardiovascular risk factors were also raised; a BMI of 24.9 is considered the upper limit of a healthy BMI and the mean BMI recorded for the cohort was 26.0, the mean WHR was at the upper limit of the healthy boundary for both sexes (0.86; healthy being less than 0.9 for men and 0.85 for women), and the mean total cholesterol was raised at 5.61 mmol/L, where less than 5.00 mmol/L is considered healthy [NHS website: https://www.nhs.uk/, date assessed 30/09/19]. Separate determinations of HDL and LDL cholesterol were not available for the cohort.

5.9.2. The most abundant plasma CER and NAE species

PEA was the NAE species at highest concentration in the cohort. PEA has signalling properties in anti-inflammation (Darmani *et al.*, 2005) and anti-nociception (Calignano *et al.*, 2001; M. Keppel Hesselink, 2012) through the peroxisome proliferator-activated receptor alpha (PPARα) (Lo Verme *et al.*, 2005), G-protein coupled receptors GPR55 and GPR119 (Godlewski *et al.*, 2009), and the transient receptor potential vanilloid type 1 (TRPV1) (Ho *et al.*, 2008).

CER[N(24)S(18)] was the most abundant CER species. This particular species has been explored as a cardiovascular disease biomarker, as it shows a negative relationship with fatal outcome (Laaksonen *et al.*, 2016) and has reduced plasma levels in patients with stable coronary artery disease (Tarasov *et al.*, 2014; Laaksonen *et al.*, 2016). However, it has also been found increased in patients that have undergone a major adverse cardiovascular event (Havulinna *et al.*, 2016; Wang *et al.*, 2017). It's abundance has also shown a positive relationship with type-2 diabetes (Bergman *et al.*, 2015; Jensen *et al.*, 2019) and HOMA-IR (Lemaitre *et al.*, 2018), with an inverse relationship to insulin sensitivity (Haus *et al.*, 2009). In further studies that have measured this CER, a significant association with CVD was not mentioned

(Pan *et al.*, 2014; Cheng *et al.*, 2015; Yu *et al.*, 2015; Hilvo *et al.*, 2018) and was found not significant as a risk marker of CVD or related to risk phenotypes in others (Lopez *et al.*, 2013; Havulinna *et al.*, 2016; de Carvalho *et al.*, 2018) (Table 1.1 and Table 5.5).

Table 5.5: Associations between blood CER[N(24)S(18)] and coronary-artery disease (parts A-D) and type-2 diabetes (part E) outcomes reported in literature.

The table highlights studies of CER[N(24)S(18)] in cardiovascular disease. Association is by hazard ratio or odds ratio, where significant includes a confidence interval (CI) of greater than 1. Positive, HR > 1.00 or a positive correlation coefficient; negative/inverse, significant HR < 1.00 or a negative correlation coefficient. Change by concentration is due to an alteration of the mean concentration from controls. MACE, major adverse cardiac events; NS, not significant; NM, not mentioned even though the species has been measured (likely means the result was NS). HR, hazard ratio; CHF, chronic heart failure; AP, angina pectoris; T2D, Type-2 Diabetes. b, depending on adjustment used. Conc., by concentration; Assoc., association; Correl., correlation. "Increased in T2D" is in comparison to athletes.

A

Fatal Outcome (by Hazard or Odds Ratio)

Reference	(Laaksonen *et al.*, 2016)	(Laaksonen *et al.*, 2016)	(Havulinna *et al.*, 2016)	(Havulinna *et al.*, 2016)	(Yu *et al.*, 2015)	(de Carvalho *et al.*, 2018)
Mortalities	81	51	Fatal Incident MACE 116	Fatal Recurrent MACE 70	200	26
Controls	Stable CAD	Stable CAD	Incident MACE 7589	Recurrent MACE 326	CHF 223	AMI 288
N(24)S(18)	NS	Negative[b]	NS	NS	NM	NS

The table highlights studies of CER[N(24)S(18)] in cardiovascular disease. Association is by hazard ratio or odds ratio, where significant includes a confidence interval (CI) of greater than 1. Positive, HR > 1.00 or a positive correlation coefficient; negative/inverse, significant HR < 1.00 or a negative correlation coefficient. Change by concentration is due to an alteration of the mean concentration from controls. MACE, major adverse cardiac events; NS, not significant; NM, not mentioned even though the species has been measured (likely means the result was NS). HR, hazard ratio; CHF, chronic heart failure; AP, angina pectoris; T2D, Type-2 Diabetes. b, depending on adjustment used. Conc., by concentration; Assoc., association; Correl., correlation. "Increased in T2D" is in comparison to athletes.

B

Fatal Outcome (by mean concentration)

Reference	(Tarasov *et al.*, 2014)	(Laaksonen *et al.*, 2016)	(Laaksonen *et al.*, 2016)	(Laaksonen *et al.*, 2016)
Mortalities (n)	258	81	51	80
Controls (n)	187	Stable CAD = 1499	Stable CAD = 1586	Stable CAD = 80
N(24)S(18)	Decreased	Decreased	Decreased	Decreased

C

Major Adverse Cardiovascular Event (by concentration)

Reference	(Pan *et al.*, 2014)	(Wang *et al.*, 2017)	(Yu *et al.*, 2015)	(Havulinna *et al.*, 2016)	(Anroedh *et al.*, 2018)
Phenotypes (n)	AMI 114	MACE 230	CHF 423	MACE 813	MACE 155
Other (n)	Unstable AP 92 and stable AP 98				
Controls (n)	52		104	6892	419
N(24)S(18)	NM	Increased	NM	Increased	NS

The table highlights studies of CER[N(24)S(18)] in cardiovascular disease. Association is by hazard ratio or odds ratio, where significant includes a confidence interval (CI) of greater than 1. Positive, HR > 1.00 or a positive correlation coefficient; negative/inverse, significant HR < 1.00 or a negative correlation coefficient. Change by concentration is due to an alteration of the mean concentration from controls. MACE, major adverse cardiac events; NS, not significant; NM, not mentioned even though the species has been measured (likely means the result was NS). HR, hazard ratio; CHF, chronic heart failure; AP, angina pectoris; T2D, Type-2 Diabetes. b, depending on adjustment used. Conc., by concentration; Assoc., association; Correl., correlation. "Increased in T2D" is in comparison to athletes.

D

Acute Coronary Syndrome (by concentration)

Reference	(Cheng *et al.*, 2015)	(Anroedh *et al.*, 2018)
ACS (n)	313	313
Stable CAD (n)	261	Stable AP = 261
N(24)S(18)	Increased in ACS	Increased in ACS

E

Type-2 Diabetes

Reference	(Lopez *et al.*, 2013)	(Haus *et al.*, 2009)	(Bergman *et al.*, 2015)	(Hilvo *et al.*, 2018)	(Hilvo *et al.*, 2018)	(Jensen *et al.*, 2019)	(Lemaitre *et al.*, 2018)	(Lemaitre *et al.*, 2018)	(Lemaitre *et al.*, 2018)	(Haus *et al.*, 2009)	(Jensen *et al.*, 2019)
T2D (n)	14	14	15	1038		610	Assoc.	Assoc.	Assoc.	Correl.	Correl.
Controls (n)	14	13	15 obese and 15 athletes	7007	3344	2145	2086				
Other info	Conc.	Conc.	Conc.	HR	HR	HR	Fasting insulin	HOMA-IR	HOMA-B	Insulin sensitivity	Fasting glucose
N(24)S(18)	NS	NS	Increased in T2D	NM	NM	Positive[c]	NS[c]	Positive	NS	Inverse	NS

5.9.3. Plasma NAE species positively associated with total cholesterol

Total plasma cholesterol previously measured for this cohort was identified as a significant cofounder ($P<0.05$) in the analysis of nine NAE species, showing a positive relationship (Appendix Table 0.7). Cholesterol is adjusted for in many bioactive lipid studies (Demirkan *et al.*, 2012; Laaksonen *et al.*, 2016). The inclusion of cholesterol as a potential cofounder allows for adjustment of a dietary influence on the lipids such as altered dietary fatty acids; the substrates to many of the lipids studied here.

Adjusting for cholesterol levels has also been used to identify cardiovascular risk associated with the lipid species additional to that of traditional clinical markers (Laaksonen *et al.*, 2016). NAEs, and in particular AEA, have shown interplay with cholesterol; NAEs regulate cholesterol trafficking in worm models (Galles *et al.*, 2018) and oxidised-LDL cholesterol induces NAE macrophage production in mice (Jiang *et al.*, 2009). However, more studies are needed to fully understand this finding.

5.9.4. Genetic analyses of N-acyl ethanolamine species

The NAE lipids were all significantly heritable with PDEA estimated as the most heritable lipid of the class. The specific role and relative importance of this lipid is unknown. Four NAE species (DHEA, LEA, PEA, and VEA) were associated with rs680379, a missense change in the gene of the NAE degradation enzyme FAAH. The association with PEA was identified previously in a single candidate gene study of mutations in *FAAH* in 114 subjects (Sipe *et al.*, 2010), which reported the same direction of effect on plasma AEA, PEA, STEA and OEA species but with P-values insignificant at genome-wide levels ($0.003<P<0.04$). OEA is the only NAE species that has been previously associated with DNA variants at GWAS significance; an eQTL of *FAAH* (rs1571138, upstream to *FAAHP1*, $P=5.15\text{x}10^{-23}$) was identified in an untargeted study of blood lipids. The SNP is in complete linkage disequilibrium with the missense SNP rs324420. OEA was the only NAE species measured in that study (Long *et al.*, 2017). Here, only a suggestive association was found between rs324420 and OEA ($P=5.80\text{x}10^{-5}$), although non-significant trends were observed in the same direction for all NAE species with genotype at this SNP (Figure 5-13).

While the *FAAH* missense SNP rs324420 is not associated with any disease endpoints identified from GWAS to date, the A allele, associated with higher NAE levels, has been reported to increase the risk of polysubstance addiction and abuse [MIM: 606581] in three candidate gene studies totaling 863 cases and 2,170 controls (Sipe *et al.*, 2002; Flanagan *et al.*, 2006; Sim *et al.*, 2013). PheWAS analysis using the Gene Atlas UK Biobank online browser however did not identify significant association in a similar number of UK Biobank cases of substance abuse/dependency (OR for A allele = 1.10; P = 0.14; 746 cases and 451,518 controls). It is possible that misclassification bias has affected the UK Biobank PheWAS; among the 451,518 UK Biobank participants assigned as controls, some reported dependencies on other substances and behaviours, such as coffee, cigarettes, prescription drugs, and gambling [UKBiobank data show case; http://biobank.ndph.ox.ac.uk/showcase/, accessed April 2019]. The potential implication of NAE species in addiction through the association with the *FAAH* SNP warrants further investigation in larger numbers of cases.

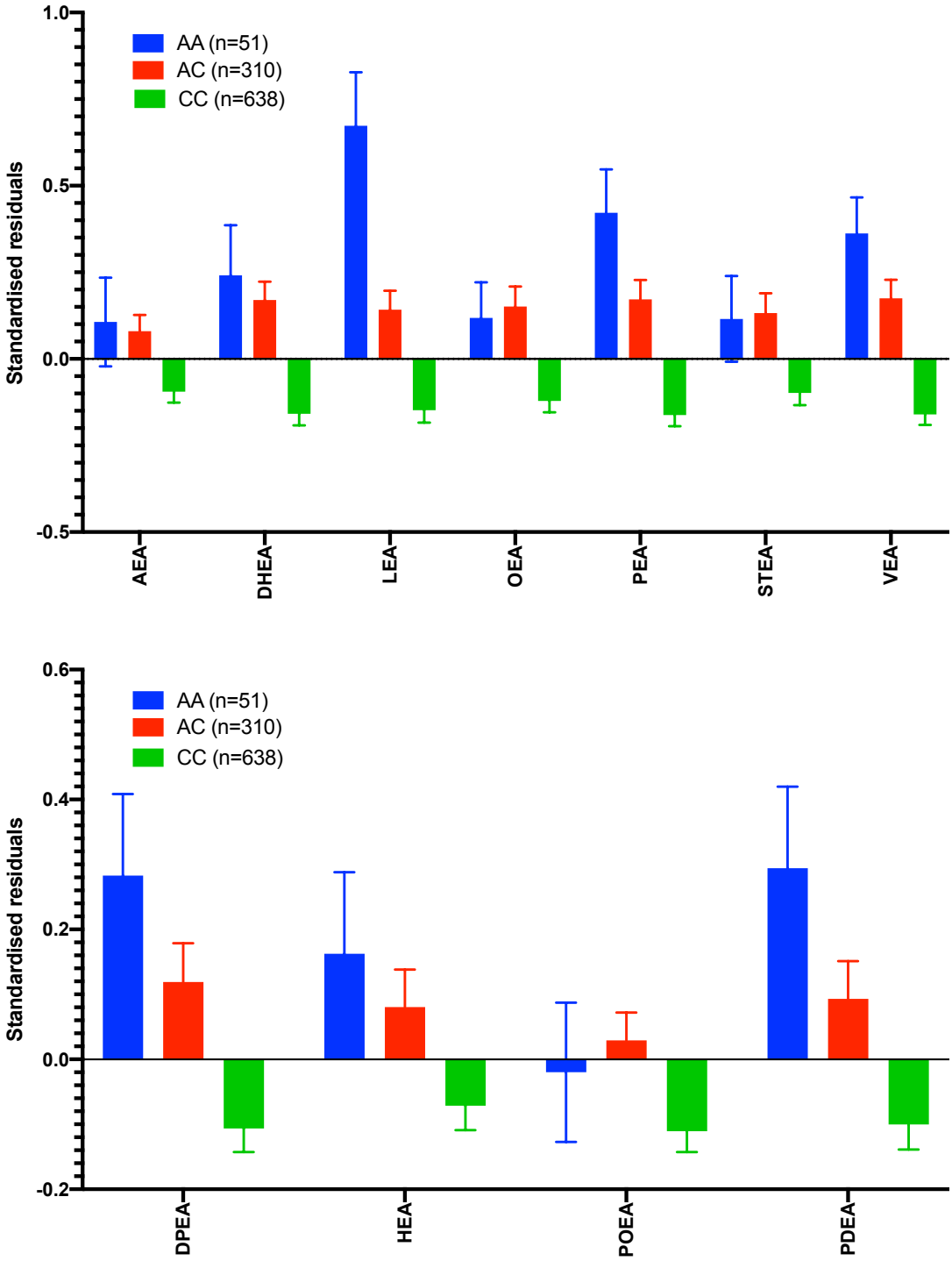


Figure 5-13: Trend in concentrations of plasma NAE species separated by *FAAH* rs324420 genotype.

Data is shown as mean ± SE of the standardised residuals of each NAE species in participants with the three genotypes (AA, AC, CC) at the *FAAH* SNP rs324420.

5.9.5. Genetic analyses of ceramides and other sphingolipid species

Here, the study assessed a larger array of species and expanded on previous estimates of heritability to show that further CER[NS] and CER[NDS] species are significantly heritable, as well as other CER species (CER[AS] and sphingoid bases C18S) that have not been assessed before.

The SNP (rs7157785) in sphingosine 1-phosphate phosphatase 1 (*SGPP1*), a CER metabolic enzyme (Figure 1-2), has been identified previously in GWAS of sphingomyelin (Hicks *et al.*, 2009; Demirkan *et al.*, 2012; Draisma *et al.*, 2015), total cholesterol (Klarin *et al.*, 2018), glycerophospholipids (Draisma *et al.*, 2015), and the ratio of an unknown blood lipid (X-08402) to cholesterol (Shin *et al.*, 2014). Other significant SNPs identified at the same locus and in linkage disequilibrium with the lead SNP, associated with the same CER species, and have been previously identified in further GWAS studies of blood phospholipids (Li *et al.*, 2018a), red cell distribution width (Astle *et al.*, 2016), sphingomyelin (Li *et al.*, 2018a), and unknown blood metabolite X-10510 (Shin *et al.*, 2014). The novel association with CER[N(24)S(16)] described is consistent with the enzyme's role in influencing CER[NS] production, through the formation of sphingosine (C18S) for CER[NS] biosynthesis (Chapter 1).

CER[N(26)S(19)] associated at GWAS significance with SNPs at a novel locus on chromosome 6, upstream to the gene encoding the inflammatory protein CD83 ($P=2.07x10^{-8}$), a member of the immunoglobulin superfamily of membrane receptors that are expressed by antigen-presenting white blood cells, leukocytes, and dendritic cells (Ju *et al.*, 2016). The protein has been suggested to be involved in T cell production and activation (Glezer *et al.*, 2015). An interaction between CD83 and CER species is currently unknown, but given the involvement of ceramide signalling in inflammation and immunity (Hannun *et al.*, 2008; Maceyka *et al.*, 2014), it would be of interest to investigate further.

Association between some CER species and a SNP (rs680379) in *SPTLC3* has been identified previously through shotgun lipidomics for a few blood CER species at GWAS significance (Hicks *et al.*, 2009; Demirkan *et al.*, 2012; Tabassum *et al.*, 2019), described in Chapter 1. Here, associations were identified for an additional

seven CER[NS] and two CER[NDS] plasma species with the same SNP and further eQTLs of *SPTLC3*. As this enzyme is the rate limiting step for the *de novo* biosynthesis of CER, this association may have wider implications. The information gathered from the eQTL analysis highlights all of the *SPTLC3* confirmed eQTLs act in the liver, which is therefore a major site for plasma CER biosynthesis. Neither PheWAS analysis in UK Biobank, nor 2SMR analysis, identified significant disease associations with the *SPTLC3* locus. A number of CER[NS] species have been studied as potential biomarkers of CVD and diabetes, and data from others has suggested that the *SPTLC3* locus is associated with these CERs (Hicks *et al.*, 2009; Demirkan *et al.*, 2012). While GWAS significant associations were not found with these lipids, the extent to which specific species have a role in CVD remains debated.

CER	Gene	Ref
CER[N(16)S(18)]	*SPTLC3*	(Hicks *et al.*, 2009)
CER[N(20)S(18)]	*CERS4*	(Hicks *et al.*, 2009; Demirkan *et al.*, 2012; Tabassum *et al.*, 2019)
CER[N(22)S(18)]	*SPTLC3*	(Hicks *et al.*, 2009; Demirkan *et al.*, 2012; Tabassum *et al.*, 2019)
CER[N(22)S(18)]	*ZNF385D*	(Tabassum *et al.*, 2019)
CER[N(22:1)S(18)]	*SPTLC3*	(Tabassum *et al.*, 2019)
CER[N(23)S(18)]	*SPTLC3*	(Hicks *et al.*, 2009; Demirkan *et al.*, 2012)
CER[N(24)S(18)]	*SPTLC3*	(Hicks *et al.*, 2009; Demirkan *et al.*, 2012)
CER[N(24)S(18)]	*ZNF385D*	(Tabassum *et al.*, 2019)
CER[N(24:1)S(18)]	*SPTLC3*	(Hicks *et al.*, 2009; Demirkan *et al.*, 2012; Tabassum *et al.*, 2019)
CER[N(22)S(19)	*SPTLC3*	Present study
CER[N(23)S(20)]	*SPTLC3*	Present study
CER[N(24)DS(19)]	*SPTLC3*	Present study
CER[N(24)DS(20)]	*SPTLC3*	Present study
CER[N(24)S(16)]	*SPTLC3*	Present study
CER[N(24)S(16)]	*SGPP1*	Present study
CER[N(24)S(19)]	*SPTLC3*	Present study
CER[N(24)S(20]	*SPTLC3*	Present study
CER[N(25)S(20)]	*SPTLC3*	Present study
CER[N(26)S(19)]	*SPTLC3*	Present study
CER[N(26)S(19)]	*CD83*	Present study

Table 5.6: List of gene associations identified from GWAS of CER species reported to date, including the present study

The current published GWAS associations for CER species to date. The unique CER species is depicted, the gene associated with the variants identified, and the reference of the publication in which the GWAS was completed are depicted.

5.9.6. Association of plasma CER species with haematological factors

A plasma CERs species, CER[N(24)S(16)], was found to have a significant causal role in the regulation of blood cell counts, as assessed by 2SMR. The analysis depicted a negative relationship with mean platelet volume and reticulocyte traits, and a positive relationship with platelet count, lymphocyte count, and red blood cell distribution width. The instrument, a SNP in *SGPP1*, increases the abundance of this shorter sphingosine base species via the creation of sphingosine from sphingosine 1-phosphate, which could also be recycled to create CER[N(24)S(16)] (Figure 1-2).

Mean blood cell volumes are higher when cells are undergoing destruction and the 2SMR results show that increases in this CER species, decreases platelet cell volumes and increases platelet cell counts. This positive relationship was also shown for lymphocyte count, potentially implicating CER with lymphocytic inflammation signalling. However, the effects of this CER seem to be blood cell-specific, as the 2SMR identified a negative relationship with multiple reticulocyte traits, which are immature red blood cells, and a positive relationship with red blood cell distribution width, a trait that has been linked to nutrient deficiency and anaemia (Salvagno *et al.*, 2015).

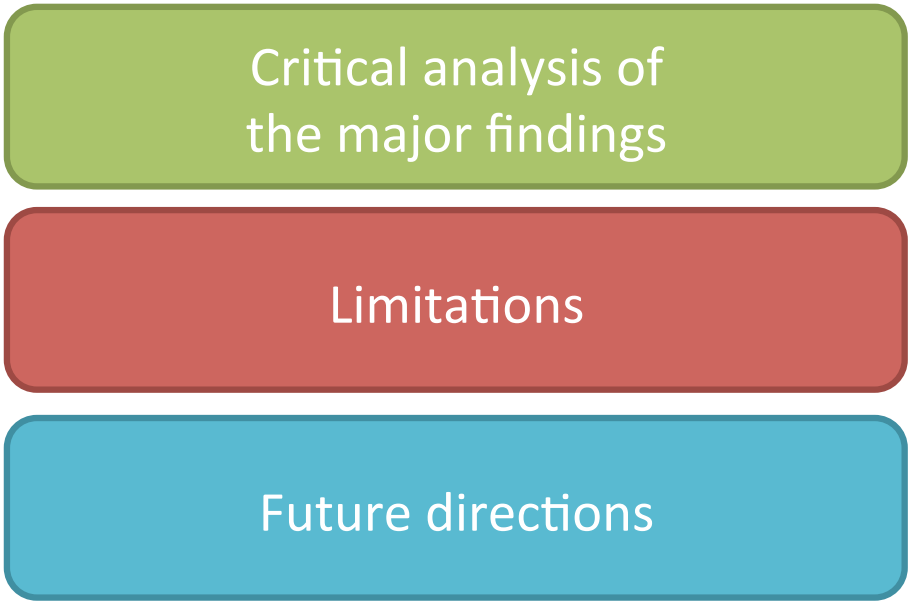
CER[N(16)S(18)] has been previously shown to stimulate erythrocyte formation through platelet activating factor (Lang *et al.*, 2005), but further studies will be required to identify the mechanism behind the association between genetically determined plasma CER levels and blood cell phenotypes. This study may provide initial evidence of more complex relationship of CERs with cell death, potentially differing in function depending on the pathway of production of the CER species, i.e. *de novo* production versus salvage/recycling of other sphingolipid mediators. Further studies examining the full profile of CER mediators are required to address this.

5.10. Conclusion

The study shows the substantial heritability estimated for an array of plasma NAE and CER species, and identifies GWAS significant associations between lipids and variants of the enzymes in their respective metabolic pathways. The results indicate that *FAAH* and *SPTLC3* are the major loci influencing plasma levels of NAEs and CERs, respectively. In addition, the study has shown novel SNP associations (*CD83, SGPP1, DEGS1)* influencing plasma CER species, which implicate CER lipids in haematological phenotypes.

Chapter 6

General discussion and future directions



6.1 Critical analysis of the major findings; *FAAH* and *SPTLC3*

The study identified significant heritability estimated for all species of the NAE and CER lipid classes. Four NAE species (DHEA, LEA, PEA, VEA) associated at GWAS with DNA variants in *FAAH*. Seven CER[NS] and two CER[NDS] associated with DNA variants of *SPTLC3* at GWAS. Further associations were identified between CER traits and *CD83*, *SGPP1*, and *DEGS1*, implicating CER species in haematological phenotypes. The association between NAE and CER lipids with CVD was not confirmed by two-sample MR, or involvement in other phenotypes that have been substantially identified in the UK Biobank study or analysed by GWAS previously and available to the GWAS Catalog database.

This study of lipidomic analysis coupled with family-based genetic analyses of heritability and GWAS, revealed *SPTLC3* and *FAAH* to be the main genetic loci influencing plasma NAE and CER lipid species in British Caucasian families. The identification of genetic variants in the enzymes of key respective biosynthetic pathways emphasises the importance of the biosynthetic enzyme *SPTLC3* and the degradation enzyme *FAAH* in maintaining respective plasma lipid concentrations.

Both associations have been identified previously; a variant in *SPTLC3* associated with four plasma CER species in three GWAS of sphingolipids (Hicks *et al.*, 2009; Demirkan *et al.*, 2012; Tabassum *et al.*, 2019), an eQTL of *FAAH* was identified for the NAE species OEA at GWAS significance (Long *et al.*, 2017), and a candidate gene study found an association with *FAAH* genotypes at rs324420 and levels of four NAE species (AEA, PEA, STEA, OEA) but not a GWAS significance. This project identified the first associations for an additional seven CER[NS] and two CER[NDS] species with *SPTLC3,* and the first association of three NAE species (LEA, VEA, DHEA) with *FAAH*.

The SNPs identified for *SPTLC3* and *FAAH* did not associate with any traits measured in the UK Biobank PheWAS analyses, nor in previous disease GWAS studies recorded on the GWAS Catalog. Therefore this study did not confirm a role for the NAE or CER pathways in CVD. Recently, a GWAS of the lipidome associated a SNP in *SPTLC3* (rs364585) with intracerebral haemorrhage (IH) risk using UK Biobank study data, where the G allele associated with decreased CERs, and

decreased IH risk (Tabassum *et al.*, 2019). The study used SAIGE for the analysis and did not provide details on sample numbers. The online UK Biobank PheWAS analyses (Gene Atlas browser) also completed this same assessment for 773 patients with IH against 451,491 controls and didn't identify a significant association with *SPTLC3* variants, nor did the Michigan PheWAS browser using SAIGE for assessment of the UK Biobank data [http://pheweb.sph.umich.edu/SAIGE-UKB/, accessed Nov 2019]. Depicted in Chapter 5, the SNP has been associated with LDL cholesterol in GWAS catalog previously, but no other CVD trait. Knockout animal studies of *SPTLC3* are currently lacking preventing an assessment of whether *SPTLC3* has a role in CVD progression. Myriocin, an inhibitor of the SPTLC enzyme, was administered to apolipoprotein E knockout mice which resulted in reductions in CER[N(18)S(18)], as well as cholesterol and triglycerides, and the mice showed a regression of pre-existing atherosclerotic lesions to the formation of a stable plaque phenotype (Park *et al.*, 2008). Further assessment of the extended CER class in a large phenotyped and genotyped cohort will be required to confirm or deny the hypothesis that CER species have a causal role in CVD.

As direct CB1 antagonist drugs have caused severe adverse psychiatric effects (Mach *et al.*, 2009), FAAH inhibitors are being evaluated as an alternative approach to modulating eCB signalling. In 2016, a FAAH inhibitor resulted in severe neurological side-effects in a Phase I trial; it was hypothesised to be due to off-target drug effects or toxicity specific to that particular drug, as other clinical trials of multiple other FAAH inhibitors have reported no serious side effects (Mallet *et al.*, 2016). However, chronic use of the drug candidates have not been assessed yet. As the functional *FAAH* SNP rs324420, which substantially impacts FAAH activity, did not associate with any adverse phenotypes in the UK Biobank, our results suggest that on-target effects of FAAH inhibitor drugs likely do not have substantial risks of causing conditions that occurred with appreciable frequency in UK Biobank. Knockout mice studies of *FAAH* have shown that animals lacking *FAAH* have an increased fear/startle reaction [http://animalab.eu/, strain TGRA6300] and cardiometabolic issues, potentially due to the increased activation of peroxisome proliferating activated receptor gamma (PPARγ), a receptor linked to NAE signalling (Brown *et al.*, 2012). This implication of NAE species with roles in cardiometabolic phenotypes may be driven through their relationship with total cholesterol levels described here.

6.2 Critical analysis of other findings; *DEGS1, SGPP1, CD83*

SNP associations influencing plasma CER species (*CD83, SGPP1, DEGS1)* were identified that have not been previously described. The phosphatase *SGPP1* and desaturase *DEGS1* associations are findings that are consistent with their enzymatic involvement in CER biochemistry. These two associations implicate CER lipids in haematological phenotypes, of which, further assessment is required to fully understand the implication, whether it is through known roles of CER in apoptosis or the expression of CER by blood cells that causes the interplay between plasma CER levels and blood cell traits.

The identification of an association between CER traits and the genomic locus of *CD83* warrants further evaluation. Expression of *CD83* is systemic, but highest in bone marrow (Uhlen *et al.*, 2015). The protein is found in transmembrane and soluble forms. The surface marker is involved in the differentiation of immune and dendritic cells (Prazma *et al.*, 2007). The soluble protein has been shown to negatively regulate the immune response, where uses for the protein include transplant/graft rejection, autoimmune diseases, and T cell proliferation (Horvatinovich *et al.*, 2017). Interestingly, the protein is highly expressed in Hodgkin lymphoma patients' cell lines and plasma, and it is excreted from tumour cells, inhibiting T cell proliferation via programmed cell death (Li *et al.*, 2018b), which may provide disease areas for further study of CER involvement. Safety of anti-CD83 antibodies has been shown in non-human primates (Li *et al.*, 2018b), where the association identified here of CER species with blood cell counts may also be of importance.

6.3 Lipid species of the Eico, NAE, and CER class are influenced by genetic factors over a range

This study assessed estimates of additive genetic variance for both heritability and GWAS. The heritability results described in Chapters 4 and 5, for the range finding and full cohort studies, respectively, showed that the estimated heritability varied for the tested lipid species. There are three assumptions when estimating heritability; (i) there is no genotype by environment covariance (e.g. parents with high IQ provide children with IQ-stimulating environment), (ii) there is no genotype by environment interaction (e.g. stressful life events cause depression in people with a polymorphism

in the serotonin transporter gene (Caspi *et al.*, 2003)), and (iii) the resemblance in relations that is due to common environmental effects is equal (Visscher *et al.*, 2008). If the assumptions are incorrect, the heritability estimate will be inflated. In this project, the participants were all Caucasians from Oxford, UK, which likely allowed for similarity in environmental factors for all immediate family members, including Western diet (Simopoulos, 2006). However, if this was not the case, diet could inflate the heritability estimates.

Heritability estimated as low, and therefore non-genetic/environmental estimates are high, can be due to measurement error (Visscher *et al.*, 2008). This can only be overcome by increased sample sizes and as the number of participants in large-scale biobank studies increases, assessment of their metabolome in conjunction with their genome would provide further information on the genetic influence of circulating metabolites. The estimates of heritability between the pedigree-based and SNP/genetic relationship matrix-based methods were very similar (R = 0.99), so it is unlikely that the chosen heritability software or structure of the cohort influenced the heritability estimates.

6.4 Limitations of the project

6.4.1 Non-fasting plasma samples

In this study, non-fasting plasma samples were analysed. While diet has been shown to not have an effect on circulating ceramide concentrations (Lankinen et al., 2016; Wang et al., 2017), certain mediators that are direct derivatives of fatty acids and belong to the NAE and Eico classes, can be altered with diet (Gouveia-Figueira et al., 2015). Recent discussions in the cardiovascular field surrounds the idea that most individuals spend the majority of their life in a postprandial state, so physiologically it may be of higher importance to assess risk of disease in that same state. It has been shown for lipoproteins and cholesterol that random non-fasting lipid profiles are not different to fasting profiles and that both are comparable in the prediction of CVD (Nordestgaard et al., 2016). Similar studies have not been completed for the lipids studied here in substantial sample sizes, and therefore it is a limitation of this project that would need to be further assessed in a repetition cohort. In this study, diet-influenced traits total blood cholesterol levels and BMI were added as a potential

predictors of the blood lipid levels. This would have potentially accounted for a small amount of dietary influence over the lipid levels, in particular the NAE species, which required adjustment for cholesterol levels. However, as the samples were non-fasting and participant's diet was not recorded, diet was not controlled for fully in this study.

6.4.2 EDTA: preferred anticoagulant for lipidomic analysis

The action of drawing blood for a sample can activate white blood cells, platelets, and cause coagulation, increasing inflammatory lipids and haemolysis (releasing lipid mediators from cells) (Quehenberger et al., 2011). The three most common anticoagulants used for plasma storage are citrate, EDTA and heparin. Citrate and EDTA chelate calcium ions involved in coagulation. Heparin activates enzyme inhibitor anti-thrombin III enzymes. It has been shown that heparin enhances plasma phospholipase A2, the enzyme that releases PUFAs from membrane phospholipids leading to increased production of the Eico lipid species studied in this project (Nakamura et al., 1995). Lithium heparin was shown to interfere with mass spectrometry analyses as it may cause matrix effects, ion suppression, enhanced mass signal, and ion adducts (Mei et al., 2003). Heparin anticoagulants have been shown to present more variable results when compared to other anticoagulants for sphingolipids, such as the CER species studied here (Hammad et al., 2010). However, analyses of all three anticoagulants have shown that mass spectrometry peaks obtained for lipid mediators were different depending on the anticoagulant used, with citrate shown to present significantly different results to heparin and EDTA in two studies (Gonzalez-Covarrubias et al., 2013; Surma et al., 2015). Thus, plasma stored in EDTA anticoagulant is less variable with less negative effects on mass spectrometry analysis as compared to citrate and heparin in lipidomics analyses.

6.4.3 Biobank storage of plasma for the analysis of bioactive lipid species

The plasma was stored at -80°C over a twenty year period of time and it is unknown if this affected the lipid species measured here. Literature suggests that many factors affect the storage of lipids: storage temperature, freeze/thaw cycles, humidity, light exposure, type of vials and caps used for extraction, and sample handling (Quehenberger et al., 2010; Hinterwirth et al., 2014; Gislefoss et al., 2015), but there is not a large range of studies looking at lipid mediator storage in plasma over a long

period of time. Although different to the species studied in this project, studies of lipoprotein levels provide insights into lipid storage effects. Modest decreases in total cholesterol and HDL cholesterol has been shown per year of storage for 26 years at -20°C in serum, and the authors suggest that it is unlikely to significantly affect cardiovascular risk stratification but could underestimate risk (Arts et al., 2014). They also found that lipid levels in their oldest cohort were lower than those quantified in a newer cohort, but their storage was at - 20°C for the first 20 years, only moving to -80°C storage in the recent years before analysis. Gislefoss et al. stored serum samples at -25°C for 29 years and found increases in HDL and LDL. They decreased the temperature to -40°C during years 12-14 which could have affected the analyses (Gislefoss et al., 2015). Research into this area would aid biobank sampling for use in lipidomic studies. Comparison with repeatedly measured (24 times) pooled quality control samples collected in 2008, showed that the cohort plasma samples did not differ greater than a %CV of 30. Three CER species, four NAE species, and three Eico species showed a substantial difference to the pooled plasma levels (Section 3.2.4). This could be potentially due to the large numbers of samples analysed for the cohort plasma (n=999 CER, n=999 NAE, n=204 Eico) compared to that of the QC standard, or due to the pooling process itself, in the creation of the pooled Quality Control samples. While storage effects have not been fully accounted for in this study and sample degradation may have occurred, it has not substantially influenced the levels of all of the lipid species in this study.

6.4.4 Sample size

As described, this cohort of extended families has previously been shown to have adequate power to detect moderate-sized genetic influences on quantitative traits (Vickers *et al.*, 2002; Baker *et al.*, 2007), and such moderate-sized genetic influences were identified here for a number of bioactive lipid mediators. The sample size analysed here (999 participants) is the largest study analysing this number of plasma NAE and CER species to date (Hinterwirth *et al.*, 2014). The lack of significant associations for all species at GWAS, however, is likely due to the sample size, and future efforts in large cohorts will be required to undertake lipidomic analyses of low concentration mediators using high throughput set ups, for example; extraction

completed via robotics, high throughput analytical platforms, and quantification through high throughput bioinformatic techniques.

6.5 Future directions

6.5.1 Large cohort analysis of Eico species

While not assessed in the full cohort, four Eico species were significantly heritable in 196 plasma samples (11,12-DHET, 14,15-DHET, 4-HDHA, TransEKODE). The assessment of this class of species in a larger sample size would likely identify a more significantly heritable lipid species, as was shown by the increased number of lipids that were significantly heritable in the full cohort analysis of NAE and CER lipids, compared to the range finding study. As GWAS associations for the four heritable lipids are currently unknown, identification of the DNA variants influencing their concentrations in plasma, potentially in the genes of the proteins of their respective biosynthetic pathways (such as CYP450 enzymes), would be of use to assess their potential involvement in disease.

6.5.2 Rare variant studies of *SPTLC3* and *FAAH* genes

The UK Biobank study has released exome sequencing data of 50,000 participants. Assessment of rare variants of the major effect loci (*FAAH* and *SPTLC3*) may identify rare, large-effect variants in a subset of the UK biobank population, such variants may influence NAE and CER biosynthesis. This has been shown in a previous study in which a Hispanic population-specific rare variant in *DEGS1* caused an increase in plasma CER[NDS] species (Blackburn *et al.*, 2019).

6.5.3 Experimental assessment of the relationship between plasma CER species and haematological traits

The causal association identified here between variants in *SGPP1,* plasma CER[N(24)S(16)], and blood cell counts could be confirmed in vitro. For example, megakaryocytes, the precursors of platelets, could be cultured and a synthetic version of CER[N(24)S(16)] could be added at physiological levels. The cultured cells could be counted to assess whether the application of CER species alters the number of platelets produced. Furthermore, follow up of the cohort participants would allow for

a study of routine multiple blood count testing, and thus an assessment could confirm if a genetic and environmental effect exists on CER and blood cell counts in the participants with known variants in *SGPP1*.

If both of these experiments confirmed the association between CER and blood cell counts, a further future analysis could assess the influence of drugs targeting the CER pathway, such as Myriocin, on participants with the known genotypes. While the variants influencing CER species have not been associated with cancers in literature, CER species involved in normal variation might mediate the risk of blood cancers and diseases involved in the over or under production of blood cells. CER species may be a potential future pharmaceutical target for blood cell disorders, if found to alter blood cell levels in a controllable manner.

6.5.4 Further analysis of the implication of variants of *FAAH* in addiction

An international collaboration with the authors of the current GWAS studies of drug addiction would allow for a meta-analysis to be completed in a substantial number of cases to identify common DNA variants influencing risk of addiction. This may confirm an association with variants in *FAAH* and furthermore the implication of NAE species in drug addiction.

6.6 Conclusion

This study provides the first heritability estimates for species of the Eico, NAE, and CER lipid classes, identifying species in each class that are particularly influenced by genetic factors. Provided is also the first GWAS significant evidence of association between SNPs in the *FAAH* gene and four plasma NAEs. Additionally, the study has extended the previously described association between SNPs in the *SPTLC3* gene and plasma CERs to a wider range of CER species. In addition, novel SNP associations (*CD83, SGPP1, DEGS1)* influencing plasma CER species, which implicate CER lipids in haematological phenotypes, were found, potentially implicating CER species as targets for blood cell disorders, however further analyses are required to understand the biological mechanism behind the association (Figure 6-1). The implication of the lipid species of the Eico, NAE, and CER classes with CVD risk remains unconfirmed.

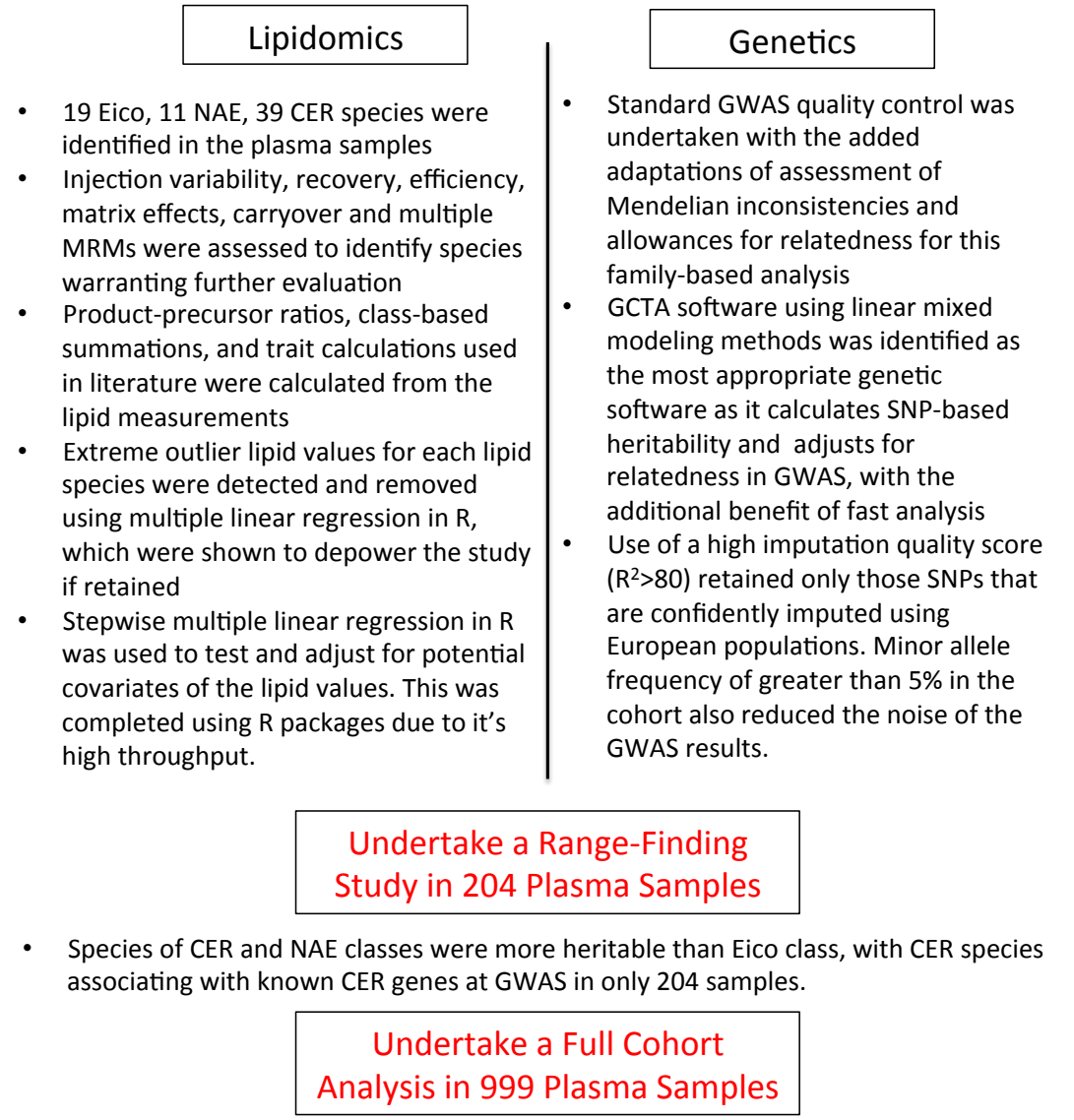


Figure 6-1: Diagrammatic overview of thesis results

At the commencement of this project, raw genotyping data and unopened plasma samples were available for the cohort. To understand the genetic influence over three classes of bioactive plasma lipids, a range-finding study was undertaken to identify those lipids at particular genetic influence in a subset of the cohort (204 plasma samples). Identification of the NAE and CER species were more substantially heritable at this stage than the Eico group. This allowed for a focused study of NAE and CER using the full cohort samples (999 plasma samples).

Lipids were identified from the three classes that were quantifiable in the plasma samples; the quality of the lipidomics data was assessed, ratios and summations for the lipid classes analysed were calculated, and extreme outlier lipid values were removed. Batch effects were identified as the most influential covariate.

With the genotyping data, standard quality control was undertaken with the adaptation of two steps for a family-based cohort. GCTA was identified as the most appropriate family-based GWAS software. Quality control thresholds were assessed for the imputed data analyses so as to only retain the most confidently imputed SNPs.

The full cohort results addressed the three main questions; 1) lipids are significantly influenced by genetics over a range (heritability estimates); 2) common variants were found at GWAS in loci of the lipid biosynthetic enzymes; 3) the lipid were not confirmed to have a causal role in cardiovascular disease, but were found to potentially influence drug addiction and blood cell traits (2-sample Mendelian Randomisation).

References

Abecasis, G.R., Cardon, L.R. and Cookson, W.O., *A general test of association for quantitative traits in nuclear families.* American Journal of Human Genetics, 2000. **66**(1): p. 279–292.

Acuna-Hidalgo, R., Veltman, J.A. and Hoischen, A., *New insights into the generation and role of de novo mutations in health and disease*. Genome Biology, 2016. **17**(241): p. 1–19.

Alvarez-Madrazo, S., Padmanabhan, S., Mayosi, B.M., et al., *Familial and phenotypic associations of the aldosterone renin ratio*. Journal of Clinical Endocrinology and Metabolism, 2009. **94**(11): p. 4324–4333.

Anroedh, S., Hilvo, M., Akkerhuis, K.M., et al., *Plasma concentrations of molecular lipid species predict long-term clinical outcome in coronary artery disease patients*. Journal of Lipid Research, 2018. **59**: p. 1729–1737.

Armbruster, D.A. and Pry, T., *Limit of blank, limit of detection and limit of quantitation.* The Clinical biochemist. Reviews, 2008. **29**: p. S49–S52.

Arnott, D., *Basics of Triple-Stage Quadrupole/Ion-Trap Mass Spectrometry: Precursor, Product and Neutral-Loss Scanning. Electrospray Ionisation and Nanospray Ionisation*. In *Proteome Research: Mass Spectrometry* pp. 11–13.

Arts, E.E.A., Popa, C.D., Smith, J.P., et al., *Serum Samples That Have Been Stored Long-Term Can Be Used as a Suitable Data Source for Developing Cardiovascular Risk Prediction Models in Large Observational Rheumatoid Arthritis Cohorts*. BioMed Research International, 2014. **2014**: p. 1–8.

Astarita, G., Kendall, A.C., Dennis, E.A., et al., *Targeted lipidomic strategies for oxygenated metabolites of polyunsaturated fatty acids*. Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids, 2015. **1851**(4): p. 456–468.

Astle, W.J., Elding, H., Jiang, T., et al., *The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease*. Cell, 2016. **167**(5): p. 1415–1429.

Auton, A., Abecasis, G.R., Altshuler, D.M., et al., *A global reference for human genetic variation*. Nature, 2015. **526**(7571): p. 68–74.

Baker, M., Gaukrodger, N., Mayosi, B.M., et al., *Association between common polymorphisms of the proopiomelanocortin gene and body fat distribution: A family study*. Diabetes, 2005. **54**(8): p. 2492–2496.

Baker, M., Rahman, T., Hall, D., et al., *The C-532T polymorphism of the angiotensinogen gene is associated with pulse pressure: A possible explanation for heterogeneity in genetic association studies of AGT and hypertension*. International Journal of Epidemiology, 2007. **36**(6): p. 1356–1362.

Bellis, C., Kulkarni, H., Mamtani, M., et al., *Human plasma lipidome is*

pleiotropically associated with cardiovascular risk factors and death. Circulation: Cardiovascular Genetics, 2014. **7**(6): p. 854–863.

Belmont, J.W., Hardenbol, P., Willis, T.D., et al., *The international HapMap project*. Nature, 2003. **426**: p. 789–796.

Benjamin, E.J., Muntner, P., Alonso, A., et al., *Heart Disease and Stroke Statistics-2019 Update: A Report From the American Heart Association*. Circulation, 2019. **139**: p. e56–e528.

Berg, K., *Lp(a) lipoprotein: An overview*. Chemistry and Physics of Lipids, 1994. **67–68**: p. 9–16.

Bergman, B.C., Brozinick, J.T., Strauss, A., et al., *Serum sphingolipids: relationships to insulin sensitivity and changes with exercise in humans.* American journal of physiology. Endocrinology and metabolism, 2015. **309**(4): p. 398–408.

Bielawski, J., Szulc, Z.M., Hannun, Y.A., et al., *Simultaneous quantitative analysis of bioactive sphingolipids by high-performance liquid chromatography-tandem mass spectrometry*. Methods, 2006. **39**(2): p. 82–91.

Blackburn, N.B., Michael, L.F., Meikle, P.J., et al., *Rare DEGS1 variant significantly alters de novo ceramide synthesis pathway.* Journal of lipid research, 2019. **60**(9): p. 1630–1639.

Boath, A., Graf, C., Lidome, E., et al., *Regulation and traffic of ceramide 1-phosphate produced by ceramide kinase: Comparative analysis to glucosylceramide and sphingomyelin*. Journal of Biological Chemistry, 2008. **283**(13): p. 8517–8526.

Bots, S.H., Peters, S.A.E. and Woodward, M., *Sex differences in coronary heart disease and stroke mortality: A global assessment of the effect of ageing between 1980 and 2010*. BMJ Global Health, 2017. **2**: p. e000298.

Bowen, K.J., Harris, W.S. and Kris-Etherton, P.M., *Omega-3 Fatty Acids and Cardiovascular Disease: Are There Benefits?* Current Treatment Options in Cardiovascular Medicine, 2016. **18**(69): p. 1–16.

Bozzoli, C., Deaton, A. and Quintana-Domeque, C., *Adult height and childhood disease*. Demography, 2009. **46**(4): p. 647–669.

Bray, P.F., Mathias, R.A., Faraday, N., et al., *Heritability of platelet function in families with premature coronary artery disease.* Journal of thrombosis and haemostasis : JTH, 2007. **5**(8): p. 1617–1623.

Brown, W.H., Gillum, M.P., Lee, H.Y., et al., *Fatty acid amide hydrolase ablation promotes ectopic lipid storage and insulin resistance due to centrally mediated hypothyroidism*. Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(37): p. 14966–14971.

Burgess, S. and Harshfield, E., *Mendelian randomization to assess causal effects of blood lipids on coronary heart disease: lessons from the past and applications to*

the future. Current opinion in endocrinology, diabetes, and obesity, 2016. **23**(2): p. 124–130.

Burton, P.R., Clayton, D.G., Cardon, L.R., et al., *Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls*. Nature, 2007. **447**(7145): p. 661–678.

Calignano, A., La Rana, G. and Piomelli, D., *Antinociceptive activity of the endogenous fatty acid amide, palmitylethanolamide*. European Journal of Pharmacology, 2001. **419**: p. 191–198.

Camacho, M., Martinez-Perez, A., Buil, A., et al., *Genetic determinants of 5-lipoxygenase pathway in a Spanish population and their relationship with cardiovascular risk*. Atherosclerosis, 2012. **224**(1): p. 129–135.

de Carvalho, L.P., Tan, S.H., Ow, G.S., et al., *Plasma Ceramides as Prognostic Biomarkers and Their Arterial and Myocardial Tissue Correlates in Acute Myocardial Infarction*. JACC: Basic to Translational Science, 2018. **3**(2): p. 163–175.

Caspi, A., Sugden, K., Moffitt, T.E., et al., *Influence of life stress on depression: Moderation by a polymorphism in the 5-HTT gene*. Science, 2003. **301**(5631): p. 386–389.

Chang, C.C., Chow, C.C., Tellier, L.C.A.M., et al., *Second-generation PLINK : rising to the challenge of larger and richer datasets*, 2015.: p. 1–16.

Checa, A., Holm, T., Sjödin, M.O.D., et al., *Lipid mediator profile in vernix caseosa reflects skin barrier development*. Scientific Reports, 2015. **5**: p. 15740.

Cheng, J.M., Suoniemi, M., Kardys, I., et al., *Plasma concentrations of molecular lipid species in relation to coronary plaque characteristics and cardiovascular outcome: Results of the ATHEROREMO-IVUS study*. Atherosclerosis, 2015. **243**(2): p. 560–566.

Chiang, K.P., Gerber, A.L., Sipe, J.C., et al., *Reduced cellular expression and activity of the P129T mutant of human fatty acid amide hydrolase: Evidence for a link between defects in the endocannabinoid system and problem drug use*. Human Molecular Genetics, 2004. **13**(18): p. 2113–2119.

Clarke, R., Peden, J.F., Hopewell, J.C., et al., *Genetic variants associated with Lp(a) lipoprotein level and coronary disease.* The New England journal of medicine, 2009. **361**(26): p. 2518–2528.

Coleman, J.R.I., Euesden, J., Patel, H., et al., *Quality control, imputation and analysis of genome-wide genotyping data from the Illumina HumanCoreExome microarray*. Briefings in Functional Genomics, 2016. **15**(4): p. 298–304.

Cordell, H.J., Bentham, J., Topf, A., et al., *Genome-wide association study of multiple congenital heart disease phenotypes identifies a susceptibility locus for atrial septal defect at chromosome 4p16*. Nature Genetics, 2013. **45**: p. 822–824.

Cunnington, M.S., Kay, C., Avery, P.J., et al., *STK39 polymorphisms and blood pressure: an association study in British Caucasians and assessment of cis-acting influences on gene expression.* BMC medical genetics, 2009. **10**(135): p. 1–9.

Cuvillier, O., Pirianov, G., Kleuser, B., et al., *Suppression of Ceramide-Mediated Programmed Cell Death by Sphingosine-1-Phosphate*. Nature, 1996. **381**: p. 800–803.

Darmani, N.A., Izzo, A.A., Degenhardt, B., et al., *Involvement of the cannabimimetic compound, N-palmitoyl-ethanolamine, in inflammatory and neuropathic conditions: Review of the available pre-clinical data, and first human studies*. Neuropharmacology, 2005. **48**: p. 1154–1163.

Das, S., Forer, L., Schönherr, S., et al., *Next-generation genotype imputation service and methods*. Nature Genetics, 2016. **48**(10): p. 1284–1287.

Davies, N.M., Holmes, M. V. and Davey Smith, G., *Reading Mendelian randomisation studies: A guide, glossary, and checklist for clinicians*. BMJ, 2018. **362**: p. k601.

Demirkan, A., van Duijn, C.M., Ugocsai, P., et al., *Genome-wide association study identifies novel loci associated with circulating phospho- and sphingolipid concentrations*. PLoS Genetics, 2012. **8**(2): p. e1002490.

Devane, W.A., Hanuš, L., Breuer, A., et al., *Isolation and structure of a brain constituent that binds to the cannabinoid receptor*. Science, 1992. **258**(5090): p. 1946–1949.

Devlin, B. and Roeder, K., *Genomic control for association studies*. Biometrics, 1999. **55**(4): p. 997–1004.

Ding, Y., Wu, C.C., Garcia, V., et al., *20-HETE induces remodeling of renal resistance arteries independent of blood pressure elevation in hypertension*. American Journal of Physiology - Renal Physiology, 2013. **305**(5): p. F753–F763.

Döring, A., Gieger, C., Mehta, D., et al., *SLC2A9 influences uric acid concentrations with pronounced sex-specific effects*. Nature Genetics, 2008. **40**(4): p. 430–436.

Draisma, H.H.M., Pool, R., Kobl, M., et al., *Genome-wide association study identifies novel genetic variants contributing to variation in blood metabolite levels*. Nature Communications, 2015. **6**: p. 7208.

Dyerberg, J., *Coronary heart disease in Greenland Inuit: a paradox. Implications for western diet patterns.* Arctic medical research, 1989. **48**(2): p. 47–54.

Engeli, S., Böhnke, J., Feldpausch, M., et al., *Activation of the peripheral endocannabinoid system in human obesity*. Diabetes, 2005. **54**(10): p. 2838–2843.

Evangelou, E., Warren, H.R., Mosen-Ansorena, D., et al., *Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits*. Nature Genetics, 2018. **50**: p. 1412–1425.

Fahy, E., Subramaniam, S., Brown, H.A., et al., *A comprehensive classification system for lipids*. Journal of Lipid Research, 2005. **46**(5): p. 839–862.

Fahy, E., Subramaniam, S., Murphy, R.C., et al., *Update of the LIPID MAPS comprehensive classification system for lipids*. J Lipid Res, 2009. **50**: p. S9-14.

Fanelli F. et al. *Profiling plasma N-Acylethanolamine levels and their ratios as a biomarker of obesity and dysmetabolism.* Mol Metab, 2018. **14**: 82-94.

FDA, *Guidance for Industry Bioanalytical Method Validation*. Biopharmaceutics, 2001. **May**: p. 1–22.

Flanagan, J.M., Gerber, A.L., Cadet, J.L., et al., *The fatty acid amide hydrolase 385 A/A (P129T) variant: Haplotype analysis of an ancient missense mutation and validation of risk for drug addiction*. Human Genetics, 2006. **120**(4): p. 581–588.

Frohlich, J., Dobiášová, M., Adler, L., et al., *Gender differences in plasma levels of lipoprotein (a) in patients with angiographically proven coronary artery disease*. Physiological Research, 2004. **53**(5): p. 481–486.

Gainer, J. V., Bellamine, A., Dawson, E.P., et al., *Functional variant of CYP4A11 20-hydroxyeicosatetraenoic acid synthase is associated with essential hypertension*. Circulation, 2005. **111**(1): p. 63–69.

Galles, C., Prez, G.M., Penkov, S., et al., *Endocannabinoids in Caenorhabditis elegans are essential for the mobilization of cholesterol from internal reserves*. Scientific Reports, 2018. **8**: p. 6398.

Gardner, K., Fulford, T., Silver, N., et al., g(HbF) *: a genetic model of fetal hemoglobin in sickle cell disease*. Blood Advances, 2018. **2**(3): p. 235–239.

Gaukrodger, N., Avery, P.J. and Keavney, B., *Plasma potassium level is associated with common genetic variation in the β-subunit of the epithelial sodium channel*. American Journal of Physiology - Regulatory Integrative and Comparative Physiology, 2008. **294**(3): p. R1068–R1072.

Gaukrodger, N., Mayosi, B.M., Imrie, H., et al., *A rare variant of the leptin gene has large effects on blood pressure and carotid intima-medial thickness: A study of 1428 individuals in 248 families*. Journal of Medical Genetics, 2005. **42**(6): p. 474–478.

Gislefoss, R.E., Grimsrud, T.K. and Mørkrid, L., *Stability of selected serum hormones and lipids after long-term storage in the Janus Serum Bank*. Clinical Biochemistry, 2015. **48**(6): p. 364–369.

Glezer, I., Rivest, S. and Xavier, A.M., *CD36, CD44, and CD83 Expression and Putative Functions in Neural Tissues*. In *Neural Surface Antigens: From Basic*

Biology Towards Biomedical Applications pp. 27–40.

Godlewski, G., Offertáler, L., Wagner, J.A., et al., *Receptors for acylethanolamides-GPR55 and GPR119*. Prostaglandins and Other Lipid Mediators, 2009. **89**(3–4): p. 105–111.

Goldinger, A., Henders, A.K., McRae, A.F., et al., *Genetic and nongenetic variation revealed for the principal components of human gene expression*. Genetics, 2013. **195**(3): p. 1117–1128.

Gonzalez-Covarrubias, V., Dane, A., Hankemeier, T., et al., *The influence of citrate, EDTA, and heparin anticoagulants to human plasma LC-MS lipidomic profiling*. Metabolomics, 2013. **9**(2): p. 337–348.

Gouveia-Figueira, S., Späth, J., Zivkovic, A.M., et al., *Profiling the oxylipin and endocannabinoid metabolome by UPLC-ESI-MS/MS in human plasma to monitor postprandial inflammation*. PLoS ONE, 2015. **10**(7): p. 1–29.

Hall, D., Mayosi, B.M., Rahman, T.J., et al., *Common variation in the CD36 (fatty acid translocase) gene is associated with left-ventricular mass.* Journal of hypertension, 2011. **29**(4): p. 690–695.

Hall, D.H., Rahman, T., Avery, P.J., et al., *INSIG-2 promoter polymorphism and obesity related phenotypes: association study in 1428 members of 248 families.* BMC medical genetics, 2006. **7**(83): p. 1–6.

Hamberg, M., Svensson, J. and Samuelsson, B., *Thromboxanes: a new group of biologically active compounds derived from prostaglandin endoperoxides*. Proceedings of the National Academy of Sciences of the United States of America, 1975. **72**(8): p. 2994–2998.

Hammad, S.M., Pierce, J.S., Soodavar, F., et al., *Blood sphingolipidomics in healthy humans: impact of sample collection methodology.* Journal of lipid research, 2010. **51**(10): p. 3074–3087.

Hampson, A.J. and Grimaldi, M., *12-Hydroxyeicosatetrenoate (12-HETE) Attenuates AMPA Receptor-Mediated Neurotoxicity: Evidence for a G-Protein-Coupled HETE Receptor*. The Journal of Neuroscience, 2002. **22**(1): p. 257–264.

Hannun, Y.A. and Obeid, L.M., *Principles of bioactive lipid signalling: lessons from sphingolipids.* Nature reviews. Molecular cell biology, 2008. **9**(2): p. 139–150.

Harper, A.R., Mayosi, B.M., Rodriguez, A., et al., *Common Variation Neighbouring Micro-RNA 22 Is Associated with Increased Left Ventricular Mass*. PLoS ONE, 2013. **8**(1): p. e55061.

Haus, J.M., Kashyap, S.R., Kasumov, T., et al., *Plasma ceramides are elevated in obese subjects with type 2 diabetes and correlate with the severity of insulin resistance*. Diabetes, 2009. **58**(2): p. 337–343.

Havulinna, A.S., Sysi-Aho, M., Hilvo, M., et al., *Circulating Ceramides Predict*

Cardiovascular Outcomes in the Population-Based FINRISK 2002 Cohort. Arteriosclerosis, Thrombosis, and Vascular Biology, 2016. **36**(12): p. 2424–2430.

Hicks, A.A., Pramstaller, P.P., Johansson, A., et al., *Genetic determinants of circulating sphingolipid concentrations in European populations*. PLoS Genetics, 2009. **5**(10): p. e1000672.

Hilvo, M., Salonurmi, T., Havulinna, A.S., et al., *Ceramide stearic to palmitic acid ratio predicts incident diabetes*. Diabetologia, 2018. **61**(6): p. 1424–1434.

Hinrichs, A.S., *The UCSC Genome Browser Database: update 2006*. Nucleic Acids Res, 2006. **1**(34): p. D590–D598.

Hinterwirth, H., Stegemann, C. and Mayr, M., *Lipidomics: Quest for molecular lipid biomarkers in cardiovascular disease*. Circulation: Cardiovascular Genetics, 2014. **7**(6): p. 941–954.

Ho, W.S.V., Barrett, D.A. and Randall, M.D., *'Entourage' effects of N-palmitoylethanolamide and N-oleoylethanolamide on vasorelaxation to anandamide occur through TRPV1 receptors*. British Journal of Pharmacology, 2008. **155**: p. 837–846.

Hoefer, I.E., Steffens, S., Ala-Korpela, M., et al., *Novel methodologies for biomarker discovery in atherosclerosis*. European Heart Journal, 2015. **36**(39): p. 2635–2642.

Hohmann, A.G., Suplita, R.L., Bolton, N.M., et al., *An endocannabinoid mechanism for stress-induced analgesia*. Nature, 2005. **435**(7045): p. 1108–1112.

Holmes, M. V., Asselbergs, F.W., Palmer, T.M., et al., *Mendelian randomization of blood lipids for coronary heart disease*. European Heart Journal, 2015. **36**(9): p. 539–550.

Horvatinovich, J.M., Grogan, E.W., Norris, M., et al., *Soluble CD83 Inhibits T Cell Activation by Binding to the TLR4/MD-2 Complex on CD14 + Monocytes*. The Journal of Immunology, 2017. **198**(6): p. 2286–2301.

Humphries, P., Kenna, P. and Farrar, G.J., *On the molecular genetics of retinitis pigmentosa*. Science, 1992. **256**: p. 804–808.

Hwang, J., Park, Y., Kim, Y., et al., *An extomopathogenic bacterium, Xenorhabdus nematophila, suppresses expression of antimicrobial peptides controlled by toll and IMD pathways by blocking eicosanoid biosynthesis*. Archives of Insect Biochemistry and Physiology, 2013. **83**(3): p. 151–169.

Illig, T., Gieger, C., Zhai, G., et al., *A genome-wide perspective of genetic variation in human metabolism*. Nature Genetics, 2010. **42**(2): p. 137–141.

Imrie, H., Freel, M., Mayosi, B.M., et al., *Association between aldosterone production and variation in the 11-beta-hydroxylase (CYP11B1) gene*. Journal of

Clinical Endocrinology and Metabolism, 2006. **91**(12): p. 5051–5056.

Innes, J.K. and Calder, P.C., *Omega-6 fatty acids and inflammation*. Prostaglandins Leukotrienes and Essential Fatty Acids, 2018. **132**: p. 41–48.

Ishikawa, M., Maekawa, K., Saito, K., et al., *Plasma and serum lipidomics of healthy white adults shows characteristic profiles by subjects' gender and age*. PLoS ONE, 2014. **9**(3): p. 1–12.

Ishikawa, M., Tajima, Y., Murayama, M., et al., *Plasma and Serum from Nonfasting Men and Women Differ in Their Lipidomic Profiles*. Biological and Pharmaceutical Bulletin, 2013. **36**(4): p. 682–685.

Jensen, P.N., Fretts, A.M., Yu, C., et al., *Circulating sphingolipids, fasting glucose, and impaired fasting glucose: The Strong Heart Family Study*. EBioMedicine, 2019. **41**: p. 44–49.

Jiang, L.S., Pu, J., Han, Z.H., et al., *Role of activated endocannabinoid system in regulation of cellular cholesterol metabolism in macrophages*. Cardiovascular Research, 2009. **81**(4): p. 805–813.

Joffres, M., Falaschetti, E., Gillespie, C., et al., *Hypertension prevalence, awareness, treatment and control in national surveys from England, the USA and Canada, and correlation with stroke and ischaemic heart disease mortality: A cross-sectional study*. BMJ Open, 2013. **3**: p. e003423.

Jones PJH., et al. *Modulation of plasma N-acylethanolamine levels and physiological parameters by dietary fatty acid composition in humans.* J Lipid Res, 2014. **55**: 2655-2664.

Joosten MM., et al. *Plasma anandamide and other N-acylethanolamines are correlated with their corresponding free fatty acid levels under both fasting and non-fasting conditions in women.* Nutr Metab, 2010. **7**:49.

Ju, X., Silveira, P.A., Hsu, W.-H., et al., *The Analysis of CD83 Expression on Human Immune Cells Identifies a Unique CD83 + -Activated T Cell Population*. The Journal of Immunology, 2016. **197**(12): p. 4613–4625.

Kalsbeek, A., Veenstra, J., Westra, J., et al., *A genome-wide association study of red-blood cell fatty acids and ratios incorporating dietary covariates: Framingham heart study offspring cohort*. PLoS ONE, 2018. **13**(4): p. e0194882.

Kauhanen D., et al. *Development and validation of a high-throughput LC–MS/MS assay for routine measurement of molecular ceramides.* Analytical and Bioanalytical Chemistry, 2016. **408**(13): p3475-3483.

Keavney, B., Danesh, J., Parish, S., et al., *Fibrinogen and coronary heart disease: Test of causality by 'Mendelian randomization'*. International Journal of Epidemiology, 2006. **35**(4): p. 935–943.

Keavney, B., Mayosi, B., Gaukrodger, N., et al., *Genetic variation at the locus*

encompassing 11-beta hydroxylase and aldosterone synthase accounts for heritability in cortisol precursor (11-deoxycortisol) urinary metabolite excretion. Journal of Clinical Endocrinology and Metabolism, 2005. **90**(2): p. 1072–1077.

Keavney, B., McKenzie, C.A., Connell, J.M.C., et al., *Measured haplotype analysis of the angiotensin-I converting enzyme gene*. Human Molecular Genetics, 1998. **7**(11): p. 1745–1751.

Kendall, A.C., Kiezel-Tsugunova, M., Brownbridge, L.C., et al., *Lipid functions in skin: Differential effects of n-3 polyunsaturated fatty acids on cutaneous ceramides, in a human skin organ culture model*. Biochimica et Biophysica Acta - Biomembranes, 2017. **1859**(9 Pt B): p. 1679–1689.

Kendall, A.C., Pilkington, S.M., Murphy, S.A., et al., *Dynamics of the human skin mediator lipidome in response to dietary ω-3 fatty acid supplementation*. FASEB, 2019. **33**(11): p. 13014–13027.

Khera, A. V, Chaffin, M., Aragam, K.G., et al., *Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations*. Nature Genetics, 2018. **50**(9): p. 1219–1224.

Klarin, D., Damrauer, S.M., Cho, K., et al., *Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program*. Nature Genetics, 2018. **50**(11): p. 1514–1523.

Laaksonen, R., Ekroos, K., Sysi-Aho, M., et al., *Plasma ceramides predict cardiovascular death in patients with stable coronary artery disease and acute coronary syndromes beyond LDL-cholesterol*. European Heart Journal, 2016. **37**(25): p. 1967–1976.

Lackner, C., Boerwinkle, E., Leffert, C.C., et al., *Molecular basis of apolipoprotein (a) isoform size heterogeneity as revealed by pulsed-field gel electrophoresis*. Journal of Clinical Investigation, 1991. **87**(6): p. 2153–2161.

Lang, P.A., Kempe, D.S., Tanneur, V., et al., *Stimulation of erythrocyte ceramide formation by platelet-activating factor*. Journal of Cell Science, 2005. **118**: p. 1233–1243.

Lankinen, M., Schwab, U., Kolehmainen, M., et al., *A Healthy Nordic Diet Alters the Plasma Lipidomic Profile in Adults with Features of Metabolic Syndrome in a Multicenter Randomized Dietary Intervention.* The Journal of nutrition, 2016. **146**: p. 662–672.

Lawes, C.M., Hoorn, S. Vander and Rodgers, A., *Global burden of blood-pressure-related disease, 2001*. The Lancet, 2008. **371**(9623): p. 1513–1518.

Lemaitre, R.N., Tanaka, T., Tang, W., et al., *Genetic loci associated with plasma phospholipid N-3 fatty acids: A Meta-Analysis of Genome-Wide association studies from the charge consortium*. PLoS Genetics, 2011. **7**(7): p. e1002193.

Lemaitre, R.N., Yu, C., Hoofnagle, A., et al., *Circulating sphingolipids, insulin,*

HOMA-IR and HOMA-B: The Strong Heart Family Study Running title: Sphingolipids and insulin resistance markers. Diabetes, 2018. **67**(8): p. 1663–1672.

Li, Y., Sekula, P., Wuttke, M., et al., *Genome-wide association studies of metabolites in patients with CKD identify multiple loci and illuminate tubular transport mechanisms*. Journal of the American Society of Nephrology, 2018a. **29**(5): p. 1513–1524.

Li, Z., Ju, X., Lee, K., et al., *Cd83 is a new potential biomarker and therapeutic target for hodgkin lymphoma*. Haematologica, 2018b. **103**(4): p. 655–665.

Libby, P., Ridker, P.M. and Hansson, G.K., *Progress and challenges in translating the biology of atherosclerosis.* Nature, 2011. **473**(7347): p. 317–325.

Lippert, C., Listgarten, J., Liu, Y., et al., *FaST linear mixed models for genome-wide association studies*. Nature Methods, 2011. **8**(10): p. 833–835.

Loh, P.R., Danecek, P., Palamara, P.F., et al., *Reference-based phasing using the Haplotype Reference Consortium panel*. Nature Genetics, 2016. **48**(11): p. 1443–1448.

Long, T., Hicks, M., Yu, H.C., et al., *Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites*. Nature Genetics, 2017. **49**: p. 568–578.

Lopez, X., Goldfine, A.B., Holland, W.L., et al., *Plasma ceramides are elevated in female children and adolescents with type 2 diabetes*. Journal of Pediatric Endocrinology and Metabolism, 2013. **26**(9–10): p. 995–998.

Lorenz, K. and Cohen, B.A., *Small- and large-effect quantitative trait locus interactions underlie variation in yeast sporulation efficiency*. Genetics, 2012. **192**(3): p. 1123–1132.

Lötzer, K., Funk, C.D. and Habenicht, A.J.R., *The 5-lipoxygenase pathway in arterial wall biology and atherosclerosis*. Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids, 2005. **1736**(1): p. 30–37.

Lu, W., Su, X., Klein, M.S., et al., *Metabolite Measurement: Pitfalls to Avoid and Practices to Follow*. Annual Review of Biochemistry, 2017. **86**: p. 277–304.

M. Keppel Hesselink, J., *New Targets in Pain, Non-Neuronal Cells, and the Role of Palmitoylethanolamide*. The Open Pain Journal, 2012. **5**: p. 12–23.

Maceyka, M. and Spiegel, S., *Sphingolipid metabolites in inflammatory disease.* Nature, 2014. **510**(7503): p. 58–67.

Mach, F., Montecucco, F. and Steffens, S., *Effect of blockage of the endocannabinoid system by CB1 antagonism on cardiovascular risk*. Pharmacological Reports, 2009. **61**(1): p. 13–21.

Mahajan, A., Go, M.J., Zhang, W., et al., *Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility*. Nature Genetics, 2014. **46**(3): p. 234–244.

Mallet, C., Dubray, C. and Dualé, C., *FAAH inhibitors in the limelight, but regrettably*. International Journal of Clinical Pharmacology and Therapeutics, 2016. **54**(7): p. 498–501.

Marees, A.T., de Kluiver, H., Stringer, S., et al., *A tutorial on conducting genome-wide association studies: Quality control and statistical analysis*. International Journal of Methods in Psychiatric Research, 2018. **27**(2): p. 1–10.

Masoodi, M., Mir, A.A., Petasis, N.A., et al., *Simultaneous lipidomic analysis of three families of bioactive lipid mediators leukotrienes, resolvins, protectins and related hydroxy-fatty acids by liquid chromatography/electrospray tandem mass spectrometry Mojgan*, 2008. **22**(2): p. 75–83.

Masoodi, M. and Nicolaou, A., *Europe PMC Funders Group Lipidomic Analysis of Twenty Seven Prostanoids and Isoprostanes by Electrospray Liquid Chromatography / Tandem Mass Spectrometry*, 2007. **20**(20): p. 3023–3029.

Massey, K.A. and Nicolaou, A., *Lipidomics of oxidized polyunsaturated fatty acids*. Free Radical Biology and Medicine, 2013. **59**: p. 45–55.

Masukawa, Y., Narita, H., Sato, H., et al., *Comprehensive quantification of ceramide species in human stratum corneum.* Journal of lipid research, 2009. **50**(8): p. 1708–1719.

Masukawa, Y., Narita, H., Shimizu, E., et al., *Characterization of overall ceramide species in human* stratum corneum. Journal of Lipid Research, 2008. **49**(7): p. 1466–1476.

Matuszewski, B.K., Constanzer, M.L. and Chavez-Eng, C.M., *Strategies for the assessment of matrix effect in quantitative bioanalytical methods based on HPLC-MS/MS*. Analytical Chemistry, 2003. **75**(13): p. 3019–3030.

Mayosi, B.M., Avery, P.J., Baker, M., et al., *Genotype at the -174G/C polymorphism of the interleukin-6 gene is associated with common carotid artery intimal-medial thickness: Family study and meta-analysis*. Stroke, 2005. **36**: p. 2215–2219.

Mayosi, B.M., Avery, P.J., Farrall, M., et al., *Genome-wide linkage analysis of electrocardiographic and echocardiographic left ventricular hypertrophy in families with hypertension*. European Heart Journal, 2008. **29**(4): p. 525–530.

Mayosi, B.M., Keavney, B., Kardos, A., et al., *Electrocardiographic measures of left ventricular hypertrophy show greater heritability than echocardiographic left ventricular mass: A family study*. European Heart Journal, 2002. **23**(24): p. 1963–1971.

Mayosi, B.M., Keavney, B., Watkins, H., et al., *Measured haplotype analysis of the*

aldosterone synthase gene and heart size. European Journal of Human Genetics, 2003. **11**: p. 395–401.

McCarthy, S., Das, S., Kretzschmar, W., et al., *A reference panel of 64,976 haplotypes for genotype imputation*. Nature Genetics, 2016. **48**(10): p. 1279–1283.

McCormick, S.P.A., *Lipoprotein(a): Biology and Clinical Importance*. The Clinical Biochemist Reviews, 2004. **25**(1): p. 69–80.

Mei, H., Hsieh, Y., Nardo, C., et al., *Investigation of matrix effects in bioanalytical high-performance liquid chromatography/tandem mass spectrometric assays: Application to drug discovery*. Rapid Communications in Mass Spectrometry, 2003. **17**(1): p. 97–103.

Meikle PJ, et al. *Plasma Lipid Profiling Shows Similar Associations With Prediabetes and Type 2 Diabetes.* PLoS ONE, 2013. **8**(9): p. e74341.

Mercado, N., Kizawa, Y., Ueda, K., et al., *Activation of transcription factor Nrf2 signalling by the sphingosine kinase inhibitor SKI-II is mediated by the formation of Keap1 dimers*. PLoS ONE, 2014. **9**(2): p. e88168.

Miller TM et al. *A rapid UPLC-MS/MS assay for eicosanoids in human plasma: Application to evaluate niacin responsivity*. PLEFA, 2017. **136:** 153-159.

Mokry, L.E., Ahmad, O., Forgetta, V., et al., *Mendelian randomisation applied to drug development in cardiovascular disease: A review*. Journal of Medical Genetics, 2015. **52**(2): p. 71–79.

Murphy, R.C. and Axelsen, P.H., *Mass spectrometric analysis of long-chain lipids*. Mass Spectrometry Reviews, 2011a. **30**(4): p. 579–599.

Murphy, R.C. and Gaskell, S.J., *New applications of mass spectrometry in lipid analysis*. Journal of Biological Chemistry, 2011b. **286**(29): p. 25427–25433.

Nagy, L., Tontonoz, P., Alvarez, J.G.A., et al., *Oxidized LDL regulates macrophage gene expression through ligand activation of PPARγ*. Cell, 1998. **93**(2): p. 229–240.

Nakamura, H., Dae Kyong Kim, Philbin, D.M., et al., *Heparin-enhanced plasma phospholipase A2 activity and prostacyclin synthesis in patients undergoing cardiac surgery*. Journal of Clinical Investigation, 1995. **95**(3): p. 1062–1070.

Nature, *Accounting for sex in the genome*. Nature Medicine, 2017. **23**: p. 1243.

Nethononda, R.M., McGurk, K.A., Whitworth, P., et al., *Marked variation in heritability estimates of left ventricular mass depending on modality of measurement*. Scientific Reports, 2019. **9**: p. 13556.

Nikpay, M., Goel, A., Won, H.-H., et al., *A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease.* Nature

genetics, 2015. **47**(10): p. 1121–1130.

Nordestgaard, B.G., Langsted, A., Mora, S., et al., *Fasting is not routinely required for determination of a lipid profile: Clinical and Laboratory implications including flagging at desirable concentration cutpoints-A joint consensus statement from the European Atherosclerosis Society and European Federat*. Clinical Chemistry, 2016. **62**(7): p. 930–946.

O'Donnell, C.J. and Nabel, E.G., *Genomics of cardiovascular disease.* The New England journal of medicine, 2011. **365**(22): p. 2098–2109.

Oni-Orisan, A., Edin, M.L., Lee, J.A., et al., *Cytochrome P450-derived epoxyeicosatrienoic acids and coronary artery disease in humans: a targeted metabolomics study*. Journal of Lipid Research, 2016. **57**(1): p. 109–119.

Ott, J., Kamatani, Y. and Lathrop, M., *Family-based designs for genome-wide association studies*. Nature Reviews Genetics, 2011. **12**: p. 465–474.

Ozaki, K., Ohnishi, Y., Iida, A., et al., *Functional SNPs in the lymphotoxin-α gene that are associated with susceptibility to myocardial infarction*. Nature Genetics, 2002. **32**: p. 650–654.

Pacher, P. and Kunos, G., *Modulating the endocannabinoid system in human health and disease - Successes and failures*. FEBS Journal, 2013. **280**(9): p. 1918–1943.

Palomino-Doza, J., Rahman, T.J., Avery, P.J., et al., *Ambulatory blood pressure is associated with polymorphic variation in P2X receptor genes*. Hypertension, 2008. **52**(5): p. 980–985.

Pan, W., Yu, J., Shi, R., et al., *Elevation of ceramide and activation of secretory acid sphingomyelinase in patients with acute coronary syndromes.* Coronary artery disease, 2014. **25**(3): p. 230–235.

Park, T.S., Rosebury, W., Kindt, E.K., et al., *Serine palmitoyltransferase inhibitor myriocin induces the regression of atherosclerotic plaques in hyperlipidemic ApoE-deficient mice*. Pharmacological Research, 2008. **58**(1): p. 45–51.

Peeters, M.W., Thomis, M.A., Maes, H.H.M., et al., *Genetic and environmental causes of tracking in explosive strength during adolescence*. Behavior Genetics, 2005. **35**(5): p. 551–563.

Perry, D.K., Carton, J., Shah, A.K., et al., *Serine palmitoyltransferase regulates de novo ceramide generation during etoposide-induced apoptosis*. Journal of Biological Chemistry, 2000. **275**(12): p. 9078–9084.

Perry, D.K. and Hannun, Y.A., *The role of ceramide in cell signaling.* Biochimica et biophysica acta, 1998. **1436**(1–2): p. 233–243.

Petersen, A.K., Krumsiek, J., Wägele, B., et al., *On the hypothesis-free testing of metabolite ratios in genome-wide and metabolome-wide association studies*. BMC Bioinformatics, 2012. **13**(120): p. 1–7.

Petroni, A., Blasevich, M., Salami, M., et al., *Inhibition of platelet aggregation and eicosanoid production by phenolic components of olive oil*. Thrombosis Research, 1995. **78**(2): p. 151–160.

Pettus, B.J., Chalfant, C.E. and Hannun, Y.A., *Ceramide in apoptosis: An overview and current perspectives*. Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids, 2002. **1585**: p. 114–125.

Piercy, K.L. and Troiano, R.P., *Physical Activity Guidelines for Americans From the US Department of Health and Human Services*. Circulation. Cardiovascular quality and outcomes, 2018. **11**(11): p. e005263.

Prazma, C.M., Yazawa, N., Fujimoto, Y., et al., *CD83 Expression Is a Sensitive Marker of Activation Required for B Cell and CD4 + T Cell Longevity In Vivo*. The Journal of Immunology, 2007. **179**: p. 4550–4562.

Public Health England, *Tackling high blood pressure: an update*. Blood Pressure System Leadership Board. PHE publications, 2018. p. 1-37. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/672554/Tackling_high_blood_pressure_an_update.pdf

Purcell, S., Neale, B., Todd-Brown, K., et al., *PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses*. The American Journal of Human Genetics, 2007. **81**(3): p. 559–575.

Quehenberger, O., Armando, A.M., Brown, A.H., et al., *Lipidomics reveals a remarkable diversity of lipids in human plasma.* Journal of lipid research, 2010. **51**(11): p. 3299–3305.

Quehenberger, O. and Dennis, E.A., *The human plasma lipidome*. New England Journal of Medicine, 2011. **365**(19): p. 1812–1823.

Rahman, T.J., Mayosi, B.M., Hall, D., et al., *Common variation at the 11-beta hydroxysteroid dehydrogenase type 1 gene is associated with left ventricular mass*. Circulation: Cardiovascular Genetics, 2011a. **4**(2): p. 156–162.

Rahman, T.J., Walker, E.A., Mayosi, B.M., et al., *Genotype at the P554L variant of the hexose-6 phosphate dehydrogenase gene is associated with carotid intima-medial thickness*. PLoS ONE, 2011b. **6**(8): p. e23248.

Rhodes, B., Meek, J., Whittaker, J.C., et al., *Quantification of the genetic component of basal C-reactive protein expression in SLE nuclear families*. Annals of Human Genetics, 2008. **72**(5): p. 611–620.

Rimm, E.B., Appel, L.J., Chiuve, S.E., et al., *Seafood Long-Chain n-3 Polyunsaturated Fatty Acids and Cardiovascular Disease: A Science Advisory From the American Heart Association*. Circulation, 2018. **138**(1): p. e35–e47.

Rodríguez de Fonseca, F., Navarro, M., Gómez, R., et al., *An anorexic lipid mediator regulated by feeding.* Nature, 2001. **414**(6860): p. 209–212.

Rokitansky, C. Von, *A manual of pathological anatomy*. Printed for the Sydenham society, 1849.: p. 1–911.

Rolim, A.E.H., Henrique-Arajo, R., Ferraz, E.G., et al., *Lipidomics in the study of lipid metabolism: Current perspectives in the omic sciences*. Gene, 2015. **554**(2): p. 131–139.

Ross, R., Glomset, J. and Harker, L., *Response to injury and atherogenesis*. Am J Pathol, 1977. **86**(3): p. 675–684.

Rubenstein, P., Walker, M., Carpenter, C., et al., *Genetics of HLA-disease associations: The use of the haplotype relative risk (HRR) and the "haplo-delta" (Dh) estimates in juvenile diabetes from three racial groups*. Human Immunology, 1981.(3): p. 384.

Ryan, V.H., Primiani, C.T., Rao, J.S., et al., *Coordination of gene expression of arachidonic and docosahexaenoic acid cascade enzymes during human brain development and aging*. PLoS ONE, 2014. **9**(6): p. e100858.

Salvagno, G.L., Sanchis-Gomar, F., Picanza, A., et al., *Red blood cell distribution width: A simple parameter with multiple clinical applications*. Critical Reviews in Clinical Laboratory Sciences, 2015. **52**(2): p. 86–105.

Sapieha, P., Stahl, A., Chen, J., et al., *5-Lipoxygenase metabolite 4-HDHA is a mediator of the antiangiogenic effect of ω-3 polyunsaturated fatty acids*. Science Translational Medicine, 2011. **3**(69): p. 69ra12.

Schaich, C.L., Shaltout, H. a., Brosnihan, K.B., et al., *Acute and chronic systemic CB1 cannabinoid receptor blockade improves blood pressure regulation and metabolic profile in hypertensive (mRen2)27 rats*. Physiological Reports, 2014. **2**(8): p. e12108–e12108.

Schuck, R.N., Theken, K.N., Edin, M.L., et al., *Cytochrome P450-derived eicosanoids and vascular dysfunction in coronary artery disease patients*. Atherosclerosis, 2013. **227**(2): p. 442–448.

Shah, S.H., Hauser, E.R., Bain, J.R., et al., *High heritability of metabolomic profiles in families burdened with premature cardiovascular disease.* Molecular systems biology, 2009. **5**(258): p. 1–7.

Shaner, R.L., Allegood, J.C., Park, H., et al., *Quantitative analysis of sphingolipids for lipidomics using triple quadrupole and quadrupole linear ion trap mass spectrometers*. J Lipid Res, 2009. **50**(8): p. 1692–1707.

Shin, S.Y., Fauman, E.B., Petersen, A.K., et al., *An atlas of genetic influences on human blood metabolites*. Nature Genetics, 2014. **46**(6): p. 543–550.

Silventoinen, K., Sammalisto, S., Perola, M., et al., *Heritability of Adult Body Height: A Comparative Study of Twin Cohorts in Eight Countries*. Twin Research, 2003. **6**: p. 399–408.

Sim, M.S., Hatim, A., Reynolds, G.P., et al., *Association of a functional FAAH polymorphism with methamphetamine-induced symptoms and dependence in a Malaysian population*. Pharmacogenomics, 2013. **14**(5): p. 505–514.

Simopoulos, A.P., *Evolutionary aspects of diet, the omega-6/omega-3 ratio and genetic variation: nutritional implications for chronic diseases*. Biomedicine and Pharmacotherapy, 2006. **60**(9): p. 502–507.

Singer, V.L., Jones, L.J., Yue, S.T., et al., *Characterization of PicoGreen reagent and development of a fluorescence- based solution assay for double-stranded DNA quantitation*. Analytical Biochemistry, 1997. **249**(2): p. 228–238.

Sipe, J.C., Chiang, K., Gerber, A.L., et al., *A missense mutation in human fatty acid amide hydrolase associated with problem drug use*. Proceedings of the National Academy of Sciences, 2002. **99**(12): p. 8394–8399.

Sipe, J.C., Scott, T.M., Murray, S., et al., *Biomarkers of endocannabinoid system activation in severe obesity*. PLoS ONE, 2010. **5**(1): p. 1–6.

Sipe, J.C., Waalen, J., Gerber, A., et al., *Overweight and obesity associated with a missense polymorphism in fatty acid amide hydrolase (FAAH)*. International Journal of Obesity, 2005. **29**(7): p. 755–759.

Siskind, L.J. and Colombini, M., *The lipids C2- and C16-ceramide form large stable channels: Implications for apoptosis*. Journal of Biological Chemistry, 2000. **275**: p. 38640–38644.

Siskind, L.J., Kolesnick, R.N. and Colombini, M., *Ceramide channels increase the permeability of the mitochondrial outer membrane to small proteins*. Journal of Biological Chemistry, 2002. **277**: p. 26796–26803.

Siskind, L.J., Mullen, T.D. and Obeid, L.M., *The role of ceramide in cell regulation*. Handbook of Cell Signaling, 2/e, 2010. **2**: p. 1201–1211.

Van Smeden, J., Bouwstra, J.A., Hoppel, L., et al., *LC/MS analysis of stratum corneum lipids: Ceramide profiling and discovery*. Journal of Lipid Research, 2011. **52**(6): p. 1211–1221.

Smith, G.D. and Hemani, G., *Mendelian randomization: Genetic anchors for causal inference in epidemiological studies*. Human Molecular Genetics, 2014. **23**(R1): p. R89–R98.

Spielman, R.S. and Ewens, W.J., *A Sibship Test for Linkage in the Presence of Association: The Sib Transmission/Disequilibrium Test*. The American Journal of Human Genetics, 1998. **62**(2): p. 450–458.

Stanley, D. and Kim, Y., *Prostaglandins and other eicosanoids in insects: Biosynthesis and biological actions*. Frontiers in Physiology, 2019. **10**: p. 1–13.

Steinbrecher, U.P., Zhang, H. and Lougheed, M., *Role of oxidatively modified LDL in atherosclerosis*. Free Radical Biology and Medicine, 1990. **9**(2): p. 155–168.

Stephenson, D.J., Hoeferlin, L.A. and Chalfant, C.E., *Lipidomics in translational research and the clinical significance of lipid-based biomarkers*. Translational Research, 2017. **189**: p. 13–29.

Sudlow, C., Gallacher, J., Green, J., et al., *UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old Age*. PLOS Medicine, 2015. **12**(3): p. e1001779.

Suhre, K., Shin, S.Y., Petersen, A.K., et al., *Human metabolic individuality in biomedical and pharmaceutical research*. Nature, 2011. **477**(7362): p. 54–60.

Sullards, M.C., *Analysis of sphingomyelin, glucosylceramide, ceramide, sphingosine, and sphingosine 1-phosphate by tandem mass spectrometry.* Methods in enzymology, 2000. **312**: p. 32–45.

Surma, M.A., Herzog, R., Vasilj, A., et al., *An automated shotgun lipidomics platform for high throughput, comprehensive, and quantitative analysis of blood plasma intact lipids*. European Journal of Lipid Science and Technology, 2015. **117**(10): p. 1540–1549.

T'Kindt, R., Jorge, L., Dumont, E., et al., *Profiling and characterizing skin ceramides using reversed-phase liquid chromatography-quadrupole time-of-flight mass spectrometry*. Analytical Chemistry, 2012. **84**(1): p. 403–411.

Tabassum, R., Rämö, J.T., Ripatti, P., et al., *Genetic architecture of human plasma lipidome and its link to cardiovascular disease*. Nature Communications, 2019. **10**(1): p. 4329.

Tanaka, T., Shen, J., Abecasis, G.R., et al., *Genome-wide association study of plasma polyunsaturated fatty acids in the InCHIANTI study*. PLoS Genetics, 2009. **5**(1): p. 1–8.

Tarasov, K., Ekroos, K., Suoniemi, M., et al., *Molecular lipids identify cardiovascular risk and are efficiently lowered by simvastatin and PCSK9 deficiency*. Journal of Clinical Endocrinology and Metabolism, 2014. **99**(1): p. 45–52.

Terao, C., Bayoumi, N., McKenzie, C.A., et al., *Quantitative variation in plasma angiotensin-i converting enzyme activity shows allelic heterogeneity in the ABO blood group locus*. Annals of Human Genetics, 2013. **77**(6): p. 465–471.

Teslovich, T.M., Musunuru, K., Smith, A. V, et al., *Biological, clinical and population relevance of 95 loci for blood lipids.* Nature, 2010. **466**(7307): p. 707–713.

Teumer, A., *Common Methods for Performing Mendelian Randomization*. Frontiers in Cardiovascular Medicine, 2018. **5**: p. 51.

Theken, K.N., Schuck, R.N., Edin, M.L., et al., *Evaluation of cytochrome P450-derived eicosanoids in humans with stable atherosclerotic cardiovascular disease*. Atherosclerosis, 2012. **222**(2): p. 530–536.

Timpson, N.J., Wade, K.H. and Smith, G.D., *Mendelian randomization: Application to cardiovascular disease*. Current Hypertension Reports, 2012. **14**(1): p. 29–37.

Topol, E.J., Smith, J., Plow, E.F., et al., *Genetic susceptibility to myocardial infarction and coronary artery disease*. Human Molecular Genetics, 2006. **15**(2): p. 117–123.

Tsakelidou, E., Virgiliou, C., Valianou, L., et al., *Sample preparation strategies for the effective quantitation of hydrophilic metabolites in serum by multi-targeted HILIC-MS/MS*. Metabolites, 2017. **7**(3): p. metabo7020013.

Turner, S., Armstrong, L.L., Bradford, Y., et al., *Quality control procedures for genome-wide association studies*. Current Protocols in Human Genetics, 2011. **68**(1): p. 1.19.1-1.19.18.

Uhlen, M., Fagerberg, L., Hallstrom, B.M., et al., *Tissue-based map of the human proteome*. Science, 2015. **347**(6220): p. 1260419–1260419.

Urquhart, P., Nicolaou, A. and Woodward, D.F., *Endocannabinoids and their oxygenation by cyclo-oxygenases, lipoxygenases and other oxygenases*. Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids, 2015. **1851**(4): p. 366–376.

Lo Verme, J., Fu, J., Astarita, G., et al., *The nuclear receptor peroxisome proliferator-activated receptor-α mediates the anti-inflammatory actions of palmitoylethanolamide*. Molecular Pharmacology, 2005. **67**: p. 15–19.

Vickers, M.A., Green, F.R., Terry, C., et al., *Genotype at a promoter polymorphism of the interleukin-6 gene is associated with baseline levels of plasma C-reactive protein*. Cardiovascular Research, 2002. **53**(4): p. 1029–1034.

Vila, L., Martinez-Perez, A., Camacho, M., et al., *Heritability of thromboxane A2 and prostaglandin E2 biosynthetic machinery in a spanish population*. Arteriosclerosis, Thrombosis, and Vascular Biology, 2010. **30**(1): p. 128–134.

Visscher, P.M., Hill, W.G. and Wray, N.R., *Heritability in the genomics era - Concepts and misconceptions*. Nature Reviews Genetics, 2008. **9**: p. 255–266.

Voight, B.F., Peloso, G.M., Orho-Melander, M., et al., *Plasma HDL cholesterol and risk of myocardial infarction: A mendelian randomisation study*. The Lancet, 2012. **380**(9841): p. 572–580.

Wall, R., Ross, R.P., Fitzgerald, G.F., et al., *Fatty acids from fish: The anti-inflammatory potential of long-chain omega-3 fatty acids*. Nutrition Reviews, 2010. **68**(5): p. 280–289.

Wang, D.D., Toledo, E., Hruby, A., et al., *Plasma ceramides, mediterranean diet, and incident cardiovascular disease in the PREDIMED trial*. Circulation, 2017. **135**(21): p. 2028–2040.

Waters Corporation, *Signal to Noise Ratio*. *Performance perspectives 2487*.

[accessed 9 October 2019].

Weir J. et al. *Plasma lipid profiling in a large population-based cohort.* J Lipid Res, 2013. **54**(10): p. 2898-2908.

Whelton, P.K., Carey, R.M., Aronow, W.S., et al., *Guidelines for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: Executive Summary*. Journal of the American Society of Hypertension, 2018. **71**(19): p. e127–e248.

Willer, C.J., Schmidt, E.M., Sengupta, S., et al., *Discovery and refinement of loci associated with lipid levels*. Nature Genetics, 2013. **45**(11): p. 1274–1285.

Wilson, P.W., Garrison, R.J., Castelli, W.P., et al., *Prevalence of coronary heart disease in the Framingham Offspring Study: role of lipoprotein cholesterols.* The American journal of cardiology, 1980. **46**(4): p. 649–654.

Wilson, R.I. and Nicoll, R. a, *Endogenous cannabinoids mediate retrograde signalling at hippocampal synapses.* Nature, 2001. **410**(6828): p. 588–592.

Wong, A., Sagar, D.R., Ortori, C. a, et al., *Simultaneous tissue profiling of eicosanoid and endocannabinoid lipid families in a rat model of osteoarthritis.* Journal of lipid research, 2014. **55**(9): p. 1902–1913.

Wright, R.S., *Recent clinical trials evaluating benefit of drug therapy for modification of HDL cholesterol*. Current Opinion in Cardiology, 2013. **28**(4): p. 389–398.

Wright, S., *Correlation and causation*. Journal of Agricultural Research, 1921. **20**: p. 557–585.

Xue, B., Yang, Z., Wang, X., et al., *Omega-3 Polyunsaturated Fatty Acids Antagonize Macrophage Inflammation via Activation of AMPK/SIRT1 Pathway*. PLoS ONE, 2012. **7**(10): p. 2–7.

Yang, J., Benyamin, B., McEvoy, B.P., et al., *Common SNPs explain a large proportion of the heritability for human height*. Nature Genetics, 2010. **42**(7): p. 565–569.

Yang, T., Peng, R., Guo, Y., et al., *The role of 14,15-dihydroxyeicosatrienoic acid levels in inflammation and its relationship to lipoproteins.* Lipids in health and disease, 2013. **12**: p. 151.

Yoshimoto, T. and Takahashi, Y., *Arachidonate 12-lipoxygenases*. Prostaglandins and Other Lipid Mediators, 2002. **68–69**: p. 245–262.

Yu, J., Pan, W., Shi, R., et al., *Ceramide Is Upregulated and Associated With Mortality in Patients With Chronic Heart Failure*. Canadian Journal of Cardiology, 2015. **31**(3): p. 357–363.

Yu, Z., Xu, F., Huse, L.M., et al., *Soluble epoxide hydrolase regulates hydrolysis of vasoactive epoxyeicosatrienoic acids.* Circulation research, 2000. **87**(11): p.

992–998.

Zaitlen, N., Kraft, P., Patterson, N., et al., *Using Extended Genealogy to Estimate Components of Heritability for 23 Quantitative and Dichotomous Traits*. PLoS Genetics, 2013. **9**(5): p. e1003520.

Zanger, U.M. and Schwab, M., *Cytochrome P450 enzymes in drug metabolism: Regulation of gene expression, enzyme activities, and impact of genetic variation*. Pharmacology and Therapeutics, 2013. **138**(1): p. 103–141.

Zhou, X., Wu, W., Chen, J., et al., *AMP-activated protein kinase is required for the anti-adipogenic effects of alpha-linolenic acid*. Nutrition and Metabolism, 2015. **12**(1): p. 6–13.

Appendix

Table 0.1: Correlation between mass spectrometry batch and the quality control plasma samples to assess for collinearity

The traits were assessed by correlation (Pearson correlation with two-tailed P-values, Graph Pad Prism 7 software). R, correlation coefficient. A Graph Pad Prism P-value of <0.0001 is depicted here as 0.000.

QC	Class	R	P-value	n
CER[A(22)S(18)]	CER	-0.62	0.000	1016
CER[A(24)S(18)]	CER	0.07	0.026	1016
CER[A(26)S(18)]	CER	0.38	0.000	946
C18DS	CER	0.39	0.000	1016
C18S	CER	0.09	0.003	1016
C18S1P	CER	0.04	0.162	1006
CER[N(16)S(18)]	CER	0.04	0.220	1016
CER[N(20)S(18)]	CER	-0.41	0.000	1016
CER[N(22)DS(18)]	CER	-0.62	0.000	1016
CER[N(22)S(18)]	CER	-0.46	0.000	1016
CER[N(22)S(19)]	CER	-0.83	0.000	1016
CER[N(23)S(18)]	CER	-0.70	0.000	1016
CER[N(23)S(20)]	CER	-0.16	0.000	1016
CER[N(24)DS(18)]	CER	-0.54	0.000	1016
CER[N(24)DS(19)]	CER	-0.29	0.000	1016
CER[N(24)DS(20)]	CER	0.20	0.000	1016
CER[N(24)S(16)]	CER	-0.74	0.000	1016
CER[N(24)S(17)]	CER	-0.74	0.000	1016
CER[N(24)S(18)]	CER	-0.72	0.000	1016
CER[N(24)S(19)]	CER	-0.76	0.000	1016
CER[N(24)S(20)]	CER	0.29	0.000	1016
CER[N(24)S(22)]	CER	0.42	0.000	1016
CER[N(25)DS(18)]	CER	0.13	0.000	1016
CER[N(25)S(20)]	CER	0.35	0.000	1016
CER[N(26)DS(18)]	CER	0.43	0.000	1016
CER[N(26)S(18)]	CER	0.30	0.000	1016
CER[N(26)S(19)]	CER	0.16	0.000	1016
CER[N(27)S(18)]	CER	0.34	0.000	1016
CER[N(28)S(18)]	CER	0.24	0.000	1016
CER[N(29)S(18)]	CER	0.05	0.083	1016
DHET1112	Eico	-0.19	0.008	204

HETE11	Eico	0.03	0.683	204
EpOME1213	Eico	-0.41	0.000	204
DiHOME1213	Eico	-0.74	0.000	204
HETE12	Eico	-0.95	0.000	204
HODE13	Eico	-0.81	0.000	204
HOTrE13	Eico	0.24	0.000	204
OxoODE13	Eico	0.14	0.049	204
DHET1415	Eico	-0.05	0.473	204
HETE15	Eico	-0.90	0.000	204
DiHDPA1920	Eico	0.03	0.656	204
HDHA4	Eico	0.43	0.000	204
HETE5	Eico	0.33	0.000	204
EpOME910	Eico	-0.70	0.000	204
DiHOME910	Eico	-0.41	0.000	204
HODE9	Eico	-0.59	0.000	204
HOTrE9	Eico	0.03	0.656	204
OxoODE9	Eico	-0.02	0.782	204
TransEKODE	Eico	-0.61	0.000	204
AEA	NAE	0.26	0.000	1016
DHEA	NAE	0.31	0.000	1016
DPEA	NAE	0.33	0.000	1016
HEA	NAE	0.41	0.000	1016
LEA	NAE	0.06	0.062	1016
OEA	NAE	-0.40	0.000	1016
PEA	NAE	0.30	0.000	1016
POEA	NAE	0.62	0.000	1016
PDEA	NAE	0.52	0.000	1016
STEA	NAE	-0.42	0.000	1016
VEA	NAE	-0.40	0.000	1016

Table 0.2: A list of published GWAS assessed by 2SMR analysis

ID, 2SMR software ID; Population, population type; UKB, UK Biobank; INT, INTERVAL; BiL, UK BiLEVE.

Author	Consortium	ID	N(Cases)	N (Controls)	N(SNP)	Pop	N (total)	Trait
Nikpay	CARDIoGRAMplusC4D	7	60801	123504	9455779	Mixed	184305	Coronary heart disease
Mahajan A	DIAGRAM	23	26488	83964	2915012	Mixed	110452	Type 2 diabetes
van der Harst P	HaemGen	275	NA	NA	2589455	Mixed	66214	Red blood cell count
Gieger C	HaemGen	1008	NA	NA	2703394	European	66867	Platelet count
Astle W	UKB+INT+BiL	1247	NA	NA	29462512	European	164454	Mean platelet volume
Astle W	UKB+INT+BiL	1248	NA	NA	29483746	European	172378	Eosinophil percentage of white cells
Astle W	UKB+INT+BiL	1249	NA	NA	29486177	European	172952	Red blood cell count
Astle W	UKB+INT+BiL	1250	NA	NA	29483230	European	172433	Mean corpuscular volume
Astle W	UKB+INT+BiL	1251	NA	NA	29465077	European	166066	Platelet count
Astle W	UKB+INT+BiL	1252	NA	NA	29484426	European	173039	Haematocrit
Astle W	UKB+INT+BiL	1253	NA	NA	29486070	European	172851	Mean corpuscular haemoglobin concentration
Astle W	UKB+INT+BiL	1254	NA	NA	29485759	European	172275	Eosinophil counts
Astle W	UKB+INT+BiL	1255	NA	NA	29463315	European	164339	Plateletcrit
Astle W	UKB+INT+BiL	1256	NA	NA	29480509	European	169545	Granulocyte percentage of myeloid white cells
Astle W	UKB+INT+BiL	1257	NA	NA	29481677	European	170494	Monocyte percentage of white cells
Astle W	UKB+INT+BiL	1258	NA	NA	29485724	European	172435	White blood cell count
Astle W	UKB+INT+BiL	1259	NA	NA	29480430	European	170761	High light scatter reticulocyte count
Astle W	UKB+INT+BiL	1260	NA	NA	29480170	European	170763	High light scatter reticulocyte percentage of red cells
Astle W	UKB+INT+BiL	1261	NA	NA	29482518	European	170384	Sum neutrophil eosinophil counts
Astle W	UKB+INT+BiL	1262	NA	NA	29481601	European	169822	Granulocyte count
Astle W	UKB+INT+BiL	1263	NA	NA	29483564	European	172925	Haemoglobin concentration
Astle W	UKB+INT+BiL	1264	NA	NA	29461105	European	164433	Platelet distribution width
Astle W	UKB+INT+BiL	1265	NA	NA	29482376	European	170536	Eosinophil percentage of granulocytes
Astle W	UKB+INT+BiL	1266	NA	NA	29484325	European	171846	White blood cell count (basophil)

Astle W	UKB+INT+BiL	1267	NA	NA	29480520	European	170690	Reticulocyte fraction of red cells
Astle W	UKB+INT+BiL	1268	NA	NA	29480759	European	170143	Sum basophil neutrophil counts
Astle W	UKB+INT+BiL	1269	NA	NA	29484006	European	171529	Red cell distribution width
Astle W	UKB+INT+BiL	1270	NA	NA	29479992	European	170641	Reticulocyte count
Astle W	UKB+INT+BiL	1271	NA	NA	29482650	European	170672	Neutrophil percentage of granulocytes
Astle W	UKB+INT+BiL	1272	NA	NA	29485063	European	171771	Sum eosinophil basophil counts
Astle W	UKB+INT+BiL	1273	NA	NA	29482454	European	170721	Monocyte count
Astle W	UKB+INT+BiL	1274	NA	NA	29478559	European	169219	Myeloid white cell count
Astle W	UKB+INT+BiL	1275	NA	NA	29484106	European	171643	Lymphocyte counts
Astle W	UKB+INT+BiL	1276	NA	NA	29479929	European	170548	Immature fraction of reticulocytes
Astle W	UKB+INT+BiL	1277	NA	NA	29481373	European	170702	Neutrophil count

Table 0.3: Descriptive statistics of the measured lipid species in the range finding study

Data is shown as mean and standard deviation. The concentrations are pmol/ml for CER and pg/ml for Eico and NAE.

Class	Lipid	Mean	SD	N
CER	A(22)S(18)	2.39	0.94	204
CER	A(24)S(18)	3.34	0.82	204
CER	A(26)S(18)	0.10	0.08	201
CER	C18DS	0.004	0.001	204
CER	C18S	0.07	0.04	198
CER	C18S1P	0.29	0.37	194
CER	N(16)S(18)	1.51	0.79	204
CER	N(20)S(18)	0.44	0.25	204
CER	N(22)DS(18)	0.80	0.50	204
CER	N(22)S(18)	6.87	3.49	204
CER	N(22)S(19)	1.52	0.88	204
CER	N(23)S(18)	57.17	21.41	204
CER	N(23)S(20)	2.30	0.74	204
CER	N(24)DS(18)	12.49	7.19	204
CER	N(24)DS(19)	3.52	2.27	204
CER	N(24)DS(20)	1.64	0.95	204
CER	N(24)S(16)	2.42	1.41	204
CER	N(24)S(17)	14.85	5.99	204
CER	N(24)S(18)	190.40	68.53	204
CER	N(24)S(19)	65.79	25.46	204
CER	N(24)S(20)	11.71	4.24	204
CER	N(24)S(22)	1.37	0.85	204
CER	N(25)DS(18)	1.34	0.76	204
CER	N(25)S(20)	1.07	0.43	204
CER	N(26)DS(18)	0.81	0.44	204
CER	N(26)S(18)	30.71	8.51	204
CER	N(26)S(19)	3.91	2.08	204
CER	N(27)S(18)	1.58	1.05	204
CER	N(28)S(18)	0.61	0.30	204
CER	N(29)S(18)	0.83	1.05	204
CER_sumorratio	ratio22to24	0.04	0.01	204
CER_sumorratio	ratio20to24	0.002	0.0007	204
CER_sumorratio	totalsphingo	421.89	125.95	204
CER_sumorratio	ratio16to24	0.01	0.003	204
CER_sumorratio	biomcers	197.77	71.60	204
CER_sumorratio	as_sum	5.83	1.69	204

CER_sumorratio	c18_sum	0.35	0.39	204
CER_sumorratio	ns_sum	395.40	118.63	204
CER_sumorratio	nds_sum	32.64	17.17	204
CER_sumorratio	s18_sum	296.27	94.76	204
CER_sumorratio	s19_sum	71.22	27.32	204
CER_sumorratio	s20_sum	15.09	5.10	204
CER_sumorratio	ds18_sum	15.76	8.50	204
CER_sumorratio	n22_sum	9.19	4.52	204
CER_sumorratio	n23_sum	59.47	21.85	204
CER_sumorratio	n24_sum	304.20	97.89	204
CER_sumorratio	n25_sum	2.41	1.00	204
CER_sumorratio	n26_sum	35.43	10.18	204
CER_sumorratio	c18ratio	16.78	10.40	204
CER_sumorratio	c18dsratio	0.08	0.05	198
CER_sumorratio	n22ratio	9.54	3.75	204
CER_sumorratio	n24ratio	18.02	8.31	204
CER_sumorratio	n24s19ratio	23.51	12.85	204
CER_sumorratio	n24s20ratio	8.35	3.52	204
CER_sumorratio	n26ratio	45.52	19.16	204
CER_sumorratio	c18s1psratio	3.78	3.82	198
CER_sumorratio	dssumc18dsratio	8727.47	6724.54	204
CER_sumorratio	c18snsratio	0.0002	0.0001	204
Eico	11,12-DHET	135.41	87.46	204
Eico	11-HETE	47.47	27.03	204
Eico	12,13-EpOME	2242.46	2175.75	184
Eico	12,13-DiHOME	3595.88	2872.39	204
Eico	12-HETE	116.54	81.56	204
Eico	13-HODE	7316.95	6630.78	204
Eico	13-HOTrE	393.54	354.11	186
Eico	13-OxoODE	488.20	304.85	204
Eico	14,15-DHET	137.63	83.70	204
Eico	15-HETE	110.62	92.51	192
Eico	19,20-DiHDPA	846.07	429.10	203
Eico	4-HDHA	130.74	91.98	201
Eico	5-HETE	131.45	133.96	189
Eico	9,10-EpOME	642.53	706.08	183
Eico	9,10-DiHOME	2995.49	2856.31	204
Eico	9-HODE	4053.73	3468.59	204
Eico	9-HOTrE	258.62	268.45	175
Eico	9-OxoODE	761.85	469.79	204
Eico	TransEKODE	139.45	107.06	203
Eico_sumorratio	SumEicos	24164.22	15458.22	204
Eico_sumorratio	LA	21949.85	14938.87	204
Eico_sumorratio	AA	662.96	385.90	204

Eico_sumorratio	DHA	973.23	451.56	203
Eico_sumorratio	aLA	580.67	590.52	204
Eico_sumorratio	omega3	1551.41	847.19	204
Eico_sumorratio	omega6	22612.81	14979.80	204
Eico_sumorratio	epdi9	21.80	179.65	183
Eico_sumorratio	epdi13	9.01	65.73	184
Eico_sumorratio	oxho9	0.28	0.32	204
Eico_sumorratio	oxho13	0.09	0.08	204
Eico_sumorratio	lox15	8268.07	7006.40	204
Eico_sumorratio	LOX1	5657.50	3816.52	204
Eico_sumorratio	cyp450	9463.41	6363.78	204
Eico_sumorratio	EPHX2	7706.34	5116.09	204
Eico_sumorratio	ALOX5	5288.04	3924.49	204
NAE	AEA	573.18	276.00	204
NAE	DHEA	491.41	207.39	204
NAE	DPEA	29.51	11.11	204
NAE	HEA	26.63	11.64	204
NAE	LEA	1035.53	412.76	204
NAE	OEA	1283.39	575.47	204
NAE	PEA	2888.98	1027.62	204
NAE	POEA	43.39	45.80	204
NAE	PDEA	39.81	17.57	204
NAE	STEA	1039.60	458.74	204
NAE	VEA	600.42	325.53	204
NAE_sumorratio	sumNEA	8086.68	2852.06	204

Table 0.4: Predictors identified for each lipid species in 204 samples

Depicted are the predictors (Pr) identified by stepwise multiple linear regression, the coefficient of each predictor (C), and the P-value (P). The class (Cl) of each species is as follows; N, NAE; C, CER; E, Eico. The predictors are as follows; a, age at enrolment; a2, age^2; qc, quality control sample measure specific to each species; B, the mass spectrometry batch; bp, hypertension status; i, trait for sample abnormality; c, cholesterol; s, sex; b, BMI. To fit the table on the page, coefficients are depicted as whole number and P-values are summarised to the nearest two decimal places. sm, summation trait; ro, ratio trait.

Lipid	Cl	adjR2	Pr	C	P	Pr	C	P	Pr	C	P	Pr	C	P	Pr	C	P	Pr	C	P	Pr	C	P	Pr	C	P	Pr	C	P	Pr	C	P
DHEA	N	0.10	a2	0	0.00	c	28	0.05																									
AEA	N	0.08	B	-34	0.00	c	54	0.00																									
DPEA	N	0.07	c	3	0.00																												
HEA	N	0.06	c	3	0.00																												
OEA	N	0.11	qc	0	0.00	c	84	0.02																									
PEA	N	0.11	B	122	0.01	c	282	0.00																									
STEA	N	0.11	c	139	0.00																												
LEA	N	0.07	qc	1	0.00	c	69	0.01																									
VEA	N	0.05	qc	0	0.01	s	63	0.16	c	36	0.08																						
smEA	N	0.08	c	743	0.00																												
POEA	N	0.02	s	13	0.04																												
PDEA	N	0.30	qc	1	0.00	c	3	0.00																									
AA	E	0.05	qc	1	0.00																												
aLA	E	0.02	qc	0	0.27	B	56	0.16	i	-49	0.34	c	55	0.20	bp	39	0.59	a	-6	0.06												
Lox1	E	0.02	qc	1	0.04																												
cyp450	E	0.01	a2	0	0.11																												
DHA	E	0.01	b	-10	0.08																												
DHET1112	E	0.10	qc	0	0.00	B	-19	0.00																									
DHET1415	E	0.02	B	-8	0.17	s	-15	0.21	a2	0	0.13	b	-2	0.16																			
DiHDPA1920	E	0.00	qc	0	0.24	B	-9	0.76																									
DiHOME910	E	0.10	B	-886	0.00																												
DiHOME1213	E	0.03	qc	1	0.08	a	-21	0.09	b	-63	0.08																						
epdi9	E	0.03	B	-35	0.01																												

epdi13	E	0.02	B	-12	0.02																								
EpOME910	E	0.12	qc	1	0.00	B	230	0.00																					
EpOME1213	E	0.01	qc	1	0.08	B	221	0.24	bp	-253	0.37	b	34	0.22															
HDHA4	E	0.02	qc	0	0.15	a2	0	0.21	c	7	0.26	b	-2	0.15															
HETE5	E	0.00	bp	22	0.18																								
HETE11	E	0.01	c	3	0.11																								
HETE12	E	0.02	B	-8	0.17	a2	0	0.06	i	8	0.27	a	4	0.09	b	-1	0.36	s	13	0.25									
HETE15	E	0.02	qc	0	0.07	b	2	0.13																					
HODE9	E	0.01	qc	1	0.05	B	459	0.11	i	-117	0.70																		
HODE13	E	0.01	s	-1677	0.07																								
HOTrE9	E	0.05	B	61	0.00																								
HOTrE13	E	0.01	a2	0	0.08																								
LA	E	0.00	s	-2554	0.22																								
lox5	E	0.02	qc	1	0.02																								
lox15	E	0.01	s	-1739	0.08																								
omega3	E	0.00	a	-4	0.24																								
omega6	E	0.00	s	-2542	0.23																								
oxho9	E	0.06	B	0	0.00																								
oxho13	E	0.05	B	0	0.00																								
OxoODE9	E	0.13	qc	1	0.00	B	-85	0.00																					
OxoODE13	E	0.03	qc	1	0.02	c	36	0.06																					
EPHX2	E	0.07	qc	0	0.06	B	-664	0.08	a	-38	0.08	b	-116	0.07															
SumEicos	E	0.00	qc	1	0.25	B	1006	0.52																					
TransEKODE	E	0.11	qc	0	0.00	c	18	0.01																					
A22S18	C	0.04	B	0	0.00																								
A24S18	C	0.14	qc	1	0.00																								

A26S18	C	0.22	qc	0	0.72	B	0	0.00	i	0	0.86	s	0	0.03	bp	0	0.92	a	0	0.00									
As_sum	C	0.11	qc	2	0.00																								
biomcers	C	0.07	B	-11	0.00	a	-1	0.04																					
C18DS	C	0.28	qc	0	0.00	B	0	0.00	i	0	0.21	a2	0	0.54	a	0	0.43	s	0	0.77	c	0	0.00	b	0	0.45	bp	0	0.88
C18S	C	0.49	qc	0	<2e-16	B	0	0.09	a2	0	0.13	b	0	0.19	c	0	0.11												
C18S1P	C	0.43	qc	2	<2e-16	b	0	0.06																					
N16S18	C	0.17	qc	0	0.06	B	0	0.04	s	0	0.07	a	0	0.19	a2	0	0.47	b	0	0.04	c	0	0.11						
N20S18	C	0.17	qc	0	0.00	B	0	0.00	c	0	0.14																		
N22DS18	C	0.09	B	0	0.00																								
N22S18	C	0.11	B	-1	0.00																								
N22S19	C	0.09	B	0	0.00	c	0	0.08																					
N23S18	C	0.02	B	-2	0.03																								
N23S20	C	0.07	a2	0	0.00																								
N24DS18	C	0.16	B	-1	0.00	c	2	0.00																					
N24DS19	C	0.19	qc	1	0.00	c	1	0.00																					
N24DS20	C	0.13	B	0	0.00	b	0	0.02	c	0	0.00																		
N24S16	C	0.17	B	0	0.00																								
N24S17	C	0.02	B	-1	0.02																								
N24S18	C	0.07	B	-11	0.00	a	-1	0.03																					
N24S19	C	0.05	B	-4	0.00																								
N24S20	C	0.09	s	-2	0.00	c	1	0.00																					
N24S22	C	0.01	qc	0	0.24	s	0	0.30	a2	0	0.19																		
N25DS18	C	0.13	c	0	0.00																								
N25S20	C	0.14	B	0	0.03	a	0	0.00																					
N26DS18	C	0.14	B	0	0.00	b	0	0.02	c	0	0.00																		
N26S18	C	0.22	qc	0	0.00	c	2	0.00																					

N26S19	C	0.04	qc	0	0.19	B	0	0.35																					
N27S18	C	0.15	qc	0	0.08	B	0	0.01	s	1	0.00	a	0	0.03	a2	0	0.08												
N28S18	C	0.15	qc	0	0.00	s	0	0.00	c	0	0.00																		
N29S18	C	0.07	s	1	0.00																								
c18sm	C	0.43	qc	1	<2e-16																								
c18dsro	C	0.56	qc	1	< 2e-16	B	0	0.00																					
c18ro	C	0.38	qc	1	< 2e-16																								
c18s1psro	C	0.15	qc	2	0.00	i	0	0.62	s	0	0.75	bp	0	0.54	a	0	0.11	a2	0	0.15	b	0	0.11	c	-1	0.01			
c18snsro	C	0.27	qc	0	0.00	a2	0	0.01	b	0	0.01																		
ds18sm	C	0.17	B	-2	0.00	c	2	0.00																					
dssmc18dsro	C	0.12	B	-1539	0.00																								
n22sm	C	0.12	B	-1	0.00																								
n22ro	C	0.09	qc	2	0.02	B	0	0.15	i	0	0.37	s	0	0.96	a	0	0.07	a2	0	0.07	b	0	0.14	c	-1	0.03			
n23sm	C	0.02	B	-2	0.03																								
n24sm	C	0.07	B	-17	0.00																								
n24ro	C	0.15	B	1	0.00	b	0	0.01	c	-2	0.00																		
n24s19ro	C	0.20	qc	0	0.05	B	1	0.01	i	-1	0.46	s	-1	0.61	bp	1	0.59	a	0	0.35	a2	0	0.37	b	0	0.12	c	-4	0.00
n24s20ro	C	0.15	qc	1	0.00	s	-1	0.02	c	-1	0.00																		
n25sm	C	0.20	B	0	0.00	a	0	0.00	c	0	0.00																		
n26sm	C	0.12	qc	0	0.00	B	-1	0.08																					
n26ro	C	0.11	B	3	0.00	bp	-3	0.23	b	-1	0.02	c	-3	0.02															
ndssm	C	0.20	B	-4	0.00	c	5	0.00																					
nssm	C	0.07	B	-20	0.00																								
ro16to24	C	0.11	i	0	0.38	s	0	0.06	a	0	0.03	b	0	0.03	c	0	0.04												
ro20to24	C	0.19	B	0	0.00	a2	0	0.00	c	0	0.00																		
ro22to24	C	0.11	B	0	0.00	a2	0	0.03	c	0	0.04																		

s18sm	C	0.06	B	-15	0.00																							
s19sm	C	0.05	qc	1	0.00	s	-5	0.15	c	2	0.26																	
s20sm	C	0.11	B	0	0.12	s	-1	0.03	a2	0	0.07	a	0	0.13	b	0	0.03	c	1	0.02								
totalsphingo	C	0.07	B	-23	0.00																							

Table 0.5: Genome-wide association study results of Eico and related species and traits of the range finding study

The table depicts the top 20 associations with all of the related Eico traits. Chr, chromosome; SE, standard error; Freq, minor allele frequency.

Lipid	Chr	SNP	Position	A1	A2	Freq	beta	SE	P-value	Gene Name	Consequence
AA	6	rs2503831	5426895	C	G	0.301546	0.556036	0.101775	4.67E-08	FARS2	intron variant
AA	6	rs11243010	5396596	T	G	0.298969	0.561488	0.102973	4.96E-08	FARS2	intron variant
AA	6	rs28372293	5405982	T	C	0.298969	0.561488	0.102973	4.96E-08	FARS2	intron variant
AA	6	rs2432756	5422123	T	C	0.298969	0.561488	0.102973	4.96E-08	FARS2	intron variant
AA	6	rs2432757	5425949	T	A	0.304124	0.541329	0.101793	1.05E-07	FARS2	intron variant
AA	6	rs2503832	5427461	G	C	0.304124	0.541329	0.101793	1.05E-07	FARS2	intron variant
AA	6	rs6597128	5395808	G	A	0.301546	0.546354	0.102984	1.13E-07	FARS2	intron variant
AA	6	rs12200334	5397070	A	G	0.301546	0.546354	0.102984	1.13E-07	FARS2	intron variant
AA	6	rs7760135	5400159	G	T	0.301546	0.546354	0.102984	1.13E-07	FARS2	intron variant
AA	6	rs12664305	5407166	C	G	0.301546	0.546354	0.102984	1.13E-07	FARS2	intron variant
AA	6	rs11759187	5409412	A	G	0.301546	0.546354	0.102984	1.13E-07	FARS2	intron variant
AA	6	rs11751802	5410152	T	C	0.301546	0.546354	0.102984	1.13E-07	FARS2	intron variant
AA	6	rs72815696	5413959	T	G	0.301546	0.546354	0.102984	1.13E-07	FARS2	intron variant
AA	6	rs12208777	5415873	G	A	0.301546	0.546354	0.102984	1.13E-07	FARS2	intron variant
AA	7	rs17167945	14216442	G	A	0.188144	0.591945	0.113449	1.81E-07	DGKB	intron variant
AA	6	rs9392080	5434995	C	T	0.311856	0.528743	0.101892	2.11E-07	FARS2	intron variant
AA	6	rs6900391	5395873	T	C	0.309278	0.519125	0.102108	3.69E-07	FARS2	intron variant
AA	6	rs4959338	5400658	A	G	0.309278	0.519125	0.102108	3.69E-07	FARS2	intron variant
AA	7	rs57625337	14215781	T	C	0.21134	0.557622	0.111267	5.40E-07	DGKB	intron variant
AA	6	rs9378947	5425772	A	C	0.322165	0.484614	0.100704	1.49E-06	FARS2	intron variant

DHA	1	rs17370861	172691322	A	C	0.417949	0.459707	0.100889	5.20E-06	Intergenic	Intergenic
DHA	2	rs13428826	28344999	G	C	0.184615	0.558016	0.123982	6.77E-06	BABAM2	intron variant
DHA	2	rs3845821	66053473	C	T	0.0564103	0.983271	0.219927	7.79E-06	AC007389.1	intron variant
DHA	19	rs8106931	18506666	G	A	0.367188	0.454017	0.101601	7.87E-06	LRRC25	intron variant
DHA	14	rs2038423	29029638	G	A	0.170157	0.56801	0.12883	1.04E-05	Intergenic	Intergenic
DHA	3	rs3902481	168684584	C	G	0.0615385	0.891086	0.202394	1.07E-05	Intergenic	Intergenic
DHA	3	rs56105327	168688578	A	G	0.0615385	0.891086	0.202394	1.07E-05	Intergenic	Intergenic
DHA	2	rs1546029	28542934	G	A	0.220513	0.506104	0.116071	1.30E-05	BABAM2	intron variant
DHA	14	rs1950797	29030709	A	C	0.172775	0.550416	0.128476	1.83E-05	Intergenic	Intergenic
DHA	14	rs12435268	29028946	T	G	0.164062	0.554091	0.129459	1.87E-05	Intergenic	Intergenic
DHA	9	rs76767633	81496978	G	A	0.091623	0.748095	0.175684	2.06E-05	Intergenic	Intergenic
DHA	1	rs163771	229469350	A	G	0.438462	-0.42427	0.0996618	2.07E-05	CCSAP	intron variant
DHA	1	rs12731355	172696875	T	C	0.402564	0.4372	0.103227	2.28E-05	Intergenic	Intergenic
DHA	1	rs34789243	172704471	A	T	0.402564	0.4372	0.103227	2.28E-05	Intergenic	Intergenic
DHA	20	rs2423999	16722776	T	C	0.112821	0.666726	0.157469	2.30E-05	SNRPB2	3_prime_UTR_variant
DHA	9	rs11523754	81473400	G	A	0.117949	0.639993	0.151467	2.39E-05	Intergenic	Intergenic
DHA	9	rs11137878	81475750	C	G	0.117949	0.639993	0.151467	2.39E-05	Intergenic	Intergenic
DHA	7	rs73276678	24389872	C	T	0.122396	0.651761	0.154606	2.49E-05	AC003044.1	intron variant
DHA	1	rs35319454	172765385	C	A	0.397436	0.431314	0.102535	2.59E-05	AL031599.1	intron variant
DHA	1	rs36060149	172769038	G	A	0.397436	0.431314	0.102535	2.59E-05	AL031599.1	intron variant
DHET1112	4	rs6551813	65154127	G	A	0.0518135	1.1172	0.206596	6.38E-08	TECRL	intron variant
DHET1112	4	rs11930994	65164696	T	C	0.0518135	1.1172	0.206596	6.38E-08	TECRL	intron variant
DHET1112	4	rs11935568	65165661	G	A	0.0518135	1.1172	0.206596	6.38E-08	TECRL	intron variant
DHET1112	4	rs1026925	65167853	C	G	0.0518135	1.1172	0.206596	6.38E-08	TECRL	intron variant

DHET1112	4	rs13104958	65176146	T	C	0.0518135	1.1172	0.206596	6.38E-08	TECRL	intron variant
DHET1112	4	rs1545868	65184724	G	A	0.0518135	1.1172	0.206596	6.38E-08	TECRL	intron variant
DHET1112	4	rs434495	65199924	G	A	0.0518135	1.1172	0.206596	6.38E-08	TECRL	intron variant
DHET1112	4	rs6824140	65212152	C	T	0.0518135	1.1172	0.206596	6.38E-08	TECRL	intron variant
DHET1112	4	rs1552211	65217836	G	T	0.0518135	1.1172	0.206596	6.38E-08	TECRL	intron variant
DHET1112	4	rs6835369	65218923	C	T	0.0518135	1.1172	0.206596	6.38E-08	TECRL	intron variant
DHET1112	13	rs1771422	98396333	T	C	0.261658	0.531482	0.104468	3.63E-07	Intergenic	Intergenic
DHET1112	13	rs1772376	98396479	A	G	0.261658	0.531482	0.104468	3.63E-07	Intergenic	Intergenic
DHET1112	13	rs1771421	98396932	G	C	0.261658	0.531482	0.104468	3.63E-07	Intergenic	Intergenic
DHET1112	13	rs1772375	98396952	T	C	0.261658	0.531482	0.104468	3.63E-07	Intergenic	Intergenic
DHET1112	13	rs1620765	98399138	A	G	0.261658	0.531482	0.104468	3.63E-07	Intergenic	Intergenic
DHET1112	13	rs1617544	98399489	A	G	0.261658	0.531482	0.104468	3.63E-07	Intergenic	Intergenic
DHET1112	13	rs2793701	98400606	A	C	0.261658	0.531482	0.104468	3.63E-07	Intergenic	Intergenic
DHET1112	4	rs72669960	106509644	G	C	0.0906736	0.830131	0.168058	7.83E-07	ARHGEF38	intron variant
DHET1112	4	rs17324954	106511488	T	G	0.0906736	0.830131	0.168058	7.83E-07	ARHGEF38	intron variant
DHET1112	13	rs2153592	98405806	T	A	0.310881	0.494196	0.101066	1.01E-06	Intergenic	Intergenic
DHET1415	10	rs113862732	80633954	C	G	0.0923077	0.980474	0.201861	1.19E-06	Intergenic	Intergenic
DHET1415	10	rs80100020	80638388	T	C	0.0923077	0.980474	0.201861	1.19E-06	Intergenic	Intergenic
DHET1415	10	rs75583478	80642070	T	G	0.0923077	0.980474	0.201861	1.19E-06	Intergenic	Intergenic
DHET1415	10	rs72474827	80645266	T	C	0.0923077	0.980474	0.201861	1.19E-06	Intergenic	Intergenic
DHET1415	10	rs74235034	80645472	T	A	0.0923077	0.980474	0.201861	1.19E-06	Intergenic	Intergenic
DHET1415	10	rs7078982	80646221	T	G	0.0871795	1.00035	0.206434	1.26E-06	Intergenic	Intergenic
DHET1415	10	rs77212942	80649622	T	C	0.0871795	1.00035	0.206434	1.26E-06	Intergenic	Intergenic
DHET1415	10	rs80297085	80650234	C	T	0.0871795	1.00035	0.206434	1.26E-06	Intergenic	Intergenic

DHET1415	10	rs75584054	80651631	G	A	0.0871795	1.00035	0.206434	1.26E-06	Intergenic	Intergenic
DHET1415	10	rs74141421	80653139	A	G	0.0871795	1.00035	0.206434	1.26E-06	Intergenic	Intergenic
DHET1415	10	rs79419066	80653298	T	C	0.0871795	1.00035	0.206434	1.26E-06	Intergenic	Intergenic
DHET1415	10	rs77116077	80653364	G	A	0.0871795	1.00035	0.206434	1.26E-06	Intergenic	Intergenic
DHET1415	20	rs435954	46967667	A	C	0.074359	0.990055	0.208121	1.96E-06	AL121888.1	intron variant
DHET1415	9	rs2291681	14790841	T	G	0.471795	-0.49392	0.105175	2.65E-06	FREM1	intron variant/
DHET1415	14	rs73331306	43828825	A	G	0.0538462	1.15937	0.248258	3.01E-06	Intergenic	Intergenic
DHET1415	9	rs7859055	94235956	A	C	0.151282	0.622091	0.138386	6.95E-06	Intergenic	Intergenic
DHET1415	8	rs4922259	16947784	G	C	0.312821	0.524927	0.117361	7.72E-06	MICU3	intron variant
DHET1415	6	rs117820567	170408983	A	G	0.0666667	0.909046	0.203999	8.35E-06	AL603783.1	non_coding_transcript
DHET1415	10	rs61195044	85578928	T	C	0.0974359	0.745734	0.168013	9.06E-06	Intergenic	Intergenic
DHET1415	12	rs74697053	98521263	T	G	0.135897	0.705427	0.159116	9.28E-06	AC016152.1	intron variant
DiHDPA1920	10	rs2940724	50139444	A	G	0.294271	0.516008	0.107194	1.48E-06	WDFY4	intron variant
DiHDPA1920	10	rs2663056	50125972	T	C	0.309278	0.495369	0.105092	2.43E-06	WDFY4	intron variant
DiHDPA1920	10	rs2663047	50128386	A	G	0.309278	0.495369	0.105092	2.43E-06	WDFY4	intron variant
DiHDPA1920	1	rs163771	229469350	A	G	0.440722	-0.45725	0.0985336	3.47E-06	CCSAP	intron variant
DiHDPA1920	11	rs123381	2835780	A	G	0.28866	0.494504	0.109233	5.98E-06	KCNQ1	intron variant
DiHDPA1920	10	rs10857657	50125133	C	T	0.31701	0.464659	0.103159	6.66E-06	WDFY4	intron variant
DiHDPA1920	16	rs60660286	31275374	T	G	0.298429	0.479284	0.108306	9.63E-06	ITGAM	intron variant
DiHDPA1920	16	rs11150616	31370090	T	G	0.295812	0.478406	0.108163	9.73E-06	ITGAX	intron variant
DiHDPA1920	16	rs11574639	31372081	G	A	0.295812	0.478406	0.108163	9.73E-06	ITGAX	intron variant
DiHDPA1920	16	rs7203472	31376105	A	G	0.295812	0.478406	0.108163	9.73E-06	ITGAX	intron variant
DiHDPA1920	16	rs62048736	84974267	C	T	0.239691	0.477822	0.109004	1.17E-05	LINC02176	downstream_gene_variant
DiHDPA1920	8	rs117293618	75591073	G	A	0.128272	0.601504	0.139265	1.57E-05	MIR2052HG	intron variant

DiHDPA1920	17	rs11655688	33664818	A	G	0.0670103	0.795366	0.185463	1.80E-05	Intergenic	Intergenic
DiHDPA1920	11	rs1690595	97044431	C	T	0.434555	-0.441505	0.103051	1.83E-05	Intergenic	Intergenic
DiHDPA1920	9	rs76767633	81496978	G	A	0.0894737	0.743265	0.17383	1.90E-05	Intergenic	Intergenic
DiHDPA1920	2	rs1546029	28542934	G	A	0.219072	0.486077	0.11385	1.96E-05	BABAM2	intron variant
DiHDPA1920	3	rs13092091	192833898	A	G	0.0684211	0.819444	0.192111	1.99E-05	Intergenic	Intergenic
DiHDPA1920	15	rs62017613	94762444	G	A	0.225131	-0.496566	0.117436	2.35E-05	Intergenic	Intergenic
DiHDPA1920	13	rs9585722	102173948	G	C	0.298969	0.440024	0.104605	2.59E-05	ITGBL1	intron variant
DiHDPA1920	15	rs1962947	94754869	G	C	0.239691	-0.47857	0.114542	2.94E-05	Intergenic	Intergenic
DiHOME1213	6	rs13207194	138354242	T	G	0.0572917	0.73138	0.16004	4.88E-06	Intergenic	Intergenic
DiHOME1213	4	rs12511352	76993900	T	C	0.0538462	0.703784	0.155105	5.69E-06	ART3	intron variant
DiHOME1213	19	rs8106327	55029944	T	C	0.0589744	0.670019	0.149618	7.53E-06	AC245036.1	upstream_gene_variant
DiHOME1213	19	rs6509883	55030376	G	A	0.0589744	0.670019	0.149618	7.53E-06	AC245036.1	upstream_gene_variant
DiHOME1213	19	rs112382747	55030736	A	G	0.0589744	0.670019	0.149618	7.53E-06	AC245036.1	upstream_gene_variant
DiHOME1213	19	rs10420116	55030897	C	A	0.0589744	0.670019	0.149618	7.53E-06	AC245036.1	upstream_gene_variant
DiHOME1213	19	rs10421435	55031501	G	A	0.0589744	0.670019	0.149618	7.53E-06	AC245036.1	upstream_gene_variant
DiHOME1213	19	rs10422190	55031519	C	G	0.0589744	0.670019	0.149618	7.53E-06	AC245036.1	upstream_gene_variant
DiHOME1213	19	rs2885683	55032548	C	G	0.0589744	0.670019	0.149618	7.53E-06	AC245036.1	upstream_gene_variant
DiHOME1213	19	rs11084349	55042124	T	G	0.0589744	0.670019	0.149618	7.53E-06	KIR3DX1	upstream_gene_variant
DiHOME1213	1	rs2310920	9414755	A	G	0.184615	0.4212	0.094885	9.04E-06	SPSB1	intron variant
DiHOME1213	1	rs733893	9415009	G	A	0.184615	0.4212	0.094885	9.04E-06	SPSB1	intron variant
DiHOME1213	1	rs3765960	9415592	A	G	0.184615	0.4212	0.094885	9.04E-06	SPSB1	intron variant
DiHOME1213	1	rs2310919	9415727	A	G	0.184615	0.4212	0.094885	9.04E-06	SPSB1	intron variant
DiHOME1213	1	rs9435243	9416151	A	G	0.184615	0.4212	0.094885	9.04E-06	SPSB1	synonymous_variant
DiHOME1213	1	rs552230	9422875	A	C	0.184615	0.4212	0.094885	9.04E-06	SPSB1	intron variant

DiHOME1213	1	rs2142575	9424208	T	C	0.184615	0.4212	0.094885	9.04E-06	SPSB1	intron variant
DiHOME1213	1	rs2142574	9424234	T	C	0.184615	0.4212	0.094885	9.04E-06	SPSB1	intron variant
DiHOME1213	1	rs938250	15339092	A	G	0.0794872	0.603604	0.138333	1.28E-05	KAZN	intron variant
DiHOME1213	1	rs6663935	15340867	G	T	0.0794872	0.603604	0.138333	1.28E-05	KAZN	intron variant
DiHOME910	10	rs11813343	68582923	A	G	0.0572917	0.890699	0.18769	2.08E-06	CTNNA3	intron variant
DiHOME910	10	rs75172124	68588163	T	C	0.0572917	0.890699	0.18769	2.08E-06	CTNNA3	intron variant
DiHOME910	10	rs115787169	68588379	A	G	0.0572917	0.890699	0.18769	2.08E-06	CTNNA3	intron variant
DiHOME910	10	rs139950600	68589346	C	G	0.0572917	0.890699	0.18769	2.08E-06	CTNNA3	intron variant
DiHOME910	14	rs4906435	104765052	C	T	0.489474	0.378146	0.0864761	1.23E-05	Intergenic	Intergenic
DiHOME910	6	rs36000519	75429720	G	A	0.106383	0.577314	0.133583	1.55E-05	AL356277.3	intron variant
DiHOME910	6	rs2502538	75421693	T	A	0.111979	0.562399	0.130139	1.55E-05	AL356277.3	intron variant
DiHOME910	16	rs154545	22743803	C	T	0.148438	0.544952	0.126207	1.58E-05	Intergenic	Intergenic
DiHOME910	3	rs76732363	24271501	A	G	0.0572917	0.788104	0.182817	1.63E-05	THRB	intron variant
DiHOME910	3	rs10510541	24274241	C	G	0.0572917	0.788104	0.182817	1.63E-05	THRB	intron variant
DiHOME910	3	rs78592786	24276745	T	C	0.0572917	0.788104	0.182817	1.63E-05	THRB	intron variant
DiHOME910	4	rs4274920	117624139	C	T	0.338542	0.371594	0.0862002	1.63E-05	Intergenic	Intergenic
DiHOME910	4	rs6533995	117626529	C	T	0.338542	0.371594	0.0862002	1.63E-05	Intergenic	Intergenic
DiHOME910	4	rs4425450	117632065	A	T	0.338542	0.371594	0.0862002	1.63E-05	AC093765.5	downstream_gene_variant
DiHOME910	2	rs949516	54548161	T	C	0.0677083	0.693168	0.160853	1.64E-05	Intergenic	Intergenic
DiHOME910	2	rs138682285	54549942	T	C	0.0677083	0.693168	0.160853	1.64E-05	Intergenic	Intergenic
DiHOME910	2	rs3769504	106477092	A	G	0.455729	0.369643	0.0864474	1.90E-05	NCK2	intron variant
DiHOME910	10	rs10906716	14655487	C	T	0.270833	0.409921	0.0964409	2.13E-05	FAM107B	intron variant
DiHOME910	17	rs8078981	51473669	A	C	0.0520833	0.828619	0.195164	2.18E-05	AC034268.2	intron variant
DiHOME910	17	rs12325912	35153830	G	T	0.169271	-0.482827	0.113819	2.21E-05	Intergenic	Intergenic

EpOME1213	9	rs2795361	100051165	G	A	0.091954	0.833523	0.155389	8.14E-08	SUGT1P4-STRA6LP	intron variant
EpOME1213	9	rs4567164	100206602	C	G	0.100575	0.783325	0.150205	1.84E-07	TDRD7	intron variant
EpOME1213	9	rs7047018	100215707	T	C	0.100575	0.783325	0.150205	1.84E-07	TDRD7	intron variant
EpOME1213	9	rs1993771	100148743	A	G	0.0977011	0.76236	0.151502	4.85E-07	AL512590.1	intron variant
EpOME1213	9	rs1037949	100157929	G	A	0.0977011	0.76236	0.151502	4.85E-07	AL512590.1	intron variant
EpOME1213	19	rs143858135	6486433	A	C	0.114035	0.78109	0.156077	5.60E-07	DENND1C	upstream_gene_variant
EpOME1213	19	rs60849443	6486679	A	C	0.114035	0.78109	0.156077	5.60E-07	DENND1C	upstream_gene_variant
EpOME1213	19	rs72985394	6488531	C	G	0.114035	0.78109	0.156077	5.60E-07	Intergenic	Intergenic
EpOME1213	7	rs7776819	29130762	G	A	0.0574713	0.967617	0.194193	6.27E-07	CPVL	intron variant
EpOME1213	9	rs4557815	100211869	T	C	0.12931	0.674284	0.136801	8.27E-07	TDRD7	intron variant
EpOME1213	9	rs7034581	100212403	C	G	0.12931	0.674284	0.136801	8.27E-07	TDRD7	intron variant
EpOME1213	9	rs6415827	100244171	G	A	0.12931	0.674284	0.136801	8.27E-07	TDRD7	intron variant
EpOME1213	4	rs28637562	182970980	C	A	0.117816	0.67156	0.14348	2.86E-06	TENM3-AS1	intron variant
EpOME1213	9	rs10739397	100178253	C	T	0.135057	0.629523	0.134645	2.93E-06	TDRD7	intron variant
EpOME1213	10	rs118107781	108587050	C	A	0.066092	0.857753	0.183737	3.04E-06	SORCS1	intron variant
EpOME1213	9	rs73397915	12591959	G	A	0.0574713	0.927729	0.200112	3.55E-06	Intergenic	Intergenic
EpOME1213	9	rs73397925	12597326	A	G	0.0574713	0.927729	0.200112	3.55E-06	Intergenic	Intergenic
EpOME1213	9	rs73397931	12602250	G	T	0.0574713	0.927729	0.200112	3.55E-06	Intergenic	Intergenic
EpOME1213	9	rs12005291	12602492	C	A	0.0574713	0.927729	0.200112	3.55E-06	Intergenic	Intergenic
EpOME1213	9	rs73397938	12603965	C	T	0.0574713	0.927729	0.200112	3.55E-06	Intergenic	Intergenic
EpOME910	13	rs1014053	104039222	A	G	0.0635838	0.887975	0.197588	6.99E-06	AL162717.1	intron variant
EpOME910	13	rs1360348	104040712	A	G	0.0635838	0.887975	0.197588	6.99E-06	AL162717.1	intron variant
EpOME910	13	rs55979101	104043015	A	G	0.0635838	0.887975	0.197588	6.99E-06	AL162717.1	intron variant
EpOME910	13	rs56317893	104043031	C	G	0.0635838	0.887975	0.197588	6.99E-06	AL162717.1	intron variant

EpOME910	13	rs9586155	104043975	C	T	0.0635838	0.887975	0.197588	6.99E-06	AL162717.1	intron variant
EpOME910	5	rs4264941	121843893	C	G	0.0606936	0.93387	0.208493	7.49E-06	Intergenic	Intergenic
EpOME910	5	rs6892295	170839672	C	T	0.0578035	0.918982	0.207104	9.11E-06	NPM1	downstream_gene_variant
EpOME910	7	rs71526355	21170229	A	G	0.0982659	0.754684	0.171029	1.02E-05	Intergenic	Intergenic
EpOME910	7	rs11769988	36196379	T	A	0.112717	0.695394	0.158988	1.22E-05	EEPD1	intron variant
EpOME910	7	rs11760954	36198976	A	G	0.112717	0.695394	0.158988	1.22E-05	EEPD1	intron variant
EpOME910	7	rs6974287	36207473	A	G	0.112717	0.695394	0.158988	1.22E-05	EEPD1	intron variant
EpOME910	11	rs10790804	126314113	C	T	0.160819	0.608013	0.139628	1.33E-05	ST3GAL4	downstream_gene_variant
EpOME910	6	rs62388202	13068417	T	C	0.0982659	0.729836	0.16928	1.62E-05	PHACTR1	intron variant
EpOME910	7	rs34134022	36208965	C	T	0.153179	0.577802	0.134047	1.63E-05	EEPD1	intron variant
EpOME910	18	rs17661827	29311943	A	G	0.0982659	0.664988	0.154815	1.74E-05	Intergenic	Intergenic
EpOME910	10	rs4345892	1844321	C	T	0.0867052	-0.779942	0.181938	1.81E-05	Intergenic	Intergenic
EpOME910	10	rs115274337	1852461	T	C	0.0867052	-0.779942	0.181938	1.81E-05	Intergenic	Intergenic
EpOME910	10	rs80107446	1854672	C	T	0.0867052	-0.779942	0.181938	1.81E-05	Intergenic	Intergenic
EpOME910	10	rs7908003	1865499	T	C	0.0867052	-0.779942	0.181938	1.81E-05	Intergenic	Intergenic
EpOME910	10	rs75733593	1865702	G	A	0.0867052	-0.779942	0.181938	1.81E-05	Intergenic	Intergenic
HDHA4	12	rs11108140	95986103	G	A	0.115789	0.636022	0.124263	3.08E-07	PGAM1P5	intron variant
HDHA4	12	rs11108142	95988375	G	T	0.115789	0.636022	0.124263	3.08E-07	PGAM1P5	intron variant
HDHA4	12	rs117507866	95989981	A	G	0.0473684	1.04847	0.205191	3.23E-07	PGAM1P5	intron variant
HDHA4	12	rs7966493	95986715	T	C	0.110526	0.633735	0.126894	5.91E-07	PGAM1P5	intron variant
HDHA4	12	rs116963178	95992910	A	T	0.0526316	0.921044	0.192857	1.79E-06	PGAM1P5	intron variant
HDHA4	7	rs3095007	20256349	A	C	0.276316	-0.422751	0.0889845	2.03E-06	MACC1	intron variant
HDHA4	7	rs3095006	20256825	C	T	0.276316	-0.422751	0.0889845	2.03E-06	MACC1	intron variant
HDHA4	7	rs2190433	20257986	T	A	0.276316	-0.422751	0.0889845	2.03E-06	MACC1	upstream_gene_variant

HDHA4	7	rs3095005	20258454	C	T	0.276316	-0.422751	0.0889845	2.03E-06	MACC1	upstream_gene_variant
HDHA4	7	rs3095004	20266770	G	A	0.276316	-0.422751	0.0889845	2.03E-06	Intergenic	Intergenic
HDHA4	20	rs75411492	59768513	T	C	0.0894737	0.663782	0.142882	3.39E-06	Intergenic	Intergenic
HDHA4	20	rs6101253	59769932	T	C	0.0921053	0.646989	0.139512	3.53E-06	Intergenic	Intergenic
HDHA4	15	rs147962900	88453406	G	A	0.0510753	0.948424	0.208691	5.50E-06	NTRK3	intron variant
HDHA4	6	rs72821438	4745271	A	C	0.0789474	0.715428	0.158547	6.41E-06	CDYL	intron variant
HDHA4	15	rs72723442	33515721	C	T	0.0594595	0.749585	0.166594	6.81E-06	Intergenic	Intergenic
HDHA4	2	rs785265	191017739	A	G	0.428947	0.377552	0.0846106	8.11E-06	C2orf88	intron variant
HDHA4	5	rs28088	52586993	A	G	0.0921053	0.623947	0.141297	1.01E-05	Intergenic	Intergenic
HDHA4	12	rs79709382	118941597	A	G	0.0947368	0.681767	0.154471	1.02E-05	Intergenic	Intergenic
HDHA4	1	rs1691245	247112207	T	C	0.457895	-0.346051	0.0794527	1.33E-05	ZNF695	intron variant/
HDHA4	5	rs38057	52559840	T	G	0.0789474	0.648947	0.149053	1.34E-05	Intergenic	Intergenic
HETE11	6	rs6597229	6798199	G	C	0.296392	0.505609	0.0983642	2.74E-07	BTF3P7	upstream_gene_variant
HETE11	6	rs62392015	6793116	C	T	0.198454	0.588034	0.120892	1.15E-06	BTF3P7	downstream_gene_variant
HETE11	8	rs28509185	102747394	T	C	0.149485	0.581277	0.120419	1.39E-06	NCALD	intron variant
HETE11	2	rs10209255	48708346	G	T	0.21134	0.525952	0.109817	1.67E-06	PPP1R21	intron variant
HETE11	2	rs59356309	48709315	G	A	0.21134	0.525952	0.109817	1.67E-06	PPP1R21	intron variant
HETE11	2	rs72822226	48713729	C	T	0.21134	0.525952	0.109817	1.67E-06	PPP1R21	intron variant
HETE11	2	rs11125173	48715483	C	T	0.21134	0.525952	0.109817	1.67E-06	PPP1R21	intron variant
HETE11	7	rs117206014	2254428	T	C	0.0515464	1.02296	0.216091	2.20E-06	MAD1L1	intron variant
HETE11	2	rs72820441	48597799	C	G	0.208763	0.526781	0.111637	2.37E-06	FOXN2	intron variant
HETE11	2	rs10454134	48648026	A	G	0.190722	0.534761	0.114415	2.96E-06	Intergenic	Intergenic
HETE11	2	rs72820474	48649658	A	T	0.190722	0.534761	0.114415	2.96E-06	Intergenic	Intergenic
HETE11	1	rs72939513	83136096	A	G	0.0578947	0.958753	0.205811	3.19E-06	AL157944.1	intron variant

HETE11	10	rs4617506	50263543	A	G	0.21134	0.528822	0.114343	3.75E-06	VSTM4	intron variant
HETE11	10	rs7074818	50298521	G	A	0.21134	0.528822	0.114343	3.75E-06	VSTM4	intron variant
HETE11	10	rs4240498	50300179	A	C	0.21134	0.528822	0.114343	3.75E-06	VSTM4	intron variant
HETE11	10	rs6537494	50300393	T	C	0.21134	0.528822	0.114343	3.75E-06	VSTM4	intron variant
HETE11	2	rs6756596	48610367	C	T	0.219072	0.51151	0.110969	4.04E-06	FOXN2	downstream_gene_variant
HETE11	2	rs80018311	48612563	G	A	0.193299	0.520676	0.114082	5.02E-06	Intergenic	Intergenic
HETE11	2	rs4547570	48620587	T	C	0.193299	0.520676	0.114082	5.02E-06	Intergenic	Intergenic
HETE11	2	rs60825643	48623535	T	G	0.193299	0.520676	0.114082	5.02E-06	Intergenic	Intergenic
HETE12	8	rs11776328	23195616	A	G	0.139175	0.601083	0.131078	4.52E-06	LOXL2	intron variant
HETE12	7	rs10262428	80570142	G	T	0.345361	0.409656	0.0930191	1.06E-05	Intergenic	Intergenic
HETE12	8	rs75508008	108036684	A	G	0.0592784	0.871298	0.198758	1.17E-05	Intergenic	Intergenic
HETE12	1	rs1360997	102367411	C	T	0.0515464	0.947574	0.216804	1.24E-05	OLFM3	intron variant
HETE12	1	rs17125689	102369185	C	G	0.0515464	0.947574	0.216804	1.24E-05	OLFM3	intron variant
HETE12	1	rs17125705	102371076	A	G	0.0515464	0.947574	0.216804	1.24E-05	OLFM3	intron variant
HETE12	1	rs12564023	102373256	T	C	0.0515464	0.947574	0.216804	1.24E-05	OLFM3	intron variant
HETE12	1	rs1415106	102374372	A	G	0.0515464	0.947574	0.216804	1.24E-05	OLFM3	intron variant
HETE12	1	rs74107128	102377044	A	G	0.0515464	0.947574	0.216804	1.24E-05	OLFM3	intron variant
HETE12	1	rs7512170	102378410	T	C	0.0515464	0.947574	0.216804	1.24E-05	OLFM3	intron variant
HETE12	1	rs17125729	102379805	T	C	0.0515464	0.947574	0.216804	1.24E-05	OLFM3	intron variant
HETE12	14	rs733314	34238454	C	A	0.185567	0.514644	0.117973	1.29E-05	NPAS3	intron variant
HETE12	14	rs71419954	34242327	G	A	0.177835	0.512092	0.119181	1.73E-05	NPAS3	intron variant
HETE12	20	rs6064445	55333918	A	T	0.0798969	0.704968	0.165399	2.02E-05	Intergenic	Intergenic
HETE12	15	rs12439732	67233330	G	C	0.273438	-0.435261	0.10244	2.15E-05	LINC02206	intron variant
HETE12	13	rs9539133	61938965	G	A	0.474227	0.384987	0.0907723	2.22E-05	Intergenic	Intergenic

HETE12	13	rs7996824	61939303	A	G	0.474227	0.384987	0.0907723	2.22E-05	Intergenic	Intergenic
HETE12	13	rs2225356	61940033	G	C	0.474227	0.384987	0.0907723	2.22E-05	Intergenic	Intergenic
HETE12	13	rs78053781	61940687	C	A	0.474227	0.384987	0.0907723	2.22E-05	Intergenic	Intergenic
HETE12	13	rs9528324	61941825	G	A	0.474227	0.384987	0.0907723	2.22E-05	Intergenic	Intergenic
HETE15	16	rs316376	9739204	G	A	0.209497	0.632766	0.11409	2.92E-08	Intergenic	Intergenic
HETE15	12	rs77779256	76302218	C	T	0.0491803	1.17719	0.225811	1.86E-07	AC078923.1	intron variant
HETE15	2	rs7601885	214880133	A	T	0.0765027	0.947149	0.182007	1.95E-07	SPAG16	intron variant
HETE15	2	rs6749347	214880852	G	A	0.0765027	0.947149	0.182007	1.95E-07	SPAG16	intron variant
HETE15	2	rs6754414	214882577	G	A	0.0765027	0.947149	0.182007	1.95E-07	SPAG16	intron variant
HETE15	2	rs2055866	214885442	A	T	0.0765027	0.947149	0.182007	1.95E-07	SPAG16	intron variant
HETE15	2	rs2055867	214885653	C	T	0.0765027	0.947149	0.182007	1.95E-07	SPAG16	intron variant
HETE15	6	rs62404513	19293557	T	C	0.0519126	1.09455	0.219521	6.16E-07	AL357052.1	intron variant
HETE15	14	rs139526010	50966202	A	T	0.0655738	0.971457	0.197539	8.75E-07	MAP4K5	intron variant
HETE15	14	rs12050306	50981217	A	G	0.0655738	0.971457	0.197539	8.75E-07	MAP4K5	intron variant
HETE15	14	rs139763750	51049397	A	G	0.0655738	0.971457	0.197539	8.75E-07	ATL1	intron variant
HETE15	14	rs149315254	51123224	T	G	0.0655738	0.971457	0.197539	8.75E-07	SAV1	intron variant
HETE15	14	rs11627553	20740871	G	A	0.0546448	1.01382	0.213709	2.10E-06	TTC5	intron variant
HETE15	5	rs75930932	6332873	A	G	0.0956284	0.745722	0.158849	2.67E-06	LINC02145	downstream_gene_variant
HETE15	1	rs12728557	116017153	A	G	0.296703	0.506645	0.108337	2.92E-06	AL512638.1	intron variant
HETE15	1	rs12735736	116018544	A	C	0.296703	0.506645	0.108337	2.92E-06	AL512638.1	intron variant
HETE15	2	rs6744864	214853399	C	T	0.068306	0.882459	0.190663	3.69E-06	SPAG16	intron variant
HETE15	2	rs55950408	214856723	A	T	0.068306	0.882459	0.190663	3.69E-06	SPAG16	intron variant
HETE15	2	rs7589549	214860886	A	G	0.068306	0.882459	0.190663	3.69E-06	SPAG16	intron variant
HETE15	6	rs2503831	5426895	C	G	0.306011	0.507567	0.109681	3.70E-06	FARS2	intron variant

HETE5	6	rs77345935	50084112	A	C	0.0558659	0.914054	0.16688	4.32E-08	AL138826.1	intron variant
HETE5	20	rs73106770	24683047	G	C	0.0893855	0.679145	0.124123	4.46E-08	Intergenic	Intergenic
HETE5	20	rs11697449	24664960	A	G	0.0977654	0.629871	0.117722	8.77E-08	AL049594.1	downstream_gene_variant
HETE5	20	rs117310958	24669237	A	G	0.0977654	0.629871	0.117722	8.77E-08	Intergenic	Intergenic
HETE5	20	rs73106747	24672404	T	G	0.0977654	0.629871	0.117722	8.77E-08	Intergenic	Intergenic
HETE5	20	rs11699014	24672796	T	C	0.0977654	0.629871	0.117722	8.77E-08	Intergenic	Intergenic
HETE5	20	rs11907830	24673027	T	G	0.0977654	0.629871	0.117722	8.77E-08	Intergenic	Intergenic
HETE5	20	rs11698940	24679412	A	G	0.0977654	0.629871	0.117722	8.77E-08	Intergenic	Intergenic
HETE5	6	rs76525928	50068259	C	T	0.0670391	0.819398	0.155846	1.46E-07	AL138826.1	downstream_gene_variant
HETE5	6	rs2224886	50081061	A	G	0.0670391	0.819398	0.155846	1.46E-07	AL138826.1	intron variant
HETE5	6	rs1324525	50081669	C	A	0.0670391	0.819398	0.155846	1.46E-07	AL138826.1	intron variant
HETE5	6	rs1324526	50082003	G	A	0.0670391	0.819398	0.155846	1.46E-07	AL138826.1	intron variant
HETE5	6	rs7764125	50082102	G	A	0.0670391	0.819398	0.155846	1.46E-07	AL138826.1	intron variant
HETE5	6	rs6908223	50082140	T	C	0.0670391	0.819398	0.155846	1.46E-07	AL138826.1	intron variant
HETE5	6	rs2031364	50082442	T	A	0.0670391	0.819398	0.155846	1.46E-07	AL138826.1	intron variant
HETE5	6	rs76307249	50084170	C	G	0.0670391	0.819398	0.155846	1.46E-07	AL138826.1	intron variant
HETE5	6	rs78680513	50086141	C	T	0.0670391	0.819398	0.155846	1.46E-07	AL138826.1	intron variant
HETE5	6	rs6458721	50087752	A	C	0.0670391	0.819398	0.155846	1.46E-07	AL138826.1	intron variant
HETE5	20	rs112919658	24669361	G	A	0.106145	0.597622	0.114628	1.85E-07	Intergenic	Intergenic
HETE5	20	rs11697715	24670396	G	A	0.106145	0.597622	0.114628	1.85E-07	Intergenic	Intergenic
HODE13	10	rs72827118	113228770	A	G	0.0710526	0.682442	0.138125	7.78E-07	Intergenic	Intergenic
HODE13	10	rs72823394	113138073	C	T	0.0721925	0.672319	0.138279	1.16E-06	Intergenic	Intergenic
HODE13	3	rs13089458	153199403	G	C	0.123711	0.502795	0.106003	2.10E-06	LINC02006	intron variant
HODE13	6	rs269443	88685731	C	T	0.0721649	0.696034	0.146832	2.13E-06	Intergenic	Intergenic

HODE13	10	rs61839375	18959086	A	G	0.108247	0.541477	0.115409	2.71E-06	ARL5B	intron variant
HODE13	10	rs61840849	18909495	C	T	0.118557	0.513326	0.109689	2.87E-06	NSUN6	intron variant
HODE13	10	rs61840850	18909805	C	T	0.118557	0.513326	0.109689	2.87E-06	NSUN6	intron variant
HODE13	10	rs7097524	18927519	C	T	0.125	0.498091	0.108516	4.43E-06	NSUN6	intron variant
HODE13	3	rs79119416	153054901	A	G	0.0824742	0.568785	0.125759	6.10E-06	Intergenic	Intergenic
HODE13	11	rs78774431	83573595	G	T	0.064433	0.640413	0.141671	6.17E-06	DLG2	intron variant
HODE13	22	rs12484807	19908543	T	C	0.159794	0.45046	0.0996969	6.23E-06	TXNRD2	intron variant
HODE13	11	rs11233769	83575703	A	G	0.0618557	0.649314	0.143746	6.27E-06	DLG2	intron variant
HODE13	18	rs9748615	55505134	G	A	0.0773196	0.643725	0.142831	6.58E-06	RSL24D1P11	upstream_gene_variant
HODE13	3	rs71308446	153204436	A	G	0.110825	0.500139	0.111802	7.70E-06	LINC02006	intron variant
HODE13	6	rs72918180	88859267	T	C	0.0859375	0.555558	0.124588	8.23E-06	CNR1	upstream_gene_variant
HODE13	2	rs1008839	37252196	A	G	0.18299	0.437537	0.0987779	9.44E-06	HEATR5B	intron variant
HODE13	2	rs6731315	37252442	A	C	0.18299	0.437537	0.0987779	9.44E-06	HEATR5B	intron variant
HODE13	16	rs12051340	79324930	C	G	0.100515	0.481789	0.109616	1.11E-05	Intergenic	Intergenic
HODE13	16	rs4627364	79332660	T	C	0.100515	0.481789	0.109616	1.11E-05	Intergenic	Intergenic
HODE13	3	rs148546011	153120075	A	G	0.0706806	0.576191	0.133229	1.53E-05	LINC02006	intron variant
HODE9	5	rs72783407	95014441	A	T	0.0540541	0.901957	0.189482	1.93E-06	SPATA9	intron variant
HODE9	2	rs2110997	37233638	G	C	0.239474	0.497424	0.10769	3.86E-06	HEATR5B	intron variant
HODE9	4	rs116483655	156081825	C	T	0.0618557	0.820434	0.184029	8.27E-06	Intergenic	Intergenic
HODE9	8	rs16906565	80077909	T	G	0.0773196	0.665172	0.150703	1.02E-05	Intergenic	Intergenic
HODE9	8	rs16906571	80078158	A	G	0.0773196	0.665172	0.150703	1.02E-05	Intergenic	Intergenic
HODE9	8	rs10504697	80079370	G	A	0.0773196	0.665172	0.150703	1.02E-05	Intergenic	Intergenic
HODE9	8	rs73260164	80087816	A	G	0.0773196	0.665172	0.150703	1.02E-05	Intergenic	Intergenic
HODE9	8	rs10108874	80095928	T	C	0.0773196	0.665172	0.150703	1.02E-05	Intergenic	Intergenic

HODE9	8	rs10109106	80096130	T	C	0.0773196	0.665172	0.150703	1.02E-05	Intergenic	Intergenic
HODE9	8	rs11991170	80096463	T	C	0.0773196	0.665172	0.150703	1.02E-05	Intergenic	Intergenic
HODE9	8	rs11991175	80096486	T	C	0.0773196	0.665172	0.150703	1.02E-05	Intergenic	Intergenic
HODE9	8	rs10108084	80101215	T	C	0.0773196	0.665172	0.150703	1.02E-05	Intergenic	Intergenic
HODE9	8	rs10448041	80101494	A	G	0.0773196	0.665172	0.150703	1.02E-05	Intergenic	Intergenic
HODE9	8	rs6473140	80105521	G	C	0.0773196	0.665172	0.150703	1.02E-05	Intergenic	Intergenic
HODE9	8	rs10081534	80107372	A	G	0.0773196	0.665172	0.150703	1.02E-05	Intergenic	Intergenic
HODE9	6	rs71558239	155483008	A	G	0.126289	0.524855	0.12103	1.45E-05	TIAM2	intron variant
HODE9	6	rs3792961	155484008	C	T	0.126289	0.524855	0.12103	1.45E-05	TIAM2	intron variant
HODE9	6	rs13219130	155460725	A	G	0.113402	0.550986	0.1272	1.48E-05	TIAM2	intron variant
HODE9	7	rs4146419	52970017	T	G	0.335052	-0.358638	0.0840232	1.97E-05	SGO1P2	upstream_gene_variant
HODE9	3	rs73043318	27754292	C	G	0.231579	0.409053	0.0960261	2.05E-05	EOMES	downstream_gene_variant
HOTrE13	6	rs13207194	138354242	T	G	0.066092	0.939899	0.196797	1.79E-06	Intergenic	Intergenic
HOTrE13	2	rs4831991	85333421	T	C	0.186441	0.594212	0.128704	3.90E-06	LSM3P3	upstream_gene_variant
HOTrE13	2	rs6759185	85334209	G	A	0.186441	0.594212	0.128704	3.90E-06	LSM3P3	upstream_gene_variant
HOTrE13	2	rs57716843	85334277	T	C	0.186441	0.594212	0.128704	3.90E-06	LSM3P3	upstream_gene_variant
HOTrE13	2	rs57363422	85334310	A	G	0.186441	0.594212	0.128704	3.90E-06	LSM3P3	upstream_gene_variant
HOTrE13	2	rs61083112	85334340	C	G	0.186441	0.594212	0.128704	3.90E-06	LSM3P3	upstream_gene_variant
HOTrE13	2	rs56101385	85334927	A	C	0.186441	0.594212	0.128704	3.90E-06	Intergenic	Intergenic
HOTrE13	2	rs7585211	85335788	C	G	0.186441	0.594212	0.128704	3.90E-06	Intergenic	Intergenic
HOTrE13	2	rs10167862	85336421	C	T	0.186441	0.594212	0.128704	3.90E-06	Intergenic	Intergenic
HOTrE13	2	rs10167871	85336432	C	T	0.186441	0.594212	0.128704	3.90E-06	Intergenic	Intergenic
HOTrE13	2	rs10179005	85336492	G	A	0.186441	0.594212	0.128704	3.90E-06	Intergenic	Intergenic
HOTrE13	2	rs10179118	85336639	G	A	0.186441	0.594212	0.128704	3.90E-06	Intergenic	Intergenic

HOTrE13	2	rs59944734	85336757	G	A	0.186441	0.594212	0.128704	3.90E-06	Intergenic	Intergenic
HOTrE13	2	rs59445246	85336764	T	A	0.186441	0.594212	0.128704	3.90E-06	Intergenic	Intergenic
HOTrE13	2	rs10168285	85336786	A	T	0.186441	0.594212	0.128704	3.90E-06	Intergenic	Intergenic
HOTrE13	2	rs10203218	85336862	A	G	0.186441	0.594212	0.128704	3.90E-06	Intergenic	Intergenic
HOTrE13	2	rs10179383	85336868	G	A	0.186441	0.594212	0.128704	3.90E-06	Intergenic	Intergenic
HOTrE13	2	rs13400944	85337089	G	A	0.186441	0.594212	0.128704	3.90E-06	Intergenic	Intergenic
HOTrE13	2	rs10205874	85337409	T	C	0.186441	0.594212	0.128704	3.90E-06	Intergenic	Intergenic
HOTrE13	2	rs7573011	85339546	C	A	0.186441	0.594212	0.128704	3.90E-06	Intergenic	Intergenic
HOTrE9	10	rs11191433	83589867	G	A	0.143713	0.669492	0.141986	2.41E-06	Intergenic	Intergenic
HOTrE9	10	rs10509436	83564432	G	T	0.149701	0.646926	0.140475	4.12E-06	Intergenic	Intergenic
HOTrE9	10	rs11595719	83565340	A	G	0.149701	0.646926	0.140475	4.12E-06	Intergenic	Intergenic
HOTrE9	10	rs11191378	83574149	T	A	0.149701	0.646926	0.140475	4.12E-06	Intergenic	Intergenic
HOTrE9	10	rs12775125	83579380	A	G	0.149701	0.646926	0.140475	4.12E-06	Intergenic	Intergenic
HOTrE9	10	rs12763547	83579572	C	T	0.149701	0.646926	0.140475	4.12E-06	Intergenic	Intergenic
HOTrE9	10	rs10883786	83586702	A	C	0.149701	0.646926	0.140475	4.12E-06	Intergenic	Intergenic
HOTrE9	10	rs61862891	83597042	A	G	0.149701	0.646926	0.140475	4.12E-06	Intergenic	Intergenic
HOTrE9	12	rs12826562	25470053	T	C	0.40303	0.487406	0.106587	4.81E-06	Intergenic	Intergenic
HOTrE9	10	rs10883755	83565922	T	G	0.152695	0.637544	0.139769	5.08E-06	Intergenic	Intergenic
HOTrE9	10	rs9787485	83566686	T	C	0.152695	0.637544	0.139769	5.08E-06	Intergenic	Intergenic
HOTrE9	10	rs764866	83581933	C	T	0.152695	0.637544	0.139769	5.08E-06	Intergenic	Intergenic
HOTrE9	10	rs12267783	83583540	C	T	0.152695	0.637544	0.139769	5.08E-06	Intergenic	Intergenic
HOTrE9	10	rs11191462	83593118	T	A	0.152695	0.637544	0.139769	5.08E-06	Intergenic	Intergenic
HOTrE9	14	rs12895726	21587865	T	C	0.119632	0.65906	0.151246	1.32E-05	AL161668.1	downstream_gene_variant
HOTrE9	14	rs12434837	21588423	T	C	0.119632	0.65906	0.151246	1.32E-05	AL161668.1	downstream_gene_variant

HOTrE9	12	rs73209558	93320674	T	C	0.347561	0.451236	0.104332	1.53E-05	EEA1	intron variant
HOTrE9	12	rs4760404	93321545	C	T	0.347561	0.451236	0.104332	1.53E-05	EEA1	intron variant
HOTrE9	12	rs56290130	93327911	A	T	0.347561	0.451236	0.104332	1.53E-05	EEA1	upstream_gene_variant
HOTrE9	7	rs2711332	109589780	C	T	0.149701	0.530882	0.122853	1.55E-05	Intergenic	Intergenic
LA	10	rs72827118	113228770	A	G	0.0710526	0.817971	0.169184	1.33E-06	Intergenic	Intergenic
LA	10	rs72823394	113138073	C	T	0.0721925	0.815536	0.169376	1.47E-06	Intergenic	Intergenic
LA	18	rs9748615	55505134	G	A	0.0773196	0.799794	0.175093	4.93E-06	RSL24D1P11	upstream_gene_variant
LA	3	rs73067298	47270336	T	A	0.05	0.917653	0.201759	5.41E-06	KIF9	intron variant
LA	6	rs56080056	126487517	C	T	0.0670103	0.832144	0.185827	7.53E-06	TRMT11	intron variant/
LA	3	rs73075642	47029994	A	G	0.0481283	0.917566	0.206306	8.68E-06	NBEAL2	intron variant
LA	13	rs61056805	22666786	G	T	0.0968586	0.69173	0.156284	9.59E-06	AL136962.1	intron variant
LA	13	rs9552594	22670379	G	T	0.0953608	0.685898	0.156013	1.10E-05	NME1P1	downstream_gene_variant
LA	15	rs1349681	70075181	A	C	0.255155	0.407469	0.0937408	1.38E-05	DRAIC	intron variant
LA	15	rs1812135	70076552	G	C	0.255155	0.407469	0.0937408	1.38E-05	DRAIC	intron variant
LA	15	rs34974792	70078818	G	A	0.255155	0.407469	0.0937408	1.38E-05	DRAIC	intron variant
LA	6	rs13207194	138354242	T	G	0.0602094	0.851216	0.196041	1.41E-05	Intergenic	Intergenic
LA	18	rs9945892	29786094	C	G	0.0876289	0.655531	0.151129	1.44E-05	MEP1B	intron variant
LA	13	rs7999220	22665512	T	C	0.100515	0.657801	0.153124	1.74E-05	AL136962.1	intron variant
LA	13	rs9552593	22665146	A	G	0.0902062	0.680903	0.159178	1.89E-05	AL136962.1	intron variant
LA	18	rs9965114	29816328	C	T	0.201031	0.46158	0.108015	1.93E-05	GAREM1	intron variant
LA	18	rs11081752	29816693	C	T	0.201031	0.46158	0.108015	1.93E-05	GAREM1	intron variant
LA	6	rs10458204	6996479	C	T	0.131443	0.583135	0.136723	2.00E-05	AL139390.1	downstream_gene_variant
LA	6	rs17142729	7000007	A	T	0.131443	0.583135	0.136723	2.00E-05	AL139390.1	downstream_gene_variant
LA	6	rs11754474	7000874	A	G	0.131443	0.583135	0.136723	2.00E-05	AL139390.1	downstream_gene_variant

OxoODE13	4	rs58081193	92124352	A	G	0.178571	0.679209	0.140541	1.35E-06	CCSER1	intron variant
OxoODE13	11	rs72993799	107623451	T	C	0.0637755	1.02314	0.22485	5.36E-06	Intergenic	Intergenic
OxoODE13	11	rs11212359	107638761	C	T	0.0637755	1.02314	0.22485	5.36E-06	Intergenic	Intergenic
OxoODE13	11	rs72999506	107639063	T	C	0.0637755	1.02314	0.22485	5.36E-06	Intergenic	Intergenic
OxoODE13	11	rs4609571	107639746	C	T	0.0637755	1.02314	0.22485	5.36E-06	Intergenic	Intergenic
OxoODE13	11	rs17107403	107640811	G	A	0.0637755	1.02314	0.22485	5.36E-06	Intergenic	Intergenic
OxoODE13	11	rs72999510	107644427	A	G	0.0637755	1.02314	0.22485	5.36E-06	Intergenic	Intergenic
OxoODE13	11	rs72999515	107644933	T	C	0.0637755	1.02314	0.22485	5.36E-06	Intergenic	Intergenic
OxoODE13	11	rs72999519	107645656	A	G	0.0637755	1.02314	0.22485	5.36E-06	Intergenic	Intergenic
OxoODE13	11	rs111925725	107647611	A	T	0.0637755	1.02314	0.22485	5.36E-06	Intergenic	Intergenic
OxoODE13	11	rs72999526	107651601	T	C	0.0637755	1.02314	0.22485	5.36E-06	Intergenic	Intergenic
OxoODE13	11	rs11212373	107657264	T	C	0.0637755	1.02314	0.22485	5.36E-06	SLC35F2	downstream_gene_variant
OxoODE13	11	rs72999536	107659186	G	A	0.0637755	1.02314	0.22485	5.36E-06	SLC35F2	downstream_gene_variant
OxoODE13	11	rs12282711	107659907	C	T	0.0637755	1.02314	0.22485	5.36E-06	SLC35F2	downstream_gene_variant
OxoODE13	12	rs12424560	96543225	A	G	0.32398	0.513933	0.11302	5.43E-06	Intergenic	Intergenic
OxoODE13	2	rs71420370	60916058	G	A	0.0612245	1.01964	0.225239	5.98E-06	Intergenic	Intergenic
OxoODE13	12	rs61943941	129812028	G	A	0.0535714	1.04198	0.230401	6.11E-06	TMEM132D	intron variant
OxoODE13	18	rs17677569	60798610	C	G	0.109694	0.758253	0.170343	8.53E-06	BCL2	intron variant
OxoODE13	18	rs72941346	60804723	G	A	0.109694	0.758253	0.170343	8.53E-06	BCL2	intron variant
OxoODE13	18	rs72941348	60805100	C	T	0.109694	0.758253	0.170343	8.53E-06	BCL2	intron variant
OxoODE9	6	rs12209958	154552636	C	T	0.0515464	1.16244	0.217926	9.60E-08	IPCEF1	intron variant
OxoODE9	5	rs141610431	32040835	T	C	0.123711	0.680946	0.14822	4.34E-06	PDZD2	intron variant
OxoODE9	5	rs138572800	32040848	A	G	0.123711	0.680946	0.14822	4.34E-06	PDZD2	intron variant
OxoODE9	7	rs117732018	24852686	C	T	0.0541237	0.938885	0.209853	7.68E-06	OSBPL3	intron variant

OxoODE9	4	rs2726641	55662427	C	T	0.164948	-0.565976	0.128841	1.12E-05	Intergenic	Intergenic
OxoODE9	6	rs12524163	110320664	T	C	0.149733	0.605907	0.138021	1.13E-05	Intergenic	Intergenic
OxoODE9	6	rs6899476	110332690	C	G	0.149733	0.605907	0.138021	1.13E-05	Intergenic	Intergenic
OxoODE9	10	rs9420133	114584019	G	A	0.0953608	0.760048	0.173216	1.14E-05	AL158212.3	non_coding_transcript
OxoODE9	7	rs10274129	24880086	T	C	0.0541237	0.913099	0.2084	1.18E-05	OSBPL3	intron variant
OxoODE9	7	rs76635739	24889040	A	G	0.0541237	0.913099	0.2084	1.18E-05	OSBPL3	intron variant
OxoODE9	6	rs3930218	154562302	A	G	0.056701	0.890734	0.208004	1.85E-05	IPCEF1	intron variant
OxoODE9	6	rs6935917	154570012	A	G	0.056701	0.890734	0.208004	1.85E-05	IPCEF1	intron variant
OxoODE9	6	rs9322454	154570651	A	G	0.056701	0.890734	0.208004	1.85E-05	IPCEF1	intron variant
OxoODE9	6	rs6918042	154573837	T	C	0.056701	0.890734	0.208004	1.85E-05	IPCEF1	intron variant
OxoODE9	3	rs4632589	191039740	T	C	0.476804	-0.417968	0.0976521	1.87E-05	UTS2B	intron variant
OxoODE9	3	rs4677722	191040174	C	T	0.476804	-0.417968	0.0976521	1.87E-05	UTS2B	intron variant
OxoODE9	22	rs80209983	27554578	C	G	0.0489691	0.930696	0.21843	2.04E-05	AL008638.1	intron variant
OxoODE9	22	rs738155	27555669	T	A	0.0489691	0.930696	0.21843	2.04E-05	AL008638.1	non_coding_transcript
OxoODE9	9	rs12000930	8877814	C	T	0.064433	0.85771	0.201646	2.10E-05	PTPRD	intron variant
OxoODE9	9	rs12001347	8878043	C	T	0.064433	0.85771	0.201646	2.10E-05	PTPRD	intron variant
TransEKODE	4	rs75592902	99324388	G	A	0.0984456	0.89773	0.165327	5.64E-08	RAP1GDS1	intron variant
TransEKODE	6	rs1535358	14161941	G	A	0.0958549	0.814707	0.169036	1.44E-06	AL133259.1	intron variant
TransEKODE	6	rs9476520	14155648	C	T	0.0595855	1.02365	0.215312	1.99E-06	RNU7-133P	upstream_gene_variant
TransEKODE	6	rs9296927	14158382	G	T	0.0595855	1.02365	0.215312	1.99E-06	RNU7-133P	downstream_gene_variant
TransEKODE	6	rs76634115	14159297	A	G	0.0595855	1.02365	0.215312	1.99E-06	RNU7-133P	downstream_gene_variant
TransEKODE	6	rs2224478	14160830	A	C	0.0595855	1.02365	0.215312	1.99E-06	RNU7-133P	downstream_gene_variant
TransEKODE	1	rs4540653	15515068	T	C	0.165803	0.653338	0.138039	2.21E-06	TMEM51	intron variant
TransEKODE	1	rs1892200	15520138	G	A	0.165803	0.653338	0.138039	2.21E-06	TMEM51	intron variant

TransEKODE	1	rs6692220	15520583	G	A	0.165803	0.653338	0.138039	2.21E-06	TMEM51	intron variant
TransEKODE	1	rs4661326	15527323	C	T	0.165803	0.653338	0.138039	2.21E-06	TMEM51	intron variant
TransEKODE	1	rs7533078	15529107	T	C	0.165803	0.653338	0.138039	2.21E-06	TMEM51	intron variant
TransEKODE	1	rs6429726	15535116	T	A	0.165803	0.653338	0.138039	2.21E-06	TMEM51	intron variant
TransEKODE	9	rs74731856	126291925	G	A	0.0544041	1.02009	0.218705	3.10E-06	DENND1A	intron variant
TransEKODE	9	rs77719717	126306001	G	C	0.0544041	1.02009	0.218705	3.10E-06	DENND1A	intron variant
TransEKODE	1	rs55767642	91946815	G	A	0.225389	0.548886	0.120112	4.88E-06	Intergenic	Intergenic
TransEKODE	1	rs1848434	91947282	C	A	0.225389	0.548886	0.120112	4.88E-06	Intergenic	Intergenic
TransEKODE	1	rs61798796	91949251	G	A	0.225389	0.548886	0.120112	4.88E-06	Intergenic	Intergenic
TransEKODE	1	rs1892198	15538470	C	T	0.145078	0.673241	0.147812	5.25E-06	TMEM51	intron variant
TransEKODE	4	rs12507680	96746348	T	C	0.0595855	0.953193	0.210671	6.05E-06	Intergenic	Intergenic
TransEKODE	4	rs116027502	98674800	A	G	0.0595855	1.02667	0.22714	6.18E-06	STPG2	intron variant
aLA	6	rs13207194	138354242	T	G	0.0598958	0.976482	0.193125	4.28E-07	Intergenic	Intergenic
aLA	3	rs6787082	176417139	A	G	0.074359	0.785001	0.168207	3.06E-06	LINC01208	intron variant
aLA	3	rs12696418	176418026	A	C	0.074359	0.785001	0.168207	3.06E-06	LINC01208	intron variant
aLA	3	rs9826338	176418794	A	G	0.074359	0.785001	0.168207	3.06E-06	LINC01208	intron variant
aLA	3	rs9850387	176419218	C	T	0.074359	0.785001	0.168207	3.06E-06	LINC01208	intron variant
aLA	3	rs3919982	176419227	T	C	0.074359	0.785001	0.168207	3.06E-06	LINC01208	intron variant
aLA	3	rs9855212	176419860	C	T	0.074359	0.785001	0.168207	3.06E-06	LINC01208	intron variant
aLA	3	rs1490075	176422652	T	C	0.074359	0.785001	0.168207	3.06E-06	LINC01208	intron variant
aLA	3	rs4857658	176423123	G	A	0.074359	0.785001	0.168207	3.06E-06	LINC01208	intron variant
aLA	3	rs13065456	176423670	G	C	0.074359	0.785001	0.168207	3.06E-06	LINC01208	intron variant
aLA	3	rs146433028	176425502	G	A	0.074359	0.785001	0.168207	3.06E-06	LINC01208	intron variant
aLA	3	rs6443364	176427160	A	G	0.074359	0.785001	0.168207	3.06E-06	LINC01208	intron variant

aLA	3	rs6763891	176428151	C	A	0.074359	0.785001	0.168207	3.06E-06	LINC01208	intron variant
aLA	3	rs9855942	176429322	A	G	0.074359	0.785001	0.168207	3.06E-06	LINC01208	intron variant
aLA	3	rs9865930	176431109	T	C	0.074359	0.785001	0.168207	3.06E-06	LINC01208	intron variant
aLA	3	rs9810728	176431123	C	T	0.074359	0.785001	0.168207	3.06E-06	LINC01208	intron variant
aLA	12	rs78435498	53618594	A	G	0.0598958	0.870549	0.193125	6.55E-06	RARG	upstream_gene_variant
aLA	12	rs79988442	53623397	A	G	0.0598958	0.870549	0.193125	6.55E-06	RARG	intron variant
aLA	2	rs4831990	85333390	A	G	0.15641	0.549152	0.12553	1.22E-05	LSM3P3	upstream_gene_variant
aLA	2	rs10208954	85335014	C	T	0.15641	0.549152	0.12553	1.22E-05	Intergenic	Intergenic
Lox1	2	rs2110997	37233638	G	C	0.239474	0.500976	0.107976	3.49E-06	HEATR5B	intron variant
Lox1	5	rs72783407	95014441	A	T	0.0540541	0.862324	0.189985	5.65E-06	SPATA9	intron variant
Lox1	7	rs6464585	139896424	T	A	0.391753	0.398023	0.092933	1.84E-05	Intergenic	Intergenic
Lox1	7	rs10279097	139902694	T	C	0.391753	0.398023	0.092933	1.84E-05	Intergenic	Intergenic
Lox1	4	rs76190277	13124083	G	T	0.056701	0.817493	0.191598	1.98E-05	Intergenic	Intergenic
Lox1	8	rs16906565	80077909	T	G	0.0773196	0.6446	0.151103	1.99E-05	Intergenic	Intergenic
Lox1	8	rs16906571	80078158	A	G	0.0773196	0.6446	0.151103	1.99E-05	Intergenic	Intergenic
Lox1	8	rs10504697	80079370	G	A	0.0773196	0.6446	0.151103	1.99E-05	Intergenic	Intergenic
Lox1	8	rs73260164	80087816	A	G	0.0773196	0.6446	0.151103	1.99E-05	Intergenic	Intergenic
Lox1	8	rs10108874	80095928	T	C	0.0773196	0.6446	0.151103	1.99E-05	Intergenic	Intergenic
Lox1	8	rs10109106	80096130	T	C	0.0773196	0.6446	0.151103	1.99E-05	Intergenic	Intergenic
Lox1	8	rs11991170	80096463	T	C	0.0773196	0.6446	0.151103	1.99E-05	Intergenic	Intergenic
Lox1	8	rs11991175	80096486	T	C	0.0773196	0.6446	0.151103	1.99E-05	Intergenic	Intergenic
Lox1	8	rs10108084	80101215	T	C	0.0773196	0.6446	0.151103	1.99E-05	Intergenic	Intergenic
Lox1	8	rs10448041	80101494	A	G	0.0773196	0.6446	0.151103	1.99E-05	Intergenic	Intergenic
Lox1	8	rs6473140	80105521	G	C	0.0773196	0.6446	0.151103	1.99E-05	Intergenic	Intergenic

Lox1	8	rs10081534	80107372	A	G	0.0773196	0.6446	0.151103	1.99E-05	Intergenic	Intergenic
Lox1	2	rs76878347	185469031	A	G	0.0902062	0.646978	0.152756	2.28E-05	ZNF804A	intron variant
Lox1	2	rs722384	185476009	G	T	0.0902062	0.646978	0.152756	2.28E-05	ZNF804A	intron variant
Lox1	2	rs72625293	185499138	C	T	0.0902062	0.646978	0.152756	2.28E-05	ZNF804A	intron variant
cyp450	6	rs10458204	6996479	C	T	0.132812	0.551962	0.116534	2.17E-06	AL139390.1	downstream_gene_variant
cyp450	6	rs17142729	7000007	A	T	0.132812	0.551962	0.116534	2.17E-06	AL139390.1	downstream_gene_variant
cyp450	6	rs11754474	7000874	A	G	0.132812	0.551962	0.116534	2.17E-06	AL139390.1	downstream_gene_variant
cyp450	6	rs4959421	7002174	C	T	0.132812	0.551962	0.116534	2.17E-06	Intergenic	Intergenic
cyp450	6	rs4959422	7002414	G	C	0.132812	0.551962	0.116534	2.17E-06	Intergenic	Intergenic
cyp450	6	rs11243133	7003510	C	A	0.132812	0.551962	0.116534	2.17E-06	Intergenic	Intergenic
cyp450	6	rs11243134	7003600	G	A	0.132812	0.551962	0.116534	2.17E-06	Intergenic	Intergenic
cyp450	6	rs11243135	7003888	A	G	0.132812	0.551962	0.116534	2.17E-06	Intergenic	Intergenic
cyp450	6	rs35922242	7006267	C	G	0.132812	0.551962	0.116534	2.17E-06	Intergenic	Intergenic
cyp450	6	rs12210411	7009380	G	A	0.132812	0.551962	0.116534	2.17E-06	Intergenic	Intergenic
cyp450	14	rs11850035	82252646	A	G	0.0885417	0.605406	0.132678	5.04E-06	AL355838.1	intron variant
cyp450	14	rs76411313	82255384	A	G	0.0885417	0.605406	0.132678	5.04E-06	AL355838.1	intron variant
cyp450	14	rs28573857	82258342	A	G	0.0885417	0.605406	0.132678	5.04E-06	AL355838.1	intron variant
cyp450	14	rs8006442	82258639	T	G	0.0885417	0.605406	0.132678	5.04E-06	AL355838.1	intron variant
cyp450	14	rs8007879	82276342	T	G	0.0885417	0.605406	0.132678	5.04E-06	AL355838.1	intron variant
cyp450	15	rs28478777	55512746	T	G	0.164062	0.462665	0.101793	5.49E-06	RAB27A	intron variant
cyp450	7	rs113639949	157648298	T	C	0.0952381	0.607799	0.134448	6.16E-06	PTPRN2	intron variant
cyp450	3	rs60899259	55323856	T	A	0.1875	0.463185	0.103061	6.98E-06	LINC02030	intron variant
cyp450	2	rs72803684	26192802	T	C	0.0718085	0.670434	0.149693	7.51E-06	KIF3C	intron variant
cyp450	16	rs62039713	79355906	G	A	0.0520833	0.74827	0.168122	8.56E-06	Intergenic	Intergenic

epdi9	3	rs7617700	36417251	G	A	0.244253	0.327624	0.00976317	7.08E-247	STAC	upstream_gene_variant
epdi9	7	rs17305725	94740381	G	A	0.0747126	0.428297	0.013567	9.84E-219	PPP1R9A	intron variant
epdi9	10	rs11188900	98503103	T	C	0.17052	0.291452	0.00933483	5.35E-214	RNU6-1274P	downstream_gene_variant
epdi9	11	rs11021824	11395983	G	T	0.311765	0.229307	0.00735766	3.10E-213	GALNT18	intron variant
epdi9	6	rs35868295	106583671	G	T	0.123563	0.323456	0.0106267	1.73E-203	ATG5	intron variant/
epdi9	13	rs9533173	43186976	C	T	0.488506	-0.20324	0.00669653	2.51E-202	TNFSF11	downstream_gene_variant
epdi9	6	rs62425215	120619602	G	C	0.235632	0.206796	0.00681598	3.42E-202	Intergenic	Intergenic
epdi9	6	rs12661355	162559158	G	A	0.144578	0.344947	0.0114622	5.75E-199	PRKN	intron variant
epdi9	6	rs114708313	31329004	T	A	0.148256	0.267596	0.00892814	2.26E-197	HLA-B	upstream_gene_variant
epdi9	3	rs62245716	36418521	G	A	0.284483	0.274338	0.0091535	2.36E-197	STAC	upstream_gene_variant
epdi9	6	rs742108	106582920	A	G	0.16954	0.302316	0.0100933	4.12E-197	ATG5	intron variant/
epdi9	6	rs9320151	106583964	T	C	0.16954	0.302316	0.0100933	4.12E-197	ATG5	intron variant/
epdi9	6	rs34408152	106586776	C	T	0.16954	0.302316	0.0100933	4.12E-197	ATG5	intron variant/
epdi9	6	rs11753719	106590062	G	C	0.16954	0.302316	0.0100933	4.12E-197	ATG5	intron variant/
epdi9	6	rs11753795	106590404	T	C	0.16954	0.302316	0.0100933	4.12E-197	ATG5	intron variant/
epdi9	6	rs11758444	106590444	G	A	0.16954	0.302316	0.0100933	4.12E-197	ATG5	intron variant/
epdi9	6	rs11758474	106590566	G	A	0.16954	0.302316	0.0100933	4.12E-197	ATG5	intron variant/
epdi9	6	rs6935163	82299640	T	C	0.146552	0.299677	0.0100083	5.47E-197	TENT5A	intron variant
epdi9	6	rs2517917	29781020	C	T	0.186782	0.27946	0.00939148	1.42E-194	MICG	upstream_gene_variant
epdi9	10	rs942790	81069268	G	C	0.342105	0.194183	0.00652695	1.69E-194	ZMIZ1	intron variant
lox15	10	rs72827118	113228770	A	G	0.0710526	0.702423	0.141912	7.43E-07	Intergenic	Intergenic
lox15	10	rs72823394	113138073	C	T	0.0721925	0.693075	0.142073	1.07E-06	Intergenic	Intergenic
lox15	6	rs269443	88685731	C	T	0.0721649	0.716454	0.151105	2.12E-06	Intergenic	Intergenic
lox15	3	rs13089458	153199403	G	C	0.123711	0.506234	0.108905	3.35E-06	LINC02006	intron variant

lox15	11	rs78774431	83573595	G	T	0.064433	0.659476	0.145674	5.98E-06	DLG2	intron variant
lox15	11	rs11233769	83575703	A	G	0.0618557	0.66902	0.147833	6.03E-06	DLG2	intron variant
lox15	10	rs61840849	18909495	C	T	0.118557	0.50752	0.11267	6.65E-06	NSUN6	intron variant
lox15	10	rs61840850	18909805	C	T	0.118557	0.50752	0.11267	6.65E-06	NSUN6	intron variant
lox15	3	rs79119416	153054901	A	G	0.0824742	0.577258	0.129298	8.02E-06	Intergenic	Intergenic
lox15	10	rs61839375	18959086	A	G	0.108247	0.52797	0.118576	8.48E-06	ARL5B	intron variant
lox15	10	rs7097524	18927519	C	T	0.125	0.494786	0.111502	9.10E-06	NSUN6	intron variant
lox15	6	rs72918180	88859267	T	C	0.0859375	0.567002	0.128117	9.61E-06	CNR1	upstream_gene_variant
lox15	18	rs9748615	55505134	G	A	0.0773196	0.646633	0.146869	1.07E-05	RSL24D1P11	upstream_gene_variant
lox15	16	rs12051340	79324930	C	G	0.100515	0.487801	0.112582	1.47E-05	Intergenic	Intergenic
lox15	16	rs4627364	79332660	T	C	0.100515	0.487801	0.112582	1.47E-05	Intergenic	Intergenic
lox15	2	rs1008839	37252196	A	G	0.18299	0.438914	0.10158	1.55E-05	HEATR5B	intron variant
lox15	2	rs6731315	37252442	A	C	0.18299	0.438914	0.10158	1.55E-05	HEATR5B	intron variant
lox15	22	rs12484807	19908543	T	C	0.159794	0.44226	0.102385	1.56E-05	TXNRD2	intron variant
lox15	3	rs71308446	153204436	A	G	0.110825	0.496068	0.114855	1.57E-05	LINC02006	intron variant
lox15	2	rs17189524	54129614	G	T	0.0876289	-0.602212	0.139673	1.62E-05	PSME4	intron variant/
lox5	2	rs2110997	37233638	G	C	0.239474	0.499115	0.108855	4.54E-06	HEATR5B	Intergenic
lox5	5	rs72783407	95014441	A	T	0.0540541	0.854633	0.191531	8.12E-06	SPATA9	Intergenic
lox5	12	rs56180838	2760898	T	C	0.0767196	0.667475	0.15434	1.53E-05	CACNA1C	Intergenic
lox5	7	rs4146419	52970017	T	G	0.335052	-0.366348	0.0849318	1.61E-05	SGO1P2	Intergenic
lox5	7	rs6464585	139896424	T	A	0.391753	0.397666	0.0936895	2.19E-05	Intergenic	Intergenic
lox5	7	rs10279097	139902694	T	C	0.391753	0.397666	0.0936895	2.19E-05	Intergenic	Intergenic
lox5	8	rs16906565	80077909	T	G	0.0773196	0.645272	0.152333	2.28E-05	Intergenic	Intergenic
lox5	8	rs16906571	80078158	A	G	0.0773196	0.645272	0.152333	2.28E-05	Intergenic	Intergenic

lox5	8	rs10504697	80079370	G	A	0.0773196	0.645272	0.152333	2.28E-05	Intergenic	Intergenic
lox5	8	rs73260164	80087816	A	G	0.0773196	0.645272	0.152333	2.28E-05	Intergenic	Intergenic
lox5	8	rs10108874	80095928	T	C	0.0773196	0.645272	0.152333	2.28E-05	Intergenic	Intergenic
lox5	8	rs10109106	80096130	T	C	0.0773196	0.645272	0.152333	2.28E-05	Intergenic	Intergenic
lox5	8	rs11991170	80096463	T	C	0.0773196	0.645272	0.152333	2.28E-05	Intergenic	Intergenic
lox5	8	rs11991175	80096486	T	C	0.0773196	0.645272	0.152333	2.28E-05	Intergenic	Intergenic
lox5	8	rs10108084	80101215	T	C	0.0773196	0.645272	0.152333	2.28E-05	Intergenic	Intergenic
lox5	8	rs10448041	80101494	A	G	0.0773196	0.645272	0.152333	2.28E-05	Intergenic	Intergenic
lox5	8	rs6473140	80105521	G	C	0.0773196	0.645272	0.152333	2.28E-05	Intergenic	Intergenic
lox5	8	rs10081534	80107372	A	G	0.0773196	0.645272	0.152333	2.28E-05	Intergenic	Intergenic
lox5	13	rs9552593	22665146	A	G	0.0902062	0.650519	0.154	2.40E-05	AL136962.1	Intergenic
lox5	2	rs1008839	37252196	A	G	0.18299	0.489976	0.117162	2.89E-05	HEATR5B	Intergenic
omega3	1	rs7526572	146502843	A	C	0.0769231	0.883375	0.173611	3.61E-07	NBPF13P	intron variant
omega3	1	rs6593800	146509537	C	A	0.0769231	0.883375	0.173611	3.61E-07	NBPF13P	intron variant
omega3	1	rs7514206	146512510	A	G	0.0769231	0.883375	0.173611	3.61E-07	NBPF13P	intron variant
omega3	1	rs12021797	146514248	T	A	0.0769231	0.883375	0.173611	3.61E-07	NBPF13P	intron variant
omega3	1	rs3753431	146514442	G	A	0.0769231	0.883375	0.173611	3.61E-07	NBPF13P	intron variant
omega3	1	rs6657631	146517465	A	G	0.0769231	0.883375	0.173611	3.61E-07	AC244394.1	upstream_gene_variant
omega3	1	rs10047061	146518179	T	G	0.0769231	0.883375	0.173611	3.61E-07	AC244394.1	upstream_gene_variant
omega3	1	rs7527352	146533724	T	C	0.0769231	0.883375	0.173611	3.61E-07	Intergenic	Intergenic
omega3	1	rs61349218	146541646	A	G	0.0769231	0.883375	0.173611	3.61E-07	Intergenic	Intergenic
omega3	1	rs6593801	146512805	G	A	0.187179	0.538102	0.115688	3.30E-06	NBPF13P	intron variant
omega3	1	rs6661540	146512899	A	G	0.187179	0.538102	0.115688	3.30E-06	NBPF13P	intron variant
omega3	1	rs12047213	146513184	C	T	0.187179	0.538102	0.115688	3.30E-06	NBPF13P	intron variant

omega3	1	rs1837981	146513783	T	C	0.187179	0.538102	0.115688	3.30E-06	NBPF13P	intron variant
omega3	1	rs1837982	146514117	C	A	0.187179	0.538102	0.115688	3.30E-06	NBPF13P	intron variant
omega3	1	rs11240120	146515813	G	A	0.187179	0.538102	0.115688	3.30E-06	NBPF13P	intron variant
omega3	1	rs4950483	146516128	C	T	0.187179	0.538102	0.115688	3.30E-06	NBPF13P	intron variant
omega3	1	rs4950484	146516199	T	G	0.187179	0.538102	0.115688	3.30E-06	NBPF13P	intron variant
omega3	1	rs6593810	146517287	G	C	0.187179	0.538102	0.115688	3.30E-06	NBPF13P	intron variant
omega3	1	rs12059102	146517790	A	G	0.187179	0.538102	0.115688	3.30E-06	AC244394.1	upstream_gene_variant
omega3	1	rs12070653	146517801	T	C	0.187179	0.538102	0.115688	3.30E-06	AC244394.1	upstream_gene_variant
omega6	10	rs72827118	113228770	A	G	0.0710526	0.812712	0.169234	1.57E-06	Intergenic	Intergenic
omega6	10	rs72823394	113138073	C	T	0.0721925	0.810244	0.169426	1.73E-06	Intergenic	Intergenic
omega6	18	rs9748615	55505134	G	A	0.0773196	0.797075	0.175144	5.34E-06	RSL24D1P11	upstream_gene_variant
omega6	3	rs73067298	47270336	T	A	0.05	0.914104	0.201818	5.92E-06	KIF9	intron variant
omega6	6	rs56080056	126487517	C	T	0.0670103	0.840791	0.185882	6.09E-06	TRMT11	intron variant/
omega6	3	rs73075642	47029994	A	G	0.0481283	0.912021	0.206367	9.90E-06	NBEAL2	intron variant
omega6	15	rs1349681	70075181	A	C	0.255155	0.413194	0.0937684	1.05E-05	DRAIC	intron variant
omega6	15	rs1812135	70076552	G	C	0.255155	0.413194	0.0937684	1.05E-05	DRAIC	intron variant
omega6	15	rs34974792	70078818	G	A	0.255155	0.413194	0.0937684	1.05E-05	DRAIC	intron variant
omega6	13	rs61056805	22666786	G	T	0.0968586	0.688275	0.15633	1.07E-05	AL136962.1	intron variant
omega6	13	rs9552594	22670379	G	T	0.0953608	0.682747	0.156059	1.21E-05	NME1P1	downstream_gene_variant
omega6	6	rs10458204	6996479	C	T	0.131443	0.593932	0.136763	1.41E-05	AL139390.1	downstream_gene_variant
omega6	6	rs17142729	7000007	A	T	0.131443	0.593932	0.136763	1.41E-05	AL139390.1	downstream_gene_variant
omega6	6	rs11754474	7000874	A	G	0.131443	0.593932	0.136763	1.41E-05	AL139390.1	downstream_gene_variant
omega6	6	rs4959421	7002174	C	T	0.131443	0.593932	0.136763	1.41E-05	Intergenic	Intergenic
omega6	6	rs4959422	7002414	G	C	0.131443	0.593932	0.136763	1.41E-05	Intergenic	Intergenic

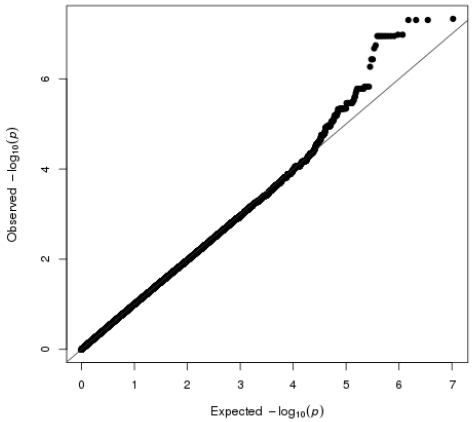
omega6	6	rs11243133	7003510	C	A	0.131443	0.593932	0.136763	1.41E-05	Intergenic	Intergenic
omega6	6	rs11243134	7003600	G	A	0.131443	0.593932	0.136763	1.41E-05	Intergenic	Intergenic
omega6	6	rs11243135	7003888	A	G	0.131443	0.593932	0.136763	1.41E-05	Intergenic	Intergenic
omega6	6	rs35922242	7006267	C	G	0.131443	0.593932	0.136763	1.41E-05	Intergenic	Intergenic
oxho13	1	rs115910760	50806455	C	T	0.0518135	0.969955	0.17912	6.12E-08	Intergenic	Intergenic
oxho13	1	rs111404284	177773801	A	C	0.0595855	0.879448	0.162682	6.45E-08	AL136114.1	intron variant
oxho13	18	rs9954493	51287076	A	G	0.168394	0.596263	0.117064	3.52E-07	Intergenic	Intergenic
oxho13	16	rs72799589	84945152	A	C	0.0508021	0.882119	0.176332	5.66E-07	CRISPLD2	downstream_gene_variant
oxho13	5	rs72778965	103868573	G	A	0.0621622	0.804173	0.163521	8.75E-07	AC099520.1	intron variant
oxho13	18	rs34395753	51204164	A	G	0.10582	0.665788	0.135679	9.24E-07	Intergenic	Intergenic
oxho13	1	rs6672992	65848034	C	A	0.147668	0.577074	0.117854	9.75E-07	DNAJC6	intron variant
oxho13	4	rs4697075	24902583	A	C	0.106218	0.61369	0.126266	1.17E-06	CCDC149	intron variant
oxho13	4	rs7688687	24905744	C	T	0.106218	0.61369	0.126266	1.17E-06	CCDC149	intron variant
oxho13	4	rs78607606	24906009	A	G	0.106218	0.61369	0.126266	1.17E-06	CCDC149	intron variant
oxho13	4	rs111933222	24906722	T	C	0.106218	0.61369	0.126266	1.17E-06	CCDC149	intron variant
oxho13	4	rs79711558	24907805	C	T	0.106218	0.61369	0.126266	1.17E-06	CCDC149	intron variant
oxho13	4	rs4697498	24908208	T	C	0.106218	0.61369	0.126266	1.17E-06	CCDC149	intron variant
oxho13	4	rs76532256	24913715	C	G	0.106218	0.61369	0.126266	1.17E-06	CCDC149	intron variant
oxho13	18	rs2339872	51452359	T	C	0.0673575	0.791706	0.163432	1.27E-06	Intergenic	Intergenic
oxho13	18	rs6508282	51513691	T	C	0.0673575	0.791706	0.163432	1.27E-06	Intergenic	Intergenic
oxho13	18	rs1995139	51522876	A	C	0.0673575	0.791706	0.163432	1.27E-06	Intergenic	Intergenic
oxho13	11	rs12808894	127599068	G	T	0.103627	0.643076	0.133986	1.59E-06	Intergenic	Intergenic
oxho13	6	rs80071633	166243751	A	G	0.0621762	0.731153	0.154963	2.38E-06	PDE10A	intron variant
oxho13	6	rs73031975	166243753	T	A	0.0621762	0.731153	0.154963	2.38E-06	PDE10A	intron variant

oxho9	8	rs34637388	75934347	T	C	0.0628272	0.688177	0.118745	6.81E-09	CRISPLD1	intron variant
oxho9	20	rs4813952	10977550	A	G	0.065445	0.539514	0.107494	5.19E-07	AL050403.2	downstream_gene_variant
oxho9	10	rs35688001	119861279	G	A	0.159686	0.401395	0.0808961	6.98E-07	CASC2	intron variant
oxho9	7	rs2715149	82452812	T	A	0.0811518	0.463749	0.0970318	1.76E-06	PCLO	intron variant
oxho9	2	rs72836397	69134929	G	A	0.0575916	0.541473	0.113606	1.88E-06	Intergenic	Intergenic
oxho9	7	rs219843	98578933	T	G	0.0471204	0.59588	0.125226	1.95E-06	TRRAP	intron variant
oxho9	2	rs72836381	69127731	C	T	0.0575916	0.542267	0.115046	2.44E-06	Intergenic	Intergenic
oxho9	9	rs12376482	126882728	G	T	0.052356	0.565509	0.120264	2.57E-06	Intergenic	Intergenic
oxho9	9	rs10986227	126883060	C	G	0.052356	0.565509	0.120264	2.57E-06	Intergenic	Intergenic
oxho9	9	rs7861404	126883429	G	A	0.052356	0.565509	0.120264	2.57E-06	Intergenic	Intergenic
oxho9	9	rs10818920	126886936	C	T	0.052356	0.565509	0.120264	2.57E-06	Intergenic	Intergenic
oxho9	9	rs10986238	126888721	A	G	0.052356	0.565509	0.120264	2.57E-06	Intergenic	Intergenic
oxho9	1	rs115910760	50806455	C	T	0.0471204	0.604225	0.129496	3.07E-06	Intergenic	Intergenic
oxho9	9	rs118145632	102042443	A	T	0.0680628	0.511488	0.111587	4.57E-06	RN7SKP225	upstream_gene_variant
oxho9	6	rs4308609	1999350	A	G	0.162304	0.333639	0.073012	4.89E-06	GMDS	intron variant
oxho9	2	rs112207171	50980560	T	C	0.052356	0.573346	0.12557	4.97E-06	NRXN1	intron variant
oxho9	2	rs111562705	51009521	C	T	0.052356	0.573346	0.12557	4.97E-06	NRXN1	intron variant
oxho9	2	rs114433761	51026014	A	G	0.052356	0.573346	0.12557	4.97E-06	NRXN1	intron variant
oxho9	9	rs7859917	18847325	C	T	0.356021	0.27027	0.0595076	5.58E-06	ADAMTSL1	intron variant
oxho9	2	rs7604154	69126964	A	G	0.0628272	0.499138	0.110545	6.32E-06	Intergenic	Intergenic
EPXH2	12	rs10902512	132691362	G	A	0.406736	-0.381972	0.0806515	2.18E-06	GALNT9	intron variant
EPXH2	12	rs4077838	132692165	G	A	0.406736	-0.381972	0.0806515	2.18E-06	GALNT9	intron variant
EPXH2	12	rs11246977	132673707	C	G	0.460106	-0.363987	0.0795616	4.76E-06	AC138466.1	upstream_gene_variant
EPXH2	12	rs6598204	132672623	A	T	0.446809	-0.36178	0.0795323	5.39E-06	AC138466.1	upstream_gene_variant

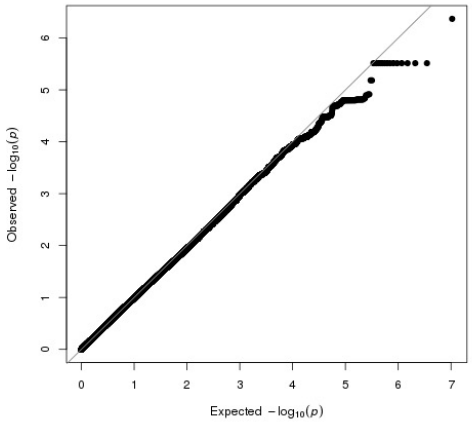
EPXH2	12	rs7961948	132685240	T	C	0.417098	-0.355914	0.0791864	6.97E-06	GALNT9	intron variant
EPXH2	12	rs10902509	132687184	A	C	0.417098	-0.355914	0.0791864	6.97E-06	GALNT9	intron variant
EPXH2	12	rs3935477	132688985	G	A	0.417098	-0.355914	0.0791864	6.97E-06	GALNT9	intron variant
EPXH2	12	rs7485291	132689898	G	A	0.414508	-0.352979	0.0789105	7.71E-06	GALNT9	intron variant
EPXH2	12	rs7485305	132689923	G	A	0.414508	-0.352979	0.0789105	7.71E-06	GALNT9	intron variant
EPXH2	4	rs10020476	139721065	T	C	0.440415	-0.344594	0.0777367	9.30E-06	AC093766.1	intron variant
EPXH2	4	rs13127451	139723260	C	T	0.440415	-0.344594	0.0777367	9.30E-06	AC093766.1	downstream_gene_variant
EPXH2	4	rs28739363	139723574	G	A	0.440415	-0.344594	0.0777367	9.30E-06	AC093766.1	downstream_gene_variant
EPXH2	4	rs6819494	139728027	G	C	0.440415	-0.344594	0.0777367	9.30E-06	AC093766.1	downstream_gene_variant
EPXH2	4	rs4532285	139723213	T	G	0.427461	-0.341139	0.077081	9.61E-06	AC093766.1	downstream_gene_variant
EPXH2	4	rs6812265	139726791	T	C	0.430052	-0.340105	0.0773582	1.10E-05	AC093766.1	downstream_gene_variant
EPXH2	4	rs6844434	139727944	C	A	0.430052	-0.340105	0.0773582	1.10E-05	AC093766.1	downstream_gene_variant
EPXH2	4	rs11724106	139721710	A	G	0.417098	-0.337451	0.0767909	1.11E-05	AC093766.1	intron variant
EPXH2	14	rs8952	50092134	T	A	0.100529	0.570976	0.131458	1.40E-05	RPL36AL	upstream_gene_variant
EPXH2	4	rs9997256	139720961	C	G	0.443005	-0.338096	0.0780418	1.48E-05	AC093766.1	intron variant
EPXH2	15	rs16976177	55509260	G	C	0.173575	0.438972	0.101408	1.50E-05	RAB27A	intron variant
SumEicos	6	rs13207194	138354242	T	G	0.0605263	0.861088	0.184832	3.18E-06	Intergenic	Intergenic
SumEicos	4	rs2309591	182945244	G	C	0.0958549	0.6455	0.147123	1.15E-05	TENM3-AS1	intron variant
SumEicos	1	rs12139677	181703276	T	G	0.279793	0.423701	0.0968921	1.23E-05	CACNA1E	intron variant
SumEicos	1	rs771329	181738708	G	C	0.352332	0.387475	0.0905848	1.89E-05	CACNA1E	intron variant
SumEicos	13	rs61056805	22666786	G	T	0.0947368	0.634532	0.148831	2.01E-05	AL136962.1	intron variant
SumEicos	1	rs697259	181736189	A	G	0.366492	0.390994	0.092121	2.19E-05	CACNA1E	intron variant
SumEicos	18	rs9945892	29786094	C	G	0.0880829	0.604281	0.142501	2.23E-05	MEP1B	intron variant
SumEicos	1	rs58928993	9432124	A	G	0.158031	0.489842	0.115556	2.25E-05	SPSB1	downstream_gene_variant

SumEicos	1	rs704331	181724209	A	G	0.38342	0.382643	0.0904013	2.31E-05	CACNA1E	intron variant
SumEicos	13	rs9552594	22670379	G	T	0.0932642	0.627642	0.148578	2.40E-05	NME1P1	downstream_gene_variant
SumEicos	4	rs6841832	182959520	T	C	0.103627	0.569486	0.135331	2.58E-05	TENM3-AS1	intron variant
SumEicos	4	rs2125505	182959894	T	C	0.103627	0.569486	0.135331	2.58E-05	TENM3-AS1	intron variant
SumEicos	4	rs4862020	182960136	C	T	0.103627	0.569486	0.135331	2.58E-05	TENM3-AS1	intron variant
SumEicos	14	rs1958387	25561403	T	C	0.331606	0.372944	0.0890414	2.81E-05	Intergenic	Intergenic
SumEicos	8	rs4132970	80412403	G	C	0.354922	0.378704	0.0908087	3.04E-05	Intergenic	Intergenic
SumEicos	13	rs7999220	22665512	T	C	0.0984456	0.607761	0.145734	3.04E-05	AL136962.1	intron variant
SumEicos	2	rs55975410	64885133	G	C	0.0777202	0.661741	0.158946	3.14E-05	SERTAD2	upstream_gene_variant
SumEicos	2	rs1008839	37252196	A	G	0.183938	0.473974	0.114281	3.36E-05	HEATR5B	intron variant
SumEicos	2	rs6731315	37252442	A	C	0.183938	0.473974	0.114281	3.36E-05	HEATR5B	intron variant
SumEicos	8	rs10102397	92143929	A	T	0.0518135	0.8138	0.196274	3.38E-05	LRRC69	intron variant

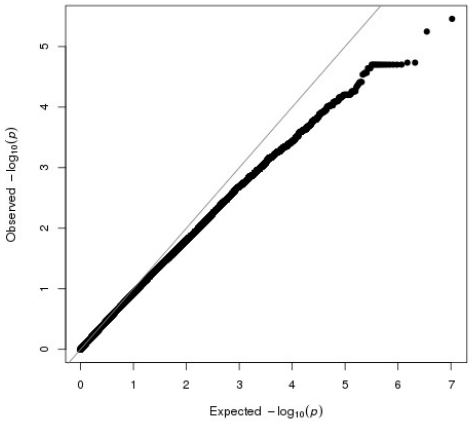
AA



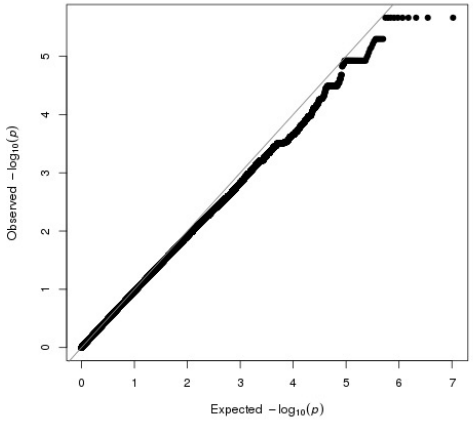
aLA



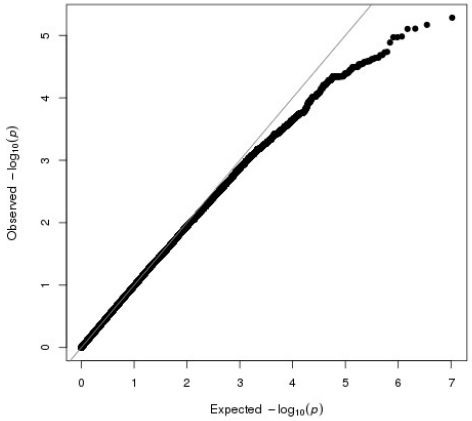
LOX1



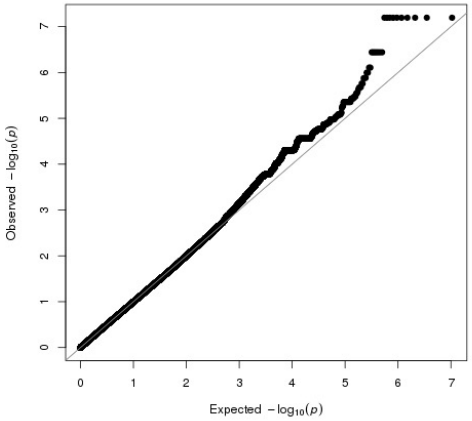
CYP450



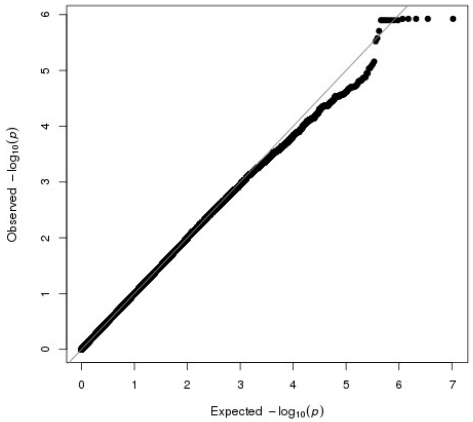
DHA



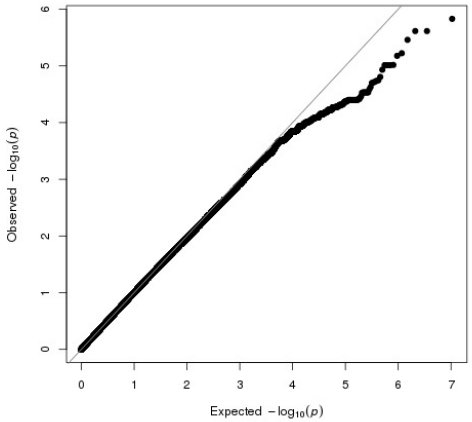
11,12-DHET



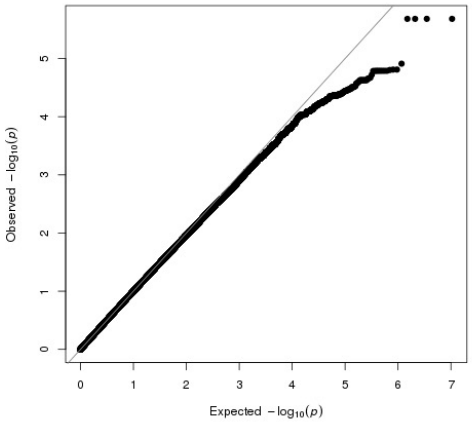
14,15-DHET



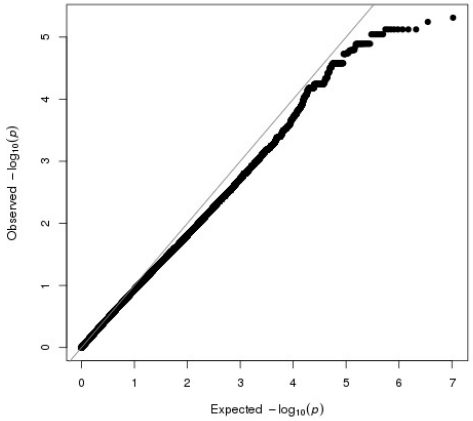
19,20-DiHDPA



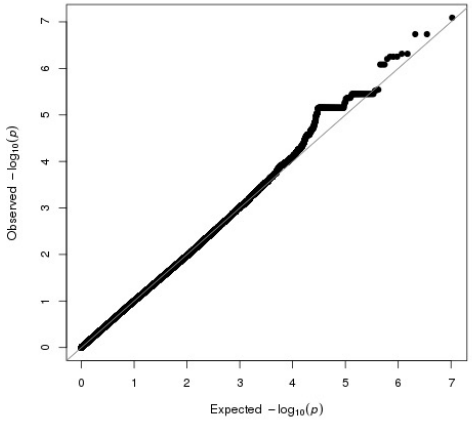
9,10-DiHOME



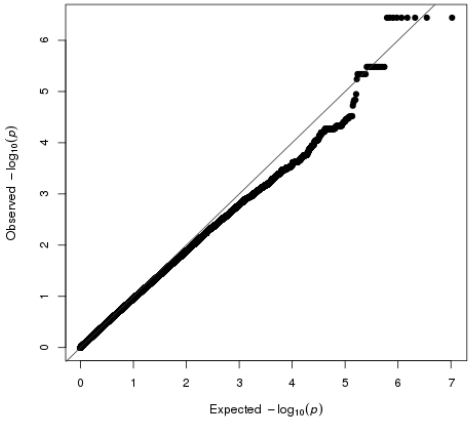
12,13-DiHOME



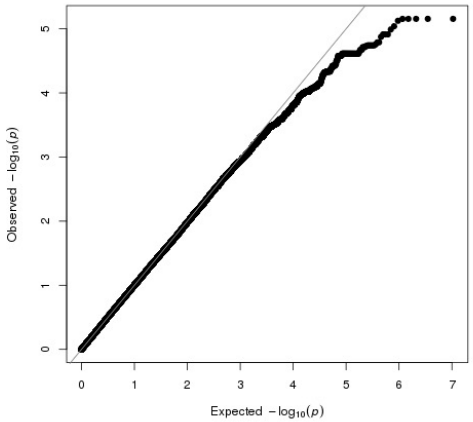
12,13-EpOME



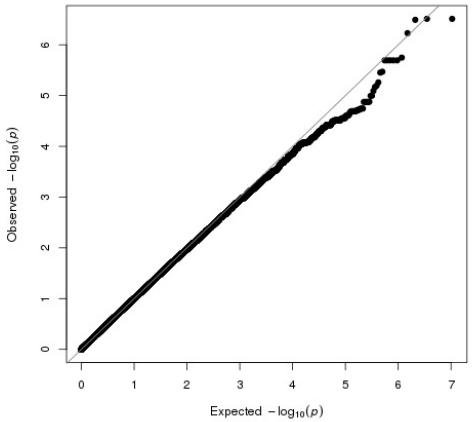
Omega-3



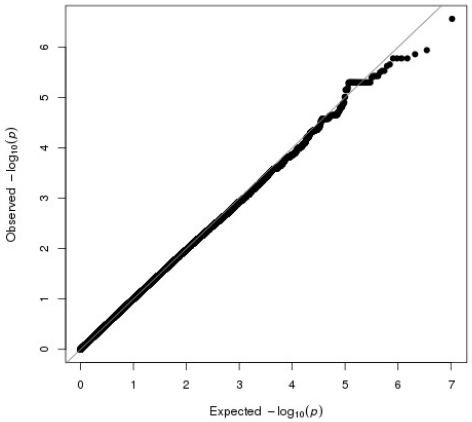
9,10-EpOME



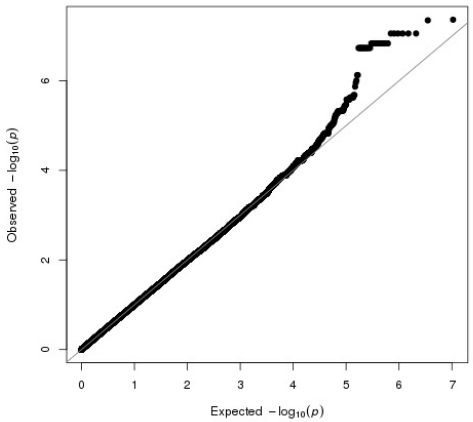
4-HDHA



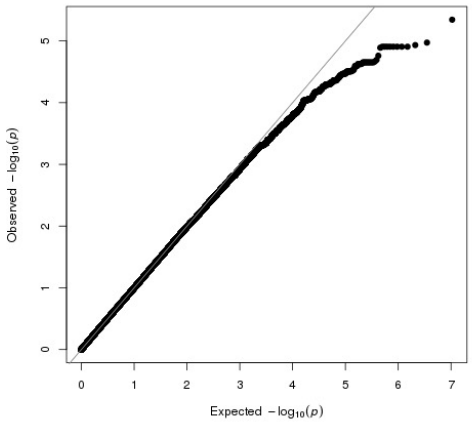
11-HETE



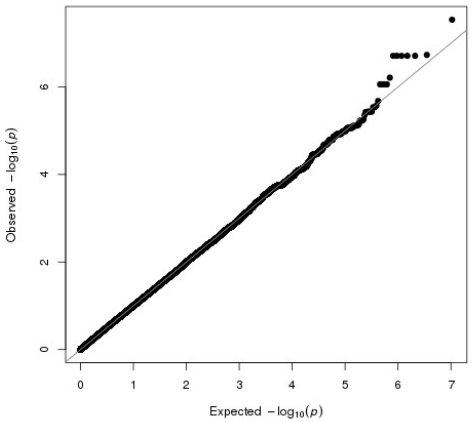
5-HETE



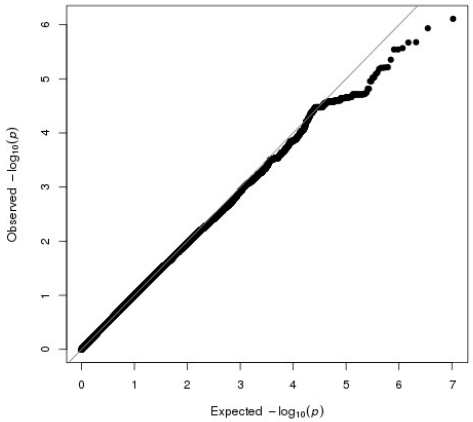
12-HETE



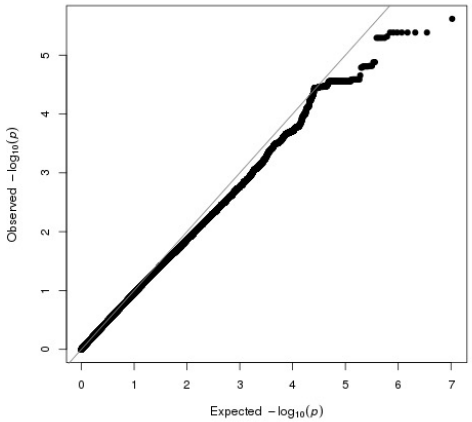
15-HETE



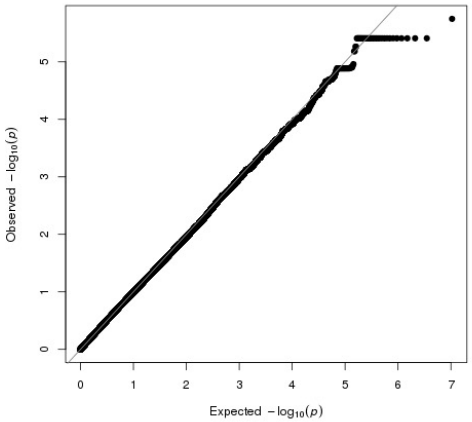
13-HODE



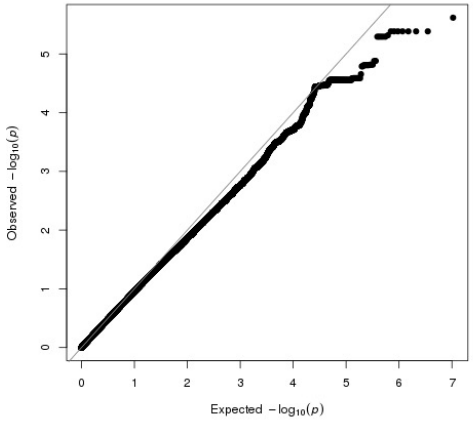
9-HODE



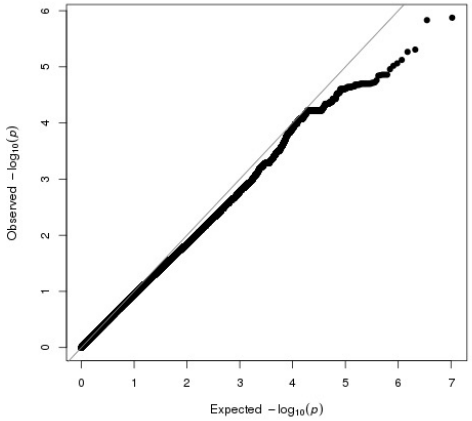
13-HOTrE



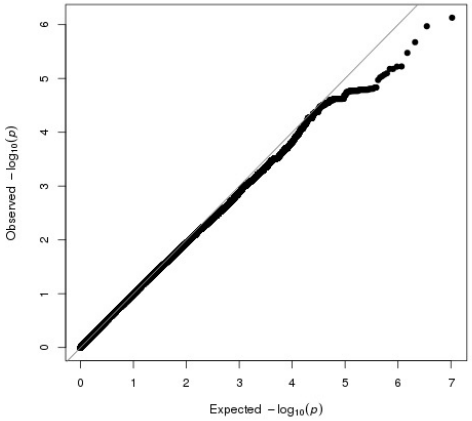
9-HOTrE



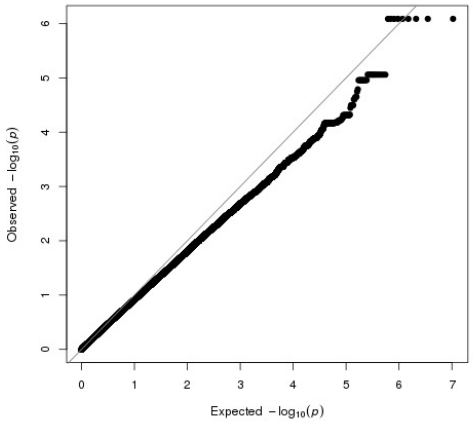
LA



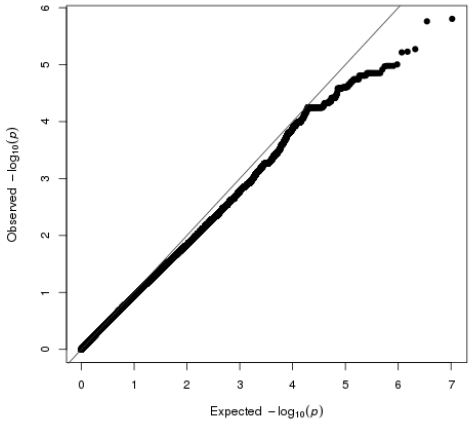
15-LOX



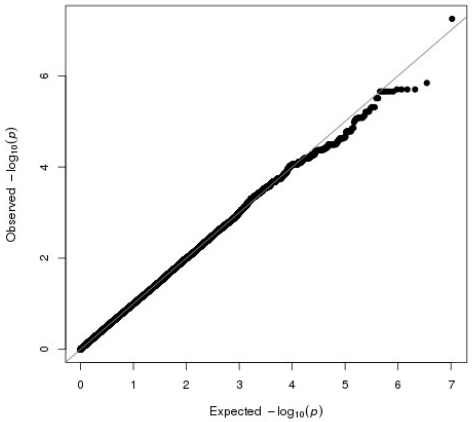
5-LOX



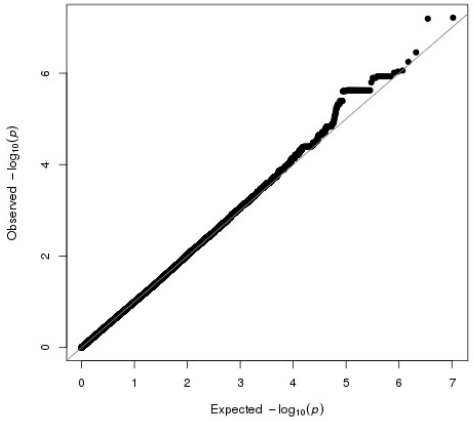
Omega-6



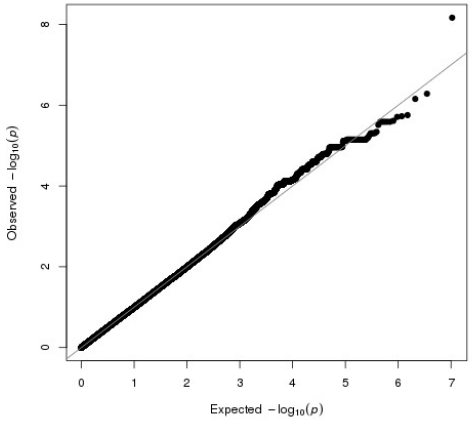
TransEKODE



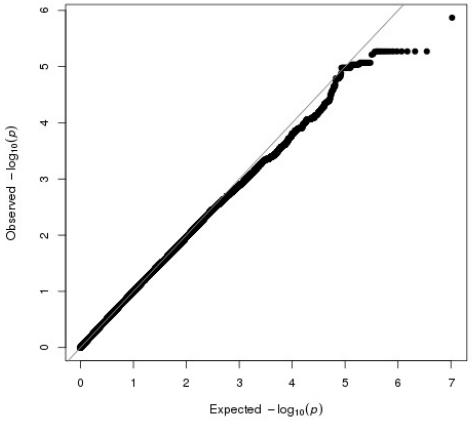
13-OXHO



9-OXHO



13-OxoODE



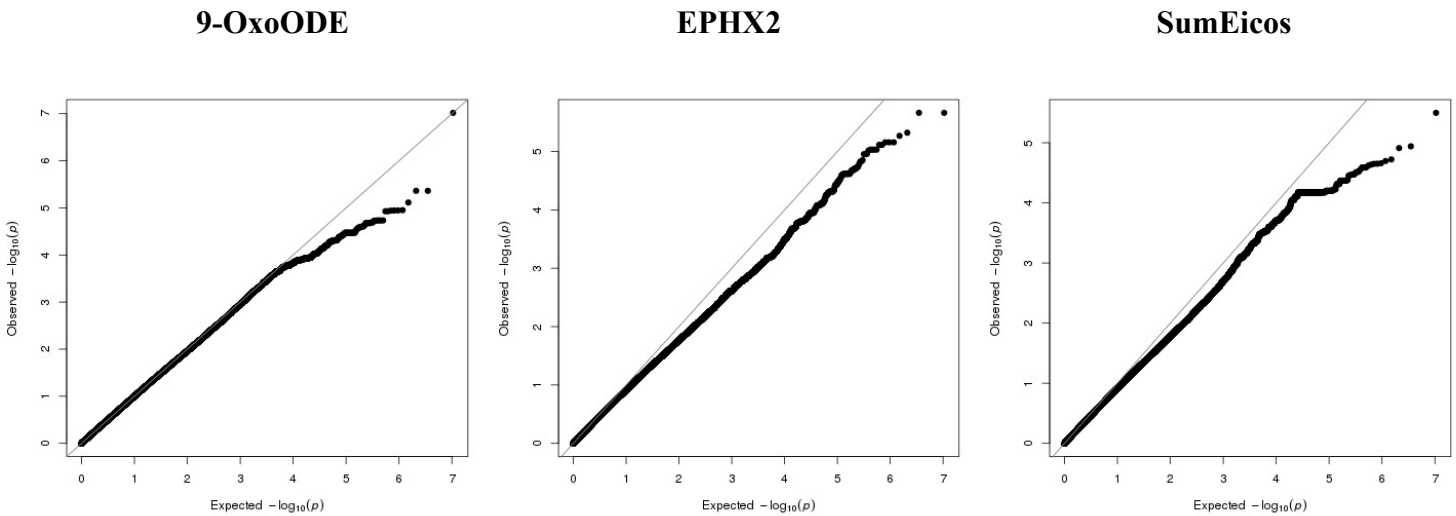


Figure 0-1: Quantile-Quantile plots of each Eico and related species and traits that underwent GWAS analysis

These results depict the genomic inflation factors (GIFs) described in Table 4.3. The plot depicts the expected versus observed P-values of association achieved at GWAS.

Table 0.6: Explanation of each lipid species and description of the measures

The concentrations are provided before adjustment or outlier removal for a total of 999 samples analysed from 196 families.

Lipid	Class	N	Mean	SD	Description
A(22)S(18)	CER	999	1.640	0.741	Alpha-hydroxy fatty acid and sphingosine base
A(24)S(18)	CER	999	2.912	0.498	Alpha-hydroxy fatty acid and sphingosine base
A(26)S(18)	CER	885	0.111	0.092	Alpha-hydroxy fatty acid and sphingosine base
C18DS	CER	999	0.284	0.183	Sphinganine base / Dihydrosphingosine base
C18S	CER	993	2.069	1.937	Sphingosine base
C18S1P	CER	987	3.973	4.604	Sphingosine 1-phosphate
N(16)S(18)	CER	999	1.591	1.247	Non-hydroxy fatty acid and sphingosine base
N(20)S(18)	CER	999	0.401	0.274	Non-hydroxy fatty acid and sphingosine base
N(22)DS(18)	CER	999	0.522	0.378	Non-hydroxy fatty acid and dihydrosphingosine base
N(22)S(18)	CER	999	5.616	3.266	Non-hydroxy fatty acid and sphingosine base
N(22)S(19)	CER	999	1.058	0.733	Non-hydroxy fatty acid and sphingosine base
N(23)S(18)	CER	999	40.522	17.172	Non-hydroxy fatty acid and sphingosine base
N(23)S(20)	CER	999	1.959	0.624	Non-hydroxy fatty acid and sphingosine base
N(24)DS(18)	CER	999	7.855	5.043	Non-hydroxy fatty acid and dihydrosphingosine base
N(24)DS(19)	CER	999	2.673	1.696	Non-hydroxy fatty acid and dihydrosphingosine base
N(24)DS(20)	CER	998	1.422	0.783	Non-hydroxy fatty acid and dihydrosphingosine base
N(24)S(16)	CER	999	1.855	1.148	Non-hydroxy fatty acid and sphingosine base
N(24)S(17)	CER	999	10.692	4.792	Non-hydroxy fatty acid and sphingosine base
N(24)S(18)	CER	999	128.396	60.974	Non-hydroxy fatty acid and sphingosine base
N(24)S(19)	CER	999	50.217	22.929	Non-hydroxy fatty acid and sphingosine base
N(24)S(20)	CER	999	11.359	4.256	Non-hydroxy fatty acid and sphingosine base
N(24)S(22)	CER	999	1.686	1.165	Non-hydroxy fatty acid and sphingosine base

N(25)DS(18)	CER	999	1.068	0.555	Non-hydroxy fatty acid and dihydrosphingosine base
N(25)S(20)	CER	999	1.337	0.668	Non-hydroxy fatty acid and sphingosine base
N(26)DS(18)	CER	999	0.775	0.412	Non-hydroxy fatty acid and dihydrosphingosine base
N(26)S(18)	CER	999	33.066	10.278	Non-hydroxy fatty acid and sphingosine base
N(26)S(19)	CER	999	4.642	3.185	Non-hydroxy fatty acid and sphingosine base
N(27)S(18)	CER	999	2.154	1.540	Non-hydroxy fatty acid and sphingosine base
N(28)S(18)	CER	999	0.877	0.545	Non-hydroxy fatty acid and sphingosine base
N(29)S(18)	CER	998	1.206	1.379	Non-hydroxy fatty acid and sphingosine base
ratio16to24	CER	999	0.016	0.025	Ratio of N(16)S(18)/N(24)S(18) investigated in literature
ratio22to24	CER	999	0.043	0.015	Ratio of N(22)S(18)/N(24)S(18) investigated in literature
ratio20to24	CER	999	0.003	0.001	Ratio of N(20)S(18)/N(24)S(18) investigated in literature
biomcers	CER	999	176.526	79.617	Sum of all N(16)S(18), N(20)S(18), N(22)S(18), N(23)S(18), N(24)S(18) assessed in literature as biomarkers
ns_sum	CER	998	298.596	102.385	Sum of all CER[NS] species
nds_sum	CER	998	14.322	7.579	Sum of all CER[NDS] species
s18_sum	CER	885	211.068	80.080	Sum of all species with a sphingosine backbone
N_s18sum	CER	998	213.764	81.121	Sum of all CER[NS] species with a sphingsine backbone
allxs18	CER	999	230.755	84.767	Sum of all species with a 18-carbon backbone (e.g. incl CER[NDS])
s19_sum	CER	999	55.917	25.129	Sum of all species with a 19-carbon sphingosine backbone
alls19	CER	999	58.590	26.173	Sum of all species with a 19-carbon backbone (e.g. incl CER[NDS])
s20_sum	CER	999	14.654	5.040	Sum of all species with a 20-carbon sphingosine backbone
alls20	CER	999	16.074	5.621	Sum of all species with a 20-carbon backbone (e.g. incl CER[NDS])
ds18_sum	CER	999	10.220	5.924	Sum of all CER[NDS] soecies with a sphingosine backbone
n22_sum	CER	999	7.196	4.109	Sum of all species with a 22-carbon fatty acid
n23_sum	CER	999	42.481	17.458	Sum of all species with a 23-carbon fatty acid
n24_sum	CER	998	215.874	84.249	Sum of all species with a 24-carbon fatty acid
n25_sum	CER	999	2.404	0.874	Sum of all species with a 25-carbon fatty acid

n26_sum	CER	999	38.483	12.864	Sum of all species with a 26-carbon fatty acid
n22ratio	CER	999	12.039	5.162	Ratio of N(22)S(18)/N(22)DS(18)
n24ratio	CER	999	19.653	11.463	Ratio of N(24)S(18)/N(24)DS(18)
n24s19ratio	CER	999	22.768	11.509	Ratio of N(24)S(19)/N(24)DS(19)
n24s20ratio	CER	998	9.353	4.036	Ratio of N(24)S(20)/N(24)DS(20)
n26ratio	CER	999	51.094	25.636	Ratio of N(26)S(18)/N(26)DS(18)
c18s1psratio	CER	984	2.386	2.432	Ratio of C18S1P/C18S
ndssumc18dsratio	CER	998	1123.003	2903.862	Ratio of CER[NDS]/C18DS
c18snsratio	CER	992	0.008	0.007	Ratio of C18S/CER[NS]
totalsphingo	CER	999	323.863	106.343	Sum of all 30 sphingolipid species
assum	CER	999	4.651	1.135	Sum of all 3 CER[AS] species
assumc18dsratio	CER	999	321.150	752.664	Ratio of CER[AS]/C18DS
nssumndssumratio	CER	997	24.275	11.620	Ratio of CER[NS]/CER[NDS]
nssum_c18sratio	CER	992	1560.549	3548.246	Ratio of CER[NS]/C18S
c18s_nssumratio	CER	992	0.008	0.007	Ratio of C18S/CER[NS]
ns18sumc18sratio	CER	992	1140.816	2608.995	Ratio of CER[NS] with a sphingosine backbone/C18S
c18sns18sumratio	CER	992	0.011	0.010	Ratio of C18S/CER[NS] with a sphingsine backbone
c18sc18s1pratio	CER	984	1.072	1.718	Ratio of C18S/C18S1P
sumn22s	CER	999	6.674	3.816	Sum of all CER[N(22)S(X)]
sumn24s	CER	999	204.206	81.146	Sum of all CER[N(24)S(X)]
sumn24ds	CER	999	11.948	6.645	Sum of all CER[N(24)DS(X)]
sumn26s	CER	999	37.708	12.649	Sum of all CER[N(26)S(X)]
sumcer	CER	997	312.634	105.857	Sum of all CER[NS] and CER[NDS] species
sumc18	CER	999	6.265	5.924	Sum of all C18 species
AEA	NAE	998	351.925	334.338	anandamide/N-arachidonoyl ethanolamide
DHEA	NAE	996	349.403	289.978	N-docosahexaenoyl ethanolamide

DPEA	NAE	990	21.565	17.458	N-docosapentaenoyl ethanolamine
HEA	NAE	968	24.067	18.573	N-heptadecanoyl ethanolamide
LEA	NAE	999	618.993	510.646	N-linoleoyl ethanolamide
OEA	NAE	999	567.691	530.576	N-oleoyl ethanolamide
PEA	NAE	999	1883.906	1355.995	N-palmitoyl ethanolamide
POEA	NAE	970	42.821	55.129	N-palmitoleoyl ethanolamide
PDEA	NAE	957	34.494	26.186	N-pentadecanoyl ethanolamide
STEA	NAE	999	494.929	446.851	N-stearoyl ethanolamide
VEA	NAE	999	252.396	258.828	vaccinoyl ethanolamide
sumEA	NAE	999	4637.154	3446.031	Sum of all 11 NAE species

Table 0.7: Predictors identified from stepwise-multiple linear regression of NAE and CER species in 999 plasma samples

Depicted are the predictors (Pr) identified by stepwise multiple linear regression, the coefficient of each predictor (C), and the P-value (P). The class (Cl) of each species is as follows; N, NAE; C, CER. The predictors are as follows; a, age at enrolment; a2, age^2; qc, quality control sample measure specific to each species; B, the mass spectrometry batch; bp, hypertension status; i, trait for sample abnormality; c, cholesterol; s, sex; b, BMI. To fit the table on the page, coefficients are depicted as whole number and P-values are summarised to the nearest two decimal places.

Lipid	Cl	adjR2	Pr	C	P	Pr	C	P	Pr	C	P	Pr	C	P	Pr	C	P	Pr	C	P	Pr	C	P	Pr	C	P	Pr	C	P
DHEA	N	0.45	qc	1	0.00	c	18	0.00	a2	0	0.00	B	3	0.00															
AEA	N	0.38	qc	1	0.00	c	18	0.00	b	2	0.14	B	-3	0.04															
DPEA	N	0.59	qc	1	0.00	c	1	0.01	b	0	0.23	B	0	0.00	s	-2	0.03	bp	0	0.48	a2	0	0.28	a	0	0.19	i	1	0.13
HEA	N	0.56	qc	1	0.00	c	1	0.00	B	0	0.00																		
OEA	N	0.64	qc	1	0.00	c	15	0.06	B	12	0.00	b	3	0.16															
PEA	N	0.68	qc	1	0.00	c	73	0.00																					
STEA	N	0.43	qc	1	0.00	c	26	0.00	B	7	0.00	s	-39	0.07	b	3	0.14												
LEA	N	0.64	qc	1	0.00	c	21	0.01	B	8	0.00	a2	0	0.00															
VEA	N	0.52	qc	1	0.00	c	9	0.05	B	3	0.01	s	17	0.14	bp	-11	0.25												
POEA	N	0.29	qc	0	0.00	i	0	0.93	s	20	0.00	B	1	0.04															
PDEA	N	0.56	qc	1	0.00	s	3	0.00	B	0	0.01	bp	-1	0.12	b	0	0.26	c	1	0.00									
A22_S18	C	0.36	qc	1	0.00	B	0	0.00	a2	0	0.00	a	0	0.00															
A24_S18	C	0.44	qc	1	0.00	B	0	0.00																					
A26_S18	C	0.15	qc	1	0.00	s	0	0.00	b	0	0.02	c	0	0.00															
C18_DS	C	0.72	qc	1	0.00	B	0	0.02	i	0	0.13	a2	0	0.03	a	0	0.14	s	0	0.04	c	0	0.00	b	0	0.00	bp	0	0.29
C18_S	C	0.46	qc	0	0.00	B	0	0.00	i	0	0.02	a2	0	0.00	a	0	0.01	s	0	0.24	c	0	0.32	b	0	0.01	bp	0	0.43
C18_S1P	C	0.37	qc	0	0.00	B	0	0.00	i	1	0.01	a2	0	0.12															
N16_S18	C	0.36	qc	1	0.00	i	0	0.25	s	0	0.01	a	0	0.00	a2	0	0.03	b	0	0.00	c	0	0.00						
N20_S18	C	0.40	qc	1	0.00	B	0	0.03	bp	0	0.19	a	0	0.00	a2	0	0.00	c	0	0.00									

N22_DS18	C	0.24	qc	0	0.00	B	0	0.00	a	0	0.00																	
N22_S18	C	0.35	qc	1	0.00	B	0	0.00	i	0	0.26	a	0	0.00	s	0	0.02	bp	0	0.81	a2	0	0.02						
N22_S19	C	0.24	qc	1	0.00	B	0	0.00	i	0	0.01	c	0	0.02															
N23_S18	C	0.35	qc	0	0.00	B	-1	0.00	a2	0	0.00	a	-1	0.00															
N23_S20	C	0.06	qc	0	0.00	B	0	0.00	s	0	0.00	a2	0	0.00															
N24_DS18	C	0.26	qc	1	0.00	B	0	0.00	s	-1	0.00	a	0	0.01	b	0	0.00	c	0	0.00									
N24_DS19	C	0.18	qc	1	0.00	B	0	0.00	s	0	0.00	b	0	0.00	c	0	0.00												
N24_DS20	C	0.07	qc	0	0.00	B	0	0.01	i	0	0.90	s	0	0.00	bp	0	0.64	a	0	0.33	a2	0	0.50	b	0	0.00	c	0	0.00
N24_S16	C	0.33	qc	1	0.00	B	0	0.03	i	0	0.10	s	0	0.00	a	0	0.08	a2	0	0.14									
N24_S17	C	0.27	qc	0	0.00	B	0	0.00	s	-1	0.00	a	0	0.01	a2	0	0.06												
N24_S18	C	0.48	qc	1	0.00	B	-2	0.00	a	-2	0.00	a2	0	0.03	c	-2	0.05												
N24_S19	C	0.14	qc	1	0.00	B	-1	0.00	i	2	0.05																		
N24_S20	C	0.10	qc	0	0.00	B	0	0.15	i	0	0.20	s	-2	0.00	bp	0	0.57	a	0	0.09	a2	0	0.17	b	0	0.01	c	0	0.00
N24_S22	C	0.11	qc	0	0.00	s	0	0.01	a2	0	0.00	b	0	0.00															
N25_DS18	C	0.17	qc	1	0.00	B	0	0.00	a	0	0.26	b	0	0.00	c	0	0.00												
N25_S20	C	0.12	qc	0	0.00	a	0	0.00	b	0	0.00	c	0	0.00															
N26_DS18	C	0.06	qc	0	0.00	b	0	0.00	c	0	0.00																		
N26_S18	C	0.15	qc	0	0.00	B	0	0.00	s	2	0.00	a	0	0.31	a2	0	0.24	c	1	0.00									
N26_S19	C	0.09	qc	0	0.00	s	0	0.01	a2	0	0.01	b	0	0.01	c	0	0.00												
N27_S18	C	0.18	qc	1	0.00	i	0	0.14	s	1	0.00	a	0	0.00	a2	0	0.02	b	0	0.02	c	0	0.01						
N28_S18	C	0.20	qc	0	0.00	s	0	0.00	b	0	0.00																		
N29_S18	C	0.13	qc	1	0.00	B	0	0.73	i	0	0.13	s	1	0.00	bp	0	0.67	a	0	0.01	a2	0	0.01	b	0	0.00	c	0	0.40
ratio16to24	C	0.54	qc	1	0.00	a	0	0.00	b	0	0.00	B	0	0.11															
ratio22to24	C	0.21	qc	1	0.00																								
ratio20to24	C	0.40	qc	1	0.00	bp	0	0.00	a	0	0.00	a2	0	0.00	b	0	0.41	c	0	0.00	B	0	0.14	i	0	0.44			
biomcers	C	0.45	qc	1	0.00	a	-2	0.00	a2	0	0.02	c	-3	0.03	B	-3	0.00												

ns_sum	C	0.39	qc	1	0.00	a	-1	0.00	B	-4	0.00																	
nds_sum	C	0.23	qc	1	0.00	s	-1	0.00	a	0	0.02	b	0	0.00	c	1	0.00	B	0	0.00	i	0	0.27						
s18_sum	C	0.47	qc	1	0.00	bp	-3	0.35	a	-2	0.02	a2	0	0.11	c	-2	0.34	B	-3	0.00									
N_s18sum	C	0.45	qc	1	0.00	a	-3	0.00	a2	0	0.01	B	-3	0.00															
allxs18	C	0.44	qc	1	0.00	a	-3	0.00	a2	0	0.01	B	-3	0.00															
s19_sum	C	0.13	qc	0	0.00	s	-5	0.00	bp	-1	0.32	a	-1	0.01	a2	0	0.02	c	1	0.03	B	-1	0.00						
alls19	C	0.14	qc	0	0.00	s	-5	0.00	bp	-1	0.34	a	-1	0.01	a2	0	0.02	b	0	0.59	c	2	0.01	B	-1	0.00			
s20_sum	C	0.08	qc	0	0.00	s	-2	0.00	bp	0	0.69	a	0	0.11	a2	0	0.15	b	0	0.02	c	0	0.00	B	0	0.10			
alls20	C	0.07	qc	0	0.00	s	-2	0.00	bp	0	0.75	a	0	0.13	a2	0	0.18	b	0	0.01	c	1	0.00	B	0	0.11			
ds18_sum	C	0.23	qc	1	0.00	s	-1	0.02	a	0	0.04	b	0	0.00	c	0	0.00	B	0	0.00	i	0	0.27						
n22_sum	C	0.35	qc	1	0.00	s	0	0.04	a	0	0.00	a2	0	0.02	c	0	0.08	B	0	0.00	i	0	0.13						
n23_sum	C	0.34	qc	0	0.00	a	-1	0.00	a2	0	0.00	B	-1	0.00															
n24_sum	C	0.42	qc	1	0.00	s	-13	0.00	bp	-4	0.28	a	-2	0.01	a2	0	0.08	b	0	0.47	c	0	0.95	B	-3	0.00			
n25_sum	C	0.08	qc	0	0.00	bp	0	0.08	a2	0	0.00	c	0	0.00															
n26_sum	C	0.15	qc	0	0.00	s	2	0.00	a2	0	0.10	c	2	0.00	B	0	0.00												
n22ratio	C	0.13	qc	1	0.00	b	0	0.00																					
n24ratio	C	0.25	qc	1	0.00	b	0	0.00	c	-1	0.00																		
n24s19ratio	C	0.15	qc	1	0.00	s	0	0.46	a	0	0.00	a2	0	0.00	b	0	0.00	c	-2	0.00	i	0	0.43						
n24s20ratio	C	0.11	qc	1	0.00	s	-1	0.00	a2	0	0.11	b	0	0.04	c	-1	0.00												
n26ratio	C	0.04	qc	1	0.00	b	-1	0.00	c	-2	0.00	B	0	0.07	i	2	0.26												
c18s1psratio	C	0.15	qc	1	0.00	bp	0	0.02	a2	0	0.02	b	0	0.02	B	0	0.00												
ndssumc18dsratio	C	0.58	qc	1	0.00	B	-27	0.03																					
c18snsratio	C	0.43	qc	0	0.00	s	0	0.74	bp	0	0.03	a	0	0.03	a2	0	0.00	b	0	0.00	c	0	0.08	B	0	0.01	i	0	0.06
totalsphingo	C	0.38	qc	1	0.00	a	-1	0.00	B	-4	0.00																		
assum	C	0.40	qc	1	0.00	bp	0	0.10	a	0	0.01	a2	0	0.03	b	0	0.30	B	0	0.00									

assumc18dsratio	C	0.68	qc	2	0.00																							
nssumndssumratio	C	0.18	qc	1	0.00	b	0	0.00	c	-2	0.00																	
nssum_c18sratio	C	0.76	qc	2	0.00	a	-13	0.00	B	-37	0.00																	
c18s_nssumratio	C	0.43	qc	0	0.00	s	0	0.74	bp	0	0.03	a	0	0.03	a2	0	0.00	b	0	0.00	c	0	0.08	B	0	0.01	i	0	0.06
ns18sumc18sratio	C	0.75	qc	1	0.00	a	-10	0.00	B	-28	0.00																	
c18sns18sumratio	C	0.45	qc	0	0.00	s	0	0.57	bp	0	0.03	a	0	0.02	a2	0	0.00	b	0	0.00	c	0	0.03	B	0	0.01	i	0	0.05
c18sc18s1pratio	C	0.07	qc	0	0.00	B	0	0.00																				
sumn22s	C	0.35	qc	1	0.00	s	0	0.04	a	0	0.00	a2	0	0.02	c	0	0.05	B	0	0.00	i	0	0.13						
sumn24s	C	0.39	qc	1	0.00	s	-12	0.00	bp	-5	0.17	a	-2	0.02	a2	0	0.14	b	0	0.71	c	-1	0.51						
sumn24ds	C	0.26	qc	1	0.00	s	-1	0.00	a	0	0.01	b	0	0.00	c	1	0.00	B	0	0.00	i	0	0.20						
sumn26s	C	0.15	qc	0	0.00	s	2	0.00	a2	0	0.11	c	2	0.00	B	0	0.00												
sumcer	C	0.40	qc	1	0.00	a	-1	0.00	B	-4	0.00																	
sumc18	C	0.54	qc	0	0.00	a2	0	0.02	B	0	0.00	i	1	0.00														
sumEA	N	0.04	bp	-652	0.00	a	-79	0.09	a2	1	0.22	b	53	0.01	c	255	0.00	B	43	0.01	i	-348	0.05						

Table 0.8: Genomic inflation factors (GIF) from GWAS results per trait.

Lipid	GIF
A(22)S(18)	1.003741383
A(24)S(18)	0.999334978
A(26)S(18)	1.008387084
C18DS	0.994580375
C18S	1.005949886
C18S1P	0.992881657
N(16)S(18)	1.007964953
N(20)S(18)	0.997653631
N(22)DS(18)	1.00814317
N(22)S(18)	1.004180929
N(22)S(19)	0.98934249
N(23)S(18)	1.00002488
N(23)S(20)	1.004615938
N(24)DS(18)	1.004428821
N(24)DS(19)	0.995193546
N(24)DS(20)	1.004494309
N(24)S(16)	1.001223708
N(24)S(17)	1.003722682
N(24)S(18)	1.003605806
N(24)S(19)	0.994436413
N(24)S(20)	0.990735753
N(24)S(22)	1.00362918
N(25)DS(18)	0.991657661
N(25)S(20)	1.006076327
N(26)DS(18)	0.996867225
N(26)S(18)	1.00172314
N(26)S(19)	0.99389785
N(27)S(18)	0.995802343
N(28)S(18)	0.993234223
N(29)S(18)	1.005617448
N_s18sum	0.999046068
alls19	0.995551355
alls20	0.995170315
allxs18	0.999460813
assum	1.01290726
assumc18dsratio	0.980548393
biomcers	0.998230925
c18s1psratio	1.000099485
c18s_nssumratio	0.994027828
c18sc18s1pratio	0.998985498
c18sns18sumratio	0.994566443
c18snsratio	0.994027828

ds18_sum	1.001293711
n22_sum	1.004714185
n22ratio	0.990513476
n23_sum	0.998202986
n24_sum	1.004012576
n24ratio	0.997532616
n24s19ratio	0.99368434
n24s20ratio	1.000953064
n25_sum	0.995923205
n26_sum	1.00039328
n26ratio	1.010175533
nds_sum	1.002414242
ndssumc18dsratio	0.982623038
ns18sumc18sratio	0.99284455
ns_sum	1.001228375
nssum_c18sratio	0.983958745
nssumndssumratio	1.004166899
ratio16to24	0.995100624
ratio20to24	0.994264607
ratio22to24	0.997299925
s18_sum	0.998119172
s19_sum	0.995411938
s20_sum	0.994478207
sumc18	1.007313249
sumcer	1.003283281
sumn22s	1.000509883
sumn24ds	1.004587869
sumn24s	1.006891446
sumn26s	0.998892319
totalsphingo	1.002049969
AEA	0.998417201
DHEA	0.991569613
DPEA	1.006165311
HEA	0.995286474
LEA	0.990958066
OEA	0.991564979
PEA	0.990402351
POEA	0.998449802
PDEA	1.000010892
STEA	0.998440487
VEA	0.992093352
sumEA	0.996076622

Table 0.9: The significant GWAS associations for the *N*-acyl ethanolamine species

A) Description of the GWAS significant associations identified. B) Description of the SNPs using the Ensemble API Client. C) Summary of the information on eQTL status as identified using the GTEX browser, including a section specifying whole blood only, GWAS Catalog information, and Gene Atlas PheWAS.

Table S5A									
Lipid	**Chr**	**SNP**	**Position**	**A1**	**A2**	**MAF**	**Beta**	**SE**	**P-value**
DHEA	1	rs324420	46870761	A	C	0.20202	0.296383	0.052927	2.14536E-08
DHEA	1	rs324422	46886782	T	C	0.230303	0.27969	0.0503418	2.76291E-08
DHEA	1	rs324418	46872698	G	A	0.219192	0.282135	0.0515253	4.35902E-08
LEA	1	rs324420	46870761	A	C	0.204728	0.314943	0.0549789	1.01361E-08
LEA	1	rs1571138	46895641	A	G	0.203219	0.3078	0.0550383	2.23871E-08
LEA	1	rs324422	46886782	T	C	0.232897	0.292222	0.0523116	2.3211E-08
PEA	1	rs324420	46870761	A	C	0.205438	0.297069	0.0508888	5.29544E-09
PEA	1	rs1571138	46895641	A	G	0.203927	0.290943	0.0509469	1.12505E-08
PEA	1	rs324418	46872698	G	A	0.222558	0.274844	0.0495773	2.96045E-08
PEA	1	rs324422	46886782	T	C	0.233635	0.267431	0.0484694	3.43796E-08
VEA	1	rs324420	46870761	A	C	0.204728	0.314814	0.0489248	1.2375E-10
VEA	1	rs1571138	46895641	A	G	0.203219	0.315102	0.04898	1.24879E-10
VEA	1	rs324418	46872698	G	A	0.221831	0.29006	0.0476481	1.15E-09
VEA	1	rs11584511	46892811	T	C	0.135815	0.340113	0.0575152	3.35E-09
VEA	1	rs324422	46886782	T	C	0.232897	0.272851	0.0465707	4.66E-09
VEA	1	rs10489770	46807597	A	G	0.136318	0.316454	0.0577336	4.22E-08
VEA	1	rs72677586	46813848	A	G	0.136318	0.316454	0.0577336	4.22E-08
VEA	1	rs10890392	46852061	G	A	0.136318	0.316454	0.0577336	4.22E-08
sumEA	1	rs324420	46870761	A	C	0.205231	0.319134	0.0526643	1.36E-09
sumEA	1	rs1571138	46895641	A	G	0.203722	0.313912	0.0527248	2.62E-09
sumEA	1	rs324418	46872698	G	A	0.222334	0.288648	0.0513043	1.84E-08
sumEA	1	rs324422	46886782	T	C	0.2334	0.281015	0.0501513	2.10E-08

Table S5B							
SNPID	**Associated Gene ID**	**Associated Transcript ID**	**Associated Gene Name**	**Associated Gene Type**	**Impact Rating**	**Variant Allele**	**Consequence Terms**
rs324420	ENSG00000117480	ENST00000243167	FAAH	Protein coding	MODERATE	A	Missense variant
rs324422	(Intergenic)	(Intergenic)	(Intergenic)	(Intergenic)	MODIFIER	T	Intergenic variant
rs324418	ENSG00000117480	ENST00000243167	FAAH	Protein coding	MODIFIER	G	Intron variant
rs1571138	ENSG00000232022	ENST00000446499	FAAHP1	pseudogene	MODIFIER	G	upstream

rs11584511	ENSG00000232022	ENST00000446499	FAAHP1	pseudogene	MODIFIER	T	upstream
rs10489770	ENSG00000117481	ENST00000307089	NSUN4	NMD	MODIFIER	A	intron_variant
rs72677586	ENSG00000117481	ENST00000307089	NSUN4	NMD	MODIFIER	A	intron_variant
rs10890392	(Intergenic)	(Intergenic)	(Intergenic)	(Intergenic)	MODIFIER	G	intergenic_variant

Table S5C						
SNPID	**eQTL**	**Tissue**	**Other**	**GTEx whole blood**	**GWAS Catalog**	**Gene Atlas**
rs324420	FAAH	Multiple	FAAHP1, LURAP1, NSUN4, RAD54L, MKNK1	FAAH, NSUN4	X	X
rs324422	FAAH	Multiple	FAAHP1, LURAP1, NSUN4, RAD54L, MKNK1, MOB3C	FAAH, NSUN4, MOB3C	X	X
rs324418	FAAH	Multiple	FAAHP1, LURAP1, NSUN4, RAD54L, MKNK1	FAAH, NSUN4	X	X
rs1571138	FAAH	Multiple	FAAHP1, LURAP1, NSUN4, RAD54L, MKNK1	FAAH, NSUN4	X	X
rs11584511	FAAH	Multiple	FAAHP1, LURAP1, NSUN4, RAD54L, MKNK1, UQCRH	FAAH, NSUN4	X	X
rs10489770	FAAH	Multiple	FAAHP1, LURAP1, NSUN4, RAD54L, MKNK1, UQCRH	FAAH	X	X
rs72677586	FAAH	Multiple	FAAHP1, LURAP1, NSUN4, RAD54L, MKNK1, UQCRH	FAAH	X	X
rs10890392	FAAH	Multiple	FAAHP1, LURAP1, NSUN4, RAD54L, MKNK1, UQCRH	FAAH	X	X

Table 0.10: The significant GWAS associations for the ceramides and related sphingolipid species

Description of the GWAS significant associations identified.

Lipid	Chr	SNP	Position	A1	A2	MAF	Beta	SE	P-value
N22S19	20	rs438568	12958687	A	G	0.364919	0.374657	0.0435367	7.59475E-18
N22S19	20	rs1321940	12959885	A	G	0.365927	0.371388	0.0435256	1.43071E-17
N22S19	20	rs364585	12962718	A	G	0.365927	0.371388	0.0435256	1.43071E-17
N22S19	20	rs168622	12966089	T	G	0.365927	0.371388	0.0435256	1.43071E-17
N22S19	20	rs680379	12969400	A	G	0.367944	0.366187	0.0433357	2.91242E-17
N22S19	20	rs686548	12973521	A	T	0.367944	0.366187	0.0433357	2.91242E-17
N22S19	20	rs4814175	12959094	A	T	0.368952	0.360224	0.0435841	1.39645E-16
N22S19	20	rs4814176	12959398	T	C	0.368952	0.360224	0.0435841	1.39645E-16
N22S19	20	rs2327452	12952964	A	C	0.323589	0.337161	0.045523	1.2979E-13
N22S19	20	rs3848746	12950606	A	G	0.324597	0.336772	0.0455756	1.47602E-13
N22S19	20	rs2327451	12953934	C	A	0.324093	0.335983	0.0455395	1.60882E-13
N22S19	20	rs4508668	12955601	T	C	0.324093	0.335983	0.0455395	1.60882E-13
N22S19	20	rs3903703	12945963	A	G	0.323589	0.333638	0.045521	2.31328E-13
N22S19	20	rs4814173	12947532	G	C	0.323589	0.333638	0.045521	2.31328E-13
N22S19	20	rs3848744	12942649	A	G	0.322077	0.332649	0.0455587	2.84478E-13
N22S19	20	rs3843765	12943737	G	A	0.322581	0.331459	0.0455747	3.51932E-13
N22S19	20	rs3848745	12944067	G	A	0.322581	0.331459	0.0455747	3.51932E-13
N22S19	20	rs6041735	12940649	T	C	0.321069	0.331519	0.0456846	3.9667E-13
N22S19	20	rs4813102	12947883	A	T	0.373992	0.303511	0.0445435	9.50448E-12
N23S20	20	rs1321940	12959885	A	G	0.366197	0.317196	0.0481867	4.62161E-11
N23S20	20	rs364585	12962718	A	G	0.366197	0.317196	0.0481867	4.62161E-11
N23S20	20	rs168622	12966089	T	G	0.366197	0.317196	0.0481867	4.62161E-11
N23S20	20	rs680379	12969400	A	G	0.368209	0.315023	0.0479728	5.14434E-11
N23S20	20	rs686548	12973521	A	T	0.368209	0.315023	0.0479728	5.14434E-11
N23S20	20	rs438568	12958687	A	G	0.365191	0.314138	0.0481863	7.06632E-11
N23S20	20	rs4814175	12959094	A	T	0.369215	0.30644	0.0482125	2.07042E-10
N23S20	20	rs4814176	12959398	T	C	0.369215	0.30644	0.0482125	2.07042E-10
N23S20	20	rs2327451	12953934	C	A	0.324447	0.284483	0.0502474	1.49919E-08
N23S20	20	rs4508668	12955601	T	C	0.324447	0.284483	0.0502474	1.49919E-08
N23S20	20	rs2327452	12952964	A	C	0.323944	0.283274	0.0502284	1.70329E-08

N23S20	20	rs3903703	12945963	A	G	0.323944	0.283154	0.0502262	1.72465E-08
N23S20	20	rs4814173	12947532	G	C	0.323944	0.283154	0.0502262	1.72465E-08
N23S20	20	rs3848746	12950606	A	G	0.32495	0.281119	0.050296	2.27993E-08
N23S20	20	rs4813102	12947883	A	T	0.374245	0.272082	0.0490454	2.89704E-08
N24DS19	20	rs1321940	12959885	A	G	0.365694	0.439829	0.0465041	3.14346E-21
N24DS19	20	rs364585	12962718	A	G	0.365694	0.439829	0.0465041	3.14346E-21
N24DS19	20	rs168622	12966089	T	G	0.365694	0.439829	0.0465041	3.14346E-21
N24DS19	20	rs438568	12958687	A	G	0.364688	0.438478	0.0465086	4.18403E-21
N24DS19	20	rs680379	12969400	A	G	0.367706	0.431625	0.0462981	1.13372E-20
N24DS19	20	rs686548	12973521	A	T	0.367706	0.431625	0.0462981	1.13372E-20
N24DS19	20	rs4814175	12959094	A	T	0.368712	0.431491	0.046545	1.85453E-20
N24DS19	20	rs4814176	12959398	T	C	0.368712	0.431491	0.046545	1.85453E-20
N24DS19	20	rs3848746	12950606	A	G	0.324447	0.412431	0.0486694	2.36864E-17
N24DS19	20	rs2327451	12953934	C	A	0.323944	0.403227	0.0486263	1.11007E-16
N24DS19	20	rs4508668	12955601	T	C	0.323944	0.403227	0.0486263	1.11007E-16
N24DS19	20	rs3903703	12945963	A	G	0.323441	0.40222	0.0486078	1.28651E-16
N24DS19	20	rs4814173	12947532	G	C	0.323441	0.40222	0.0486078	1.28651E-16
N24DS19	20	rs2327452	12952964	A	C	0.323441	0.401717	0.0486079	1.40353E-16
N24DS19	20	rs3843765	12943737	G	A	0.321932	0.399228	0.0487272	2.54535E-16
N24DS19	20	rs3848745	12944067	G	A	0.321932	0.399228	0.0487272	2.54535E-16
N24DS19	20	rs6041735	12940649	T	C	0.320423	0.399361	0.0488276	2.86207E-16
N24DS19	20	rs3848744	12942649	A	G	0.321429	0.397723	0.0487092	3.20791E-16
N24DS19	20	rs4813102	12947883	A	T	0.373742	0.343345	0.0474717	4.73709E-13
N24DS19	20	rs6078854	12960153	A	T	0.401408	-0.299991	0.0455804	4.65471E-11
N24DS19	20	rs4544513	12954215	T	C	0.346579	-0.307759	0.0469301	5.46021E-11
N24DS19	20	rs6109637	12954804	T	C	0.346579	-0.307759	0.0469301	5.46021E-11
N24DS19	20	rs382003	12963171	A	G	0.297787	-0.293357	0.0486072	1.58708E-09
N24DS19	20	rs360539	12966440	G	T	0.297284	-0.28859	0.0484437	2.56569E-09
N24DS19	20	rs73079703	12941782	T	C	0.296781	-0.289123	0.0487693	3.05936E-09
N24DS19	20	rs8183164	12942600	A	C	0.296781	-0.289123	0.0487693	3.05936E-09
N24DS19	20	rs6131414	12945669	A	G	0.296781	-0.289123	0.0487693	3.05936E-09
N24DS19	20	rs7272107	12946328	A	G	0.296781	-0.289123	0.0487693	3.05936E-09
N24DS19	20	rs73079713	12947141	T	C	0.296781	-0.289123	0.0487693	3.05936E-09
N24DS19	20	rs6134734	12952640	T	A	0.296278	-0.286445	0.0487463	4.19678E-09
N24DS19	20	rs6109634	12953314	G	A	0.296278	-0.286445	0.0487463	4.19678E-09

N24DS19	20	rs3848748	12957587	C	G	0.294266	-0.286133	0.0488082	4.56212E-09
N24DS19	20	rs3848749	12962089	C	T	0.294266	-0.286133	0.0488082	4.56212E-09
N24DS19	20	rs6131417	12967751	A	G	0.293763	-0.281393	0.0486441	7.26325E-09
N24DS19	20	rs6134740	12968649	C	G	0.293763	-0.281393	0.0486441	7.26325E-09
N24DS19	20	rs6134741	12970842	A	G	0.293763	-0.281393	0.0486441	7.26325E-09
N24DS19	20	rs59131252	12975257	A	C	0.294769	-0.272744	0.0486133	2.01773E-08
N24DS20	20	rs680379	12969400	A	G	0.368077	0.302499	0.0472078	1.47623E-10
N24DS20	20	rs686548	12973521	A	T	0.368077	0.302499	0.0472078	1.47623E-10
N24DS20	20	rs1321940	12959885	A	G	0.366062	0.302453	0.0474188	1.79019E-10
N24DS20	20	rs364585	12962718	A	G	0.366062	0.302453	0.0474188	1.79019E-10
N24DS20	20	rs168622	12966089	T	G	0.366062	0.302453	0.0474188	1.79019E-10
N24DS20	20	rs438568	12958687	A	G	0.365055	0.299704	0.0474216	2.61572E-10
N24DS20	20	rs4814175	12959094	A	T	0.369084	0.299142	0.0474504	2.8947E-10
N24DS20	20	rs4814176	12959398	T	C	0.369084	0.299142	0.0474504	2.8947E-10
N24DS20	20	rs3848746	12950606	A	G	0.324773	0.278741	0.0495759	1.88216E-08
N24DS20	20	rs2327451	12953934	C	A	0.32427	0.276601	0.0495285	2.34122E-08
N24DS20	20	rs4508668	12955601	T	C	0.32427	0.276601	0.0495285	2.34122E-08
N24DS20	20	rs2327452	12952964	A	C	0.323766	0.27492	0.0495103	2.81149E-08
N24DS20	20	rs3903703	12945963	A	G	0.323766	0.274307	0.0495086	3.01452E-08
N24DS20	20	rs4814173	12947532	G	C	0.323766	0.274307	0.0495086	3.01452E-08
N24S16	14	rs7160525	64232220	A	G	0.135081	0.361768	0.0583558	5.66951E-10
N24S16	14	rs17101394	64232386	A	G	0.135081	0.361768	0.0583558	5.66951E-10
N24S16	14	rs8008068	64233717	G	A	0.135081	0.361768	0.0583558	5.66951E-10
N24S16	14	rs8008070	64233720	T	A	0.135081	0.361768	0.0583558	5.66951E-10
N24S16	14	rs8012828	64233980	T	C	0.135081	0.361768	0.0583558	5.66951E-10
N24S16	14	rs34609767	64234034	G	T	0.135081	0.361768	0.0583558	5.66951E-10
N24S16	14	rs4902243	64234243	G	A	0.135081	0.361768	0.0583558	5.66951E-10
N24S16	14	rs7157785	64235556	T	G	0.135081	0.361768	0.0583558	5.66951E-10
N24S16	14	rs34817779	64236003	T	C	0.135081	0.361768	0.0583558	5.66951E-10
N24S16	14	rs35372182	64236157	G	A	0.135081	0.361768	0.0583558	5.66951E-10
N24S16	14	rs12897637	64239351	C	T	0.135081	0.361768	0.0583558	5.66951E-10
N24S16	14	rs12878001	64239629	G	T	0.135081	0.361768	0.0583558	5.66951E-10
N24S16	20	rs3848746	12950606	A	G	0.324597	0.26998	0.0447687	1.63414E-09
N24S16	20	rs2327452	12952964	A	C	0.323589	0.267025	0.0447076	2.33352E-09
N24S16	20	rs2327451	12953934	C	A	0.324093	0.263883	0.0447245	3.63081E-09

N24S16	20	rs4508668	12955601	T	C	0.324093	0.263883	0.0447245	3.63081E-09
N24S16	20	rs3903703	12945963	A	G	0.323589	0.263219	0.0447063	3.91516E-09
N24S16	20	rs4814173	12947532	G	C	0.323589	0.263219	0.0447063	3.91516E-09
N24S16	20	rs6041735	12940649	T	C	0.321069	0.263134	0.0448237	4.34729E-09
N24S16	20	rs438568	12958687	A	G	0.364415	0.250766	0.0429843	5.41461E-09
N24S16	20	rs680379	12969400	A	G	0.36744	0.249217	0.0427935	5.75521E-09
N24S16	20	rs686548	12973521	A	T	0.36744	0.249217	0.0427935	5.75521E-09
N24S16	20	rs3848744	12942649	A	G	0.322077	0.260405	0.044731	5.82991E-09
N24S16	20	rs1321940	12959885	A	G	0.365423	0.247629	0.042985	8.37106E-09
N24S16	20	rs364585	12962718	A	G	0.365423	0.247629	0.042985	8.37106E-09
N24S16	20	rs168622	12966089	T	G	0.365423	0.247629	0.042985	8.37106E-09
N24S16	20	rs3843765	12943737	G	A	0.322581	0.257248	0.0447474	8.98339E-09
N24S16	20	rs3848745	12944067	G	A	0.322581	0.257248	0.0447474	8.98339E-09
N24S16	20	rs4814175	12959094	A	T	0.368448	0.239437	0.0430044	2.58073E-08
N24S16	20	rs4814176	12959398	T	C	0.368448	0.239437	0.0430044	2.58073E-08
N24S19	20	rs438568	12958687	A	G	0.365792	0.469332	0.0430075	1.00128E-27
N24S19	20	rs1321940	12959885	A	G	0.366801	0.466724	0.0430031	1.92443E-27
N24S19	20	rs364585	12962718	A	G	0.366801	0.466724	0.0430031	1.92443E-27
N24S19	20	rs168622	12966089	T	G	0.366801	0.466724	0.0430031	1.92443E-27
N24S19	20	rs680379	12969400	A	G	0.368819	0.461062	0.0428137	4.82E-27
N24S19	20	rs686548	12973521	A	T	0.368819	0.461062	0.0428137	4.82E-27
N24S19	20	rs4814175	12959094	A	T	0.369828	0.453684	0.043043	5.6352E-26
N24S19	20	rs4814176	12959398	T	C	0.369828	0.453684	0.043043	5.6352E-26
N24S19	20	rs3848746	12950606	A	G	0.325429	0.432838	0.0449339	5.8135E-22
N24S19	20	rs2327452	12952964	A	C	0.32442	0.431043	0.0448767	7.61201E-22
N24S19	20	rs2327451	12953934	C	A	0.324924	0.429693	0.0448933	1.05433E-21
N24S19	20	rs4508668	12955601	T	C	0.324924	0.429693	0.0448933	1.05433E-21
N24S19	20	rs3903703	12945963	A	G	0.32442	0.428464	0.0448754	1.32428E-21
N24S19	20	rs4814173	12947532	G	C	0.32442	0.428464	0.0448754	1.32428E-21
N24S19	20	rs6041735	12940649	T	C	0.321393	0.420927	0.0450735	9.75368E-21
N24S19	20	rs3848744	12942649	A	G	0.322402	0.419634	0.0449655	1.03543E-20
N24S19	20	rs3843765	12943737	G	A	0.322906	0.418262	0.0449818	1.42467E-20
N24S19	20	rs3848745	12944067	G	A	0.322906	0.418262	0.0449818	1.42467E-20
N24S19	20	rs4813102	12947883	A	T	0.374369	0.379056	0.0438697	5.59692E-18
N24S19	20	rs608994	12980885	G	A	0.307887	0.295747	0.0444707	2.92304E-11

N24S19	20	rs6078854	12960153	A	T	0.402119	-0.278476	0.0422331	4.28768E-11
N24S19	20	rs3910136	12962261	A	T	0.335015	-0.258289	0.0440702	4.60506E-09
N24S19	20	rs6041755	12973617	T	C	0.334006	-0.251944	0.0440555	1.0728E-08
N24S19	20	rs4544513	12954215	T	C	0.34662	-0.24564	0.0434844	1.61444E-08
N24S19	20	rs6109637	12954804	T	C	0.34662	-0.24564	0.0434844	1.61444E-08
N24S19	20	rs3848754	12971345	C	T	0.284561	-0.25678	0.0465701	3.51103E-08
N24S19	20	rs3848755	12971437	C	T	0.284561	-0.25678	0.0465701	3.51103E-08
N24S19	20	rs13037956	12974302	A	C	0.284561	-0.25678	0.0465701	3.51103E-08
N24S19	20	rs6074538	12974493	T	C	0.284561	-0.25678	0.0465701	3.51103E-08
N24S19	20	rs6078866	12974567	G	A	0.284561	-0.25678	0.0465701	3.51103E-08
N24S19	20	rs6074539	12974665	A	G	0.284561	-0.25678	0.0465701	3.51103E-08
N24S20	20	rs680379	12969400	A	G	0.368976	0.381364	0.0492653	9.86247E-15
N24S20	20	rs686548	12973521	A	T	0.368976	0.381364	0.0492653	9.86247E-15
N24S20	20	rs1321940	12959885	A	G	0.366968	0.381831	0.0494835	1.19732E-14
N24S20	20	rs364585	12962718	A	G	0.366968	0.381831	0.0494835	1.19732E-14
N24S20	20	rs168622	12966089	T	G	0.366968	0.381831	0.0494835	1.19732E-14
N24S20	20	rs438568	12958687	A	G	0.365964	0.376938	0.0494832	2.58624E-14
N24S20	20	rs4814175	12959094	A	T	0.36998	0.371939	0.049508	5.79186E-14
N24S20	20	rs4814176	12959398	T	C	0.36998	0.371939	0.049508	5.79186E-14
N24S20	20	rs2327451	12953934	C	A	0.325803	0.339827	0.0515323	4.26852E-11
N24S20	20	rs4508668	12955601	T	C	0.325803	0.339827	0.0515323	4.26852E-11
N24S20	20	rs3903703	12945963	A	G	0.325301	0.337809	0.0515117	5.45658E-11
N24S20	20	rs4814173	12947532	G	C	0.325301	0.337809	0.0515117	5.45658E-11
N24S20	20	rs2327452	12952964	A	C	0.325301	0.337621	0.051513	5.59796E-11
N24S20	20	rs3848746	12950606	A	G	0.326305	0.337278	0.0515843	6.21872E-11
N24S20	20	rs3843765	12943737	G	A	0.323795	0.332308	0.0516252	1.21908E-10
N24S20	20	rs3848745	12944067	G	A	0.323795	0.332308	0.0516252	1.21908E-10
N24S20	20	rs3848744	12942649	A	G	0.323293	0.330108	0.0516064	1.5883E-10
N24S20	20	rs6041735	12940649	T	C	0.322289	0.32553	0.0517142	3.07808E-10
N24S20	20	rs4813102	12947883	A	T	0.375502	0.304422	0.0503062	1.43619E-09
N25S20	20	rs680379	12969400	A	G	0.367706	0.293801	0.0479832	9.18333E-10
N25S20	20	rs686548	12973521	A	T	0.367706	0.293801	0.0479832	9.18333E-10
N25S20	20	rs1321940	12959885	A	G	0.365694	0.293694	0.0481973	1.10417E-09
N25S20	20	rs364585	12962718	A	G	0.365694	0.293694	0.0481973	1.10417E-09
N25S20	20	rs168622	12966089	T	G	0.365694	0.293694	0.0481973	1.10417E-09

N25S20	20	rs438568	12958687	A	G	0.364688	0.293123	0.0481945	1.1863E-09
N25S20	20	rs4814175	12959094	A	T	0.368712	0.288831	0.0482083	2.08171E-09
N25S20	20	rs4814176	12959398	T	C	0.368712	0.288831	0.0482083	2.08171E-09
N26S19	20	rs438568	12958687	A	G	0.363269	0.325927	0.0447845	3.39629E-13
N26S19	20	rs1321940	12959885	A	G	0.364279	0.324906	0.0447765	3.98235E-13
N26S19	20	rs364585	12962718	A	G	0.364279	0.324906	0.0447765	3.98235E-13
N26S19	20	rs168622	12966089	T	G	0.364279	0.324906	0.0447765	3.98235E-13
N26S19	20	rs680379	12969400	A	G	0.366297	0.319887	0.0445781	7.18432E-13
N26S19	20	rs686548	12973521	A	T	0.366297	0.319887	0.0445781	7.18432E-13
N26S19	20	rs4814175	12959094	A	T	0.367306	0.318806	0.0448248	1.14155E-12
N26S19	20	rs4814176	12959398	T	C	0.367306	0.318806	0.0448248	1.14155E-12
N26S19	6	rs6940658	14238511	C	G	0.0882947	0.481412	0.074604	1.09734E-10
N26S19	6	rs4333409	14240330	A	C	0.0882947	0.481412	0.074604	1.09734E-10
N26S19	6	rs2039310	14250304	A	G	0.0882947	0.481412	0.074604	1.09734E-10
N26S19	6	rs9382948	14251751	A	G	0.0882947	0.481412	0.074604	1.09734E-10
N26S19	6	rs6910045	14255808	G	A	0.0882947	0.481412	0.074604	1.09734E-10
N26S19	6	rs9367828	14236554	G	A	0.0872856	0.479767	0.0754176	1.99826E-10
N26S19	6	rs9370735	14244517	T	G	0.0872856	0.479767	0.0754176	1.99826E-10
N26S19	6	rs6940973	14238655	C	T	0.0857719	0.465654	0.0759222	8.60712E-10
N26S19	6	rs1537152	14240435	C	A	0.0857719	0.465654	0.0759222	8.60712E-10
N26S19	6	rs1537151	14240594	C	A	0.0857719	0.465654	0.0759222	8.60712E-10
N26S19	6	rs9396477	14234971	C	G	0.0787084	0.43973	0.0777369	1.54356E-08
N26S19	20	rs3848746	12950606	A	G	0.323411	0.263021	0.0468486	1.97374E-08
N26S19	20	rs6041735	12940649	T	C	0.319879	0.263456	0.0469477	2.00364E-08
N26S19	6	rs12208698	14237070	G	T	0.073663	0.457551	0.0816213	2.07314E-08
N26S19	6	rs12190393	14239216	A	G	0.073663	0.457551	0.0816213	2.07314E-08
N26S19	6	rs12212956	14240006	G	A	0.073663	0.457551	0.0816213	2.07314E-08
N26S19	6	rs12207359	14244274	T	G	0.073663	0.457551	0.0816213	2.07314E-08
N26S19	6	rs75762794	14245458	C	G	0.073663	0.457551	0.0816213	2.07314E-08
N26S19	6	rs2876349	14246531	T	C	0.073663	0.457551	0.0816213	2.07314E-08
N26S19	6	rs12213267	14247608	C	G	0.073663	0.457551	0.0816213	2.07314E-08
N26S19	6	rs115366574	14250710	C	T	0.073663	0.457551	0.0816213	2.07314E-08
N26S19	6	rs79263173	14253258	T	A	0.073663	0.457551	0.0816213	2.07314E-08
N26S19	20	rs2327452	12952964	A	C	0.322402	0.26148	0.0467913	2.29423E-08
N26S19	20	rs3848744	12942649	A	G	0.320383	0.261192	0.0468927	2.54741E-08

N26S19	20	rs2327451	12953934	C	A	0.322906	0.259595	0.0468082	2.92399E-08
N26S19	20	rs4508668	12955601	T	C	0.322906	0.259595	0.0468082	2.92399E-08
N26S19	20	rs3843765	12943737	G	A	0.320888	0.259293	0.0469092	3.24709E-08
N26S19	20	rs3848745	12944067	G	A	0.320888	0.259293	0.0469092	3.24709E-08
N26S19	20	rs3903703	12945963	A	G	0.322402	0.258485	0.0467902	3.30722E-08
N26S19	20	rs4814173	12947532	G	C	0.322402	0.258485	0.0467902	3.30722E-08
N24S19ratio	1	rs4653568	224396329	A	G	0.353179	0.292434	0.0461572	2.36406E-10
N24S19ratio	1	rs4654000	224396487	C	T	0.353179	0.292434	0.0461572	2.36406E-10
N24S19ratio	1	rs9793489	224397459	T	A	0.353179	0.292434	0.0461572	2.36406E-10
N24S19ratio	1	rs2011117	224398639	C	T	0.353179	0.292434	0.0461572	2.36406E-10
N24S19ratio	1	rs908801	224398817	T	C	0.353179	0.292434	0.0461572	2.36406E-10
N24S19ratio	1	rs6681673	224398910	T	C	0.353179	0.292434	0.0461572	2.36406E-10
N24S19ratio	1	rs4654003	224399383	T	C	0.353179	0.292434	0.0461572	2.36406E-10
N24S19ratio	1	rs6682292	224400175	A	G	0.353179	0.292434	0.0461572	2.36406E-10
N24S19ratio	1	rs12038372	224363881	T	C	0.410101	0.258043	0.0453781	1.29654E-08
N24S19ratio	1	rs6426143	224318875	G	A	0.39697	0.253265	0.0452329	2.1541E-08
N24S19ratio	1	rs6682551	224352258	C	T	0.391742	0.251419	0.0450659	2.4202E-08
N24S19ratio	1	rs10799505	224351023	T	C	0.395267	0.24962	0.0451549	3.23753E-08
N24S19ratio	1	rs10916403	224352551	C	A	0.395267	0.24962	0.0451549	3.23753E-08
N24S19ratio	1	rs55664906	224355682	A	C	0.395267	0.24962	0.0451549	3.23753E-08
N24S19ratio	1	rs12076788	224355701	C	A	0.395267	0.24962	0.0451549	3.23753E-08
N24S19ratio	1	rs4653563	224314776	A	C	0.39577	0.248927	0.0452131	3.6785E-08
N24S19ratio	1	rs4653564	224314814	C	T	0.39577	0.248927	0.0452131	3.6785E-08
N24S19ratio	1	rs6426139	224315128	A	G	0.39577	0.248927	0.0452131	3.6785E-08
N24S19ratio	1	rs10753454	224316813	C	T	0.39577	0.248927	0.0452131	3.6785E-08
N24S19ratio	1	rs7542293	224316881	A	G	0.39577	0.248927	0.0452131	3.6785E-08
N24S19ratio	1	rs7542474	224317043	A	C	0.39577	0.248927	0.0452131	3.6785E-08
N24S19ratio	1	rs6698041	224317551	G	A	0.39577	0.248927	0.0452131	3.6785E-08
N24S19ratio	1	rs7519433	224319003	T	C	0.39577	0.248927	0.0452131	3.6785E-08
N24S19ratio	1	rs12730611	224319825	G	C	0.39577	0.248927	0.0452131	3.6785E-08
N24S19ratio	1	rs2014782	224321492	T	C	0.39577	0.248927	0.0452131	3.6785E-08
N24S19ratio	1	rs869945	224321637	G	A	0.39577	0.248927	0.0452131	3.6785E-08
N24S19ratio	1	rs6691405	224322879	A	G	0.39577	0.248927	0.0452131	3.6785E-08
N24S19ratio	1	rs6685783	224322942	T	C	0.39577	0.248927	0.0452131	3.6785E-08
N24S19ratio	1	rs10916355	224323548	T	C	0.39577	0.248927	0.0452131	3.6785E-08

N24S19ratio	1	rs66506223	224325436	C	G	0.39577	0.248927	0.0452131	3.6785E-08
N24S19ratio	1	rs61827699	224325460	G	A	0.39577	0.248927	0.0452131	3.6785E-08
N24S19ratio	1	rs10916367	224329969	G	A	0.396274	0.247343	0.0451633	4.33499E-08
N24S19ratio	1	rs4653986	224330818	A	G	0.396274	0.247343	0.0451633	4.33499E-08
N24S19ratio	1	rs7518839	224331779	G	A	0.396274	0.247343	0.0451633	4.33499E-08
N24S19ratio	1	rs13374070	224332078	A	G	0.396274	0.247343	0.0451633	4.33499E-08
N24S19ratio	1	rs55736782	224334072	G	A	0.396274	0.247343	0.0451633	4.33499E-08
N24S19ratio	1	rs997297	224335492	A	T	0.396274	0.247343	0.0451633	4.33499E-08
N24S19ratio	1	rs997296	224335497	C	G	0.396274	0.247343	0.0451633	4.33499E-08
N24S19ratio	1	rs7531891	224335956	A	G	0.396274	0.247343	0.0451633	4.33499E-08
N24S19ratio	1	rs1492694	224337153	G	A	0.396274	0.247343	0.0451633	4.33499E-08
N24S19ratio	1	rs4653991	224337208	G	A	0.396274	0.247343	0.0451633	4.33499E-08
N24S19ratio	1	rs7546235	224338017	T	C	0.396274	0.247343	0.0451633	4.33499E-08
N24S19ratio	1	rs10916371	224338162	A	G	0.396274	0.247343	0.0451633	4.33499E-08
N24S19ratio	1	rs7524705	224339000	A	G	0.396274	0.247343	0.0451633	4.33499E-08
N24S19ratio	1	rs1826421	224339447	A	G	0.396274	0.247343	0.0451633	4.33499E-08
N24S19ratio	1	rs12563153	224339647	A	G	0.396274	0.247343	0.0451633	4.33499E-08
N24S19ratio	1	rs8328	224346959	G	A	0.396274	0.247343	0.0451633	4.33499E-08
N24S19ratio	1	rs7526252	224348818	T	C	0.396274	0.247343	0.0451633	4.33499E-08
N25Sum	20	rs1321940	12959885	A	G	0.365191	0.289696	0.0479366	1.50976E-09
N25Sum	20	rs364585	12962718	A	G	0.365191	0.289696	0.0479366	1.50976E-09
N25Sum	20	rs168622	12966089	T	G	0.365191	0.289696	0.0479366	1.50976E-09
N25Sum	20	rs4814175	12959094	A	T	0.368209	0.288593	0.0479673	1.78272E-09
N25Sum	20	rs4814176	12959398	T	C	0.368209	0.288593	0.0479673	1.78272E-09
N25Sum	20	rs438568	12958687	A	G	0.364185	0.287816	0.0479374	1.92521E-09
N25Sum	20	rs680379	12969400	A	G	0.367203	0.285702	0.0477235	2.14271E-09
N25Sum	20	rs686548	12973521	A	T	0.367203	0.285702	0.0477235	2.14271E-09
S19Sum	20	rs438568	12958687	A	G	0.365792	0.483828	0.0432293	4.45656E-29
S19Sum	20	rs1321940	12959885	A	G	0.366801	0.482428	0.0432244	6.32768E-29
S19Sum	20	rs364585	12962718	A	G	0.366801	0.482428	0.0432244	6.32768E-29
S19Sum	20	rs168622	12966089	T	G	0.366801	0.482428	0.0432244	6.32768E-29
S19Sum	20	rs680379	12969400	A	G	0.368819	0.475533	0.0430341	2.18888E-28
S19Sum	20	rs686548	12973521	A	T	0.368819	0.475533	0.0430341	2.18888E-28
S19Sum	20	rs4814175	12959094	A	T	0.369828	0.468758	0.043266	2.36728E-27
S19Sum	20	rs4814176	12959398	T	C	0.369828	0.468758	0.043266	2.36728E-27

S19Sum	20	rs3848746	12950606	A	G	0.325429	0.442746	0.0451725	1.11208E-22
S19Sum	20	rs2327452	12952964	A	C	0.32442	0.440637	0.0451153	1.56152E-22
S19Sum	20	rs2327451	12953934	C	A	0.324924	0.439777	0.045132	1.95249E-22
S19Sum	20	rs4508668	12955601	T	C	0.324924	0.439777	0.045132	1.95249E-22
S19Sum	20	rs3903703	12945963	A	G	0.32442	0.438144	0.045114	2.68232E-22
S19Sum	20	rs4814173	12947532	G	C	0.32442	0.438144	0.045114	2.68232E-22
S19Sum	20	rs3848744	12942649	A	G	0.322402	0.42941	0.0452056	2.1184E-21
S19Sum	20	rs3843765	12943737	G	A	0.322906	0.428529	0.0452219	2.63895E-21
S19Sum	20	rs3848745	12944067	G	A	0.322906	0.428529	0.0452219	2.63895E-21
S19Sum	20	rs6041735	12940649	T	C	0.321393	0.428393	0.0453155	3.27512E-21
S19Sum	20	rs4813102	12947883	A	T	0.374369	0.391123	0.0441069	7.47276E-19
S19Sum	20	rs6078854	12960153	A	T	0.402119	-0.295564	0.0424598	3.37791E-12
S19Sum	20	rs608994	12980885	G	A	0.307887	0.300951	0.0447195	1.69967E-11
S19Sum	20	rs4544513	12954215	T	C	0.34662	-0.263011	0.0437192	1.78886E-09
S19Sum	20	rs6109637	12954804	T	C	0.34662	-0.263011	0.0437192	1.78886E-09
S19Sum	20	rs382003	12963171	A	G	0.298184	-0.263466	0.0453068	6.0575E-09
S19Sum	20	rs360539	12966440	G	T	0.297679	-0.261227	0.0451525	7.23288E-09
S19Sum	20	rs3910136	12962261	A	T	0.335015	-0.254967	0.0443103	8.70942E-09
S19Sum	20	rs73079703	12941782	T	C	0.297175	-0.253683	0.0454613	2.40243E-08
S19Sum	20	rs8183164	12942600	A	C	0.297175	-0.253683	0.0454613	2.40243E-08
S19Sum	20	rs6131414	12945669	A	G	0.297175	-0.253683	0.0454613	2.40243E-08
S19Sum	20	rs7272107	12946328	A	G	0.297175	-0.253683	0.0454613	2.40243E-08
S19Sum	20	rs73079713	12947141	T	C	0.297175	-0.253683	0.0454613	2.40243E-08
S19Sum	20	rs6041755	12973617	T	C	0.334006	-0.247109	0.044295	2.42327E-08
S19Sum	20	rs6134734	12952640	T	A	0.29667	-0.249622	0.0454395	3.94041E-08
S19Sum	20	rs6109634	12953314	G	A	0.29667	-0.249622	0.0454395	3.94E-08
S19Sum	20	rs3848748	12957587	C	G	0.294652	-0.249321	0.0454975	4.26E-08
S19Sum	20	rs3848749	12962089	C	T	0.294652	-0.249321	0.0454975	4.26E-08
S19Sum	20	rs3848754	12971345	C	T	0.284561	-0.255417	0.0468243	4.90E-08
S19Sum	20	rs3848755	12971437	C	T	0.284561	-0.255417	0.0468243	4.90E-08
S19Sum	20	rs13037956	12974302	A	C	0.284561	-0.255417	0.0468243	4.90E-08
S19Sum	20	rs6074538	12974493	T	C	0.284561	-0.255417	0.0468243	4.90E-08
S19Sum	20	rs6078866	12974567	G	A	0.284561	-0.255417	0.0468243	4.90E-08
S19Sum	20	rs6074539	12974665	A	G	0.284561	-0.255417	0.0468243	4.90E-08
S20Sum	20	rs680379	12969400	A	G	0.369107	0.40291	0.0496971	5.17E-16

S20Sum	20	rs686548	12973521	A	T	0.369107	0.40291	0.0496971	5.17E-16
S20Sum	20	rs1321940	12959885	A	G	0.367101	0.403861	0.0499178	5.94E-16
S20Sum	20	rs364585	12962718	A	G	0.367101	0.403861	0.0499178	5.94E-16
S20Sum	20	rs168622	12966089	T	G	0.367101	0.403861	0.0499178	5.94E-16
S20Sum	20	rs438568	12958687	A	G	0.366098	0.399976	0.0499174	1.12E-15
S20Sum	20	rs4814175	12959094	A	T	0.37011	0.393035	0.0499442	3.56E-15
S20Sum	20	rs4814176	12959398	T	C	0.37011	0.393035	0.0499442	3.56E-15
S20Sum	20	rs2327451	12953934	C	A	0.325476	0.352063	0.052026	1.31E-11
S20Sum	20	rs4508668	12955601	T	C	0.325476	0.352063	0.052026	1.31E-11
S20Sum	20	rs2327452	12952964	A	C	0.324975	0.35107	0.0520066	1.47E-11
S20Sum	20	rs3903703	12945963	A	G	0.324975	0.34981	0.0520046	1.74E-11
S20Sum	20	rs4814173	12947532	G	C	0.324975	0.34981	0.0520046	1.74E-11
S20Sum	20	rs3848746	12950606	A	G	0.325978	0.350143	0.0520774	1.77E-11
S20Sum	20	rs3843765	12943737	G	A	0.32347	0.344118	0.052119	4.04E-11
S20Sum	20	rs3848745	12944067	G	A	0.32347	0.344118	0.052119	4.04E-11
S20Sum	20	rs3848744	12942649	A	G	0.322969	0.343134	0.0521001	4.52E-11
S20Sum	20	rs6041735	12940649	T	C	0.321966	0.339062	0.0522108	8.35E-11
S20Sum	20	rs4813102	12947883	A	T	0.375125	0.317817	0.050788	3.91E-10

Table 0.11: Ensembl and GTEx summary of the significant SNPs identified by GWAS association for ceramides and related sphingolipid species

Description of the SNPs using the Ensemble API Client and summary of the information on eQTL status as identified using the GTEX browser, including their identification in liver or whole blood.

SNP ID	Gene Name	Gene Type	Variant Allele	Consequence	GTEx	Liver or whole blood?	Other
rs438568	*LINC01723*	lncRNA	G	intron	*SPTLC3*	Liver	*X*
rs1321940	*LINC01723*	lncRNA	G	intron	*SPTLC3*	Liver	*ISM1 (pancreas)*
rs364585	*LINC01723*	lncRNA	G	downstream	*SPTLC3*	Liver	*X*
rs168622	*LINC01723*	lncRNA	G	intron	*SPTLC3*	Liver	*X*
rs680379	*LINC01723*	lncRNA	G	intron	*SPTLC3*	Liver	*X*
rs686548	*LINC01723*	lncRNA	T	intron	*SPTLC3*	Liver	*ISM1 (pancreas)*
rs4814175	*LINC01723*	lncRNA	T	intron	*SPTLC3*	Liver	*X*
rs4814176	*LINC01723*	lncRNA	C	intron	*SPTLC3*	Liver	*X*
rs2327452	*LINC01723*	lncRNA	C	intron	*SPTLC3*	Liver	*X*
rs3848746	*LINC01723*	lncRNA	G	intron	*SPTLC3*	Liver	*X*
rs2327451	*LINC01723*	lncRNA	A	intron	*SPTLC3*	Liver	*X*
rs4508668	*LINC01723*	lncRNA	C	intron	*SPTLC3*	Liver	*X*
rs3903703	*LINC01723*	lncRNA	G	intron	*SPTLC3*	Liver	*X*
rs4814173	*LINC01723*	lncRNA	A	intron	*SPTLC3*	Liver	*X*
rs3848744	*LINC01723*	lncRNA	G	intron	*SPTLC3*	Liver	*X*
rs3843765	*LINC01723*	lncRNA	A	intron	*SPTLC3*	Liver	*X*
rs3848745	*LINC01723*	lncRNA	A	intron	*SPTLC3*	Liver	*X*
rs6041735	*LINC01723*	lncRNA	C	intron	*SPTLC3*	Liver	*X*
rs4813102	*LINC01723*	lncRNA	T	intron	*SPTLC3*	Liver	*X*
rs6078854	*LINC01723*	lncRNA	A	downstream	*SPTLC3*	X	*X*
rs4544513	*LINC01723*	lncRNA	T	intron	*SPTLC3*	Liver	*X*
rs6109637	*LINC01723*	lncRNA	T	intron	*SPTLC3*	Liver	*X*
rs382003	*LINC01723*	lncRNA	A	downstream	*SPTLC3*	Liver	*X*
rs360539	*LINC01723*	lncRNA	C	intron	*SPTLC3*	Liver	*X*
rs73079703	*LINC01723*	lncRNA	T	intron	*X*	X	*X*

rs8183164	*LINC01723*	lncRNA	A	intron	*X*	X	*X*
rs6131414	*LINC01723*	lncRNA	A	intron	*X*	X	*X*
rs7272107	*LINC01723*	lncRNA	A	intron	*X*	X	*X*
rs73079713	*LINC01723*	lncRNA	T	intron	*X*	X	*X*
rs6134734	*LINC01723*	lncRNA	T	intron	*X*	X	*X*
rs6109634	*LINC01723*	lncRNA	G	intron	*X*	X	*X*
rs3848748	*LINC01723*	lncRNA	C	intron	*X*	X	*X*
rs3848749	*LINC01723*	lncRNA	C	downstream	*X*	X	*X*
rs6131417	*LINC01723*	lncRNA	A	intron	*X*	X	*X*
rs6134740	*LINC01723*	lncRNA	C	intron	*X*	X	*X*
rs6134741	*LINC01723*	lncRNA	A	intron	*X*	X	*X*
rs59131252	*LINC01723*	lncRNA	A	intron	*X*	X	*X*
rs7160525	*AL161670.1*	pseudogene	A	downstream	*X*	X	*X*
rs17101394	*AL161670.1*	pseudogene	A	downstream	*X*	X	*X*
rs8008068	*AL161670.1*	pseudogene	G	downstream	*X*	X	*X*
rs8008070	*AL161670.1*	pseudogene	T	downstream	*X*	X	*X*
rs8012828	*Intergenic*	Intergenic	T	intergenic	*X*	X	*X*
rs34609767	*Intergenic*	Intergenic	G	intergenic	*X*	X	*X*
rs4902243	*Intergenic*	Intergenic	G	intergenic	*X*	X	*X*
rs7157785	*Intergenic*	Intergenic	T	intergenic	*X*	X	*X*
rs34817779	*Intergenic*	Intergenic	T	intergenic	*X*	X	*X*
rs35372182	*Intergenic*	Intergenic	G	intergenic	*X*	X	*X*
rs12897637	*Intergenic*	Intergenic	C	intergenic	*X*	X	*X*
rs12878001	*Intergenic*	Intergenic	G	intergenic	*X*	X	*X*
rs608994	*LINC01723*	lncRNA	A	intron	*SPTLC3*	Liver	*X*
rs3910136	*LINC01723*	lncRNA	A	downstream	*NA*	NA	*NA*
rs6041755	*LINC01723*	lncRNA	T	intron	*SPTLC3*	X	*X*
rs3848754	*LINC01723*	lncRNA	C	intron	*SPTLC3*	X	*X*
rs3848755	*LINC01723*	lncRNA	C	intron	*X*	X	*X*
rs13037956	*LINC01723*	lncRNA	A	intron	*SPTLC3*	X	*X*
rs6074538	*LINC01723*	lncRNA	T	intron	*SPTLC3*	X	*X*
rs6078866	*LINC01723*	lncRNA	G	intron	*SPTLC3*	X	*X*
rs6074539	*LINC01723*	lncRNA	A	intron	*X*	X	*X*

rs6940658	*Intergenic*	Intergenic	G	intergenic	*X*	X	*X*
rs4333409	*Intergenic*	Intergenic	C	intergenic	*X*	X	*X*
rs2039310	*Intergenic*	Intergenic	C	intergenic	*X*	X	*X*
rs9382948	*Intergenic*	Intergenic	G	intergenic	*X*	X	*X*
rs6910045	*Intergenic*	Intergenic	A	intergenic	*RP3-500L14.2*	X	*X*
rs9367828	*AL353152.1*	lncRNA	A	upstream	*X*	X	*X*
rs9370735	*Intergenic*	Intergenic	G	intergenic	*X*	X	*X*
rs6940973	*Intergenic*	Intergenic	T	intergenic	*X*	X	*X*
rs1537152	*Intergenic*	Intergenic	A	intergenic	*X*	X	*X*
rs1537151	*Intergenic*	Intergenic	A	intergenic	*X*	X	*X*
rs9396477	*AL353152.1*	lncRNA	G	upstream	*X*	X	*X*
rs12208698	*Intergenic*	Intergenic	A	intergenic	*X*	X	*X*
rs12190393	*Intergenic*	Intergenic	A	intergenic	*X*	X	*X*
rs12212956	*Intergenic*	Intergenic	G	intergenic	*X*	X	*X*
rs12207359	*Intergenic*	Intergenic	T	intergenic	*X*	X	*X*
rs75762794	*Intergenic*	Intergenic	C	intergenic	*X*	X	*X*
rs2876349	*Intergenic*	Intergenic	A	intergenic	*X*	X	*X*
rs12213267	*Intergenic*	Intergenic	C	intergenic	*X*	X	*X*
rs115366574	*Intergenic*	Intergenic	C	intergenic	*X*	X	*X*
rs79263173	*Intergenic*	Intergenic	T	intergenic	*X*	X	*X*
rs4653568	*AC092809.2*	lncRNA	G	downstream	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*
rs4654000	*AC092809.2*	lncRNA	G	exon	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8, GTP2IP20*
rs9793489	*AC092809.2*	lncRNA	A	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8, GTP2IP20*
rs2011117	*AC092809.2*	lncRNA	T	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8, GTP2IP20*
rs908801	*AC092809.2*	lncRNA	A	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8, GTP2IP20*
rs6681673	*AC092809.2*	lncRNA	C	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8, GTP2IP20*
rs4654003	*AC092809.2*	lncRNA	C	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8, GTP2IP20*

rs6682292	*AC092809.2*	lncRNA	G	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8, GTP2IP20*
rs12038372	*DEGS1*	protein_coding	G	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*
rs6426143	*FBXO28*	protein_coding	A	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*
rs6682551	*FBXO28*	protein_coding	C	downstream	*FBXO28*	blood	*RP11-365O16.3*
rs10799505	*FBXO28*	protein_coding	T	downstream	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8, CNIH3*
rs10916403	*FBXO28*	protein_coding	C	downstream	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8, CNIH3*
rs55664906	*Intergenic*	Intergenic	A	intergenic	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*
rs12076788	*Intergenic*	Intergenic	A	intergenic	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*
rs4653563	*FBXO28*	protein_coding	C	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*
rs4653564	*FBXO28*	protein_coding	T	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*
rs6426139	*FBXO28*	protein_coding	G	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*
rs10753454	*FBXO28*	protein_coding	G	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*
rs7542293	*FBXO28*	protein_coding	G	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*
rs7542474	*FBXO28*	protein_coding	C	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*
rs6698041	*FBXO28*	protein_coding	A	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*
rs7519433	*FBXO28*	protein_coding	T	intron	*FBXO28*	blood	*RP11-365O16.3*
rs12730611	*FBXO28*	protein_coding	C	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*
rs2014782	*FBXO28*	protein_coding	T	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*
rs869945	*FBXO28*	protein_coding	G	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*
rs6691405	*FBXO28*	protein_coding	G	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*
rs6685783	*FBXO28*	protein_coding	C	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*
rs10916355	*FBXO28*	protein_coding	T	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*
rs66506223	*FBXO28*	protein_coding	G	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*

rs61827699	*FBXO28*	protein_coding	A	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*
rs10916367	*FBXO28*	protein_coding	A	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*
rs4653986	*FBXO28*	protein_coding	G	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*
rs7518839	*FBXO28*	protein_coding	A	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*
rs13374070	*FBXO28*	protein_coding	A	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*
rs55736782	*FBXO28*	protein_coding	G	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*
rs997297	*FBXO28*	protein_coding	T	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*
rs997296	*FBXO28*	protein_coding	A	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*
rs7531891	*FBXO28*	protein_coding	A	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*
rs1492694	*FBXO28*	protein_coding	G	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*
rs4653991	*FBXO28*	protein_coding	A	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*
rs7546235	*FBXO28*	protein_coding	T	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*
rs10916371	*FBXO28*	protein_coding	G	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*
rs7524705	*FBXO28*	protein_coding	A	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*
rs1826421	*FBXO28*	protein_coding	G	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*
rs12563153	*FBXO28*	protein_coding	A	intron	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*
rs8328	*FBXO28*	protein_coding	A	3_prime_UTR	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*
rs7526252	*FBXO28*	protein_coding	C	3_prime_UTR	*FBXO28*	blood	*DEGS1, RP11-365O16.3, CAPN8*

Table 0.12: GWAS Catalog searches, Gene Atlas PheWAS, and UCSC Genome Browser summaries of the significant SNPs identified by GWAS association for ceramides and related sphingolipid species

The identification of the SNPs in the browsers are described.

SNP ID	GWAS Catalog	Gene Atlas	UCSC
rs438568	X	X	NA
rs1321940	X	X	NA
rs364585	LDL cholesterol	X	NA
rs168622	X	X	NA
rs680379	SL levels, FA levels, Glycerophospholipid levels	X	NA
rs686548	Serum metabolite ratios in chronic kidney disease	X	NA
rs4814175	X	X	NA
rs4814176	SL levels, blood metabolites	X	NA
rs2327452	X	X	NA
rs3848746	X	X	NA
rs2327451	X	X	NA
rs4508668	X	X	NA
rs3903703	FA levels	X	NA
rs4814173	X	X	NA
rs3848744	X	X	NA
rs3843765	X	X	NA
rs3848745	X	X	NA
rs6041735	X	X	NA
rs4813102	X	X	NA
rs6078854	X	X	NA
rs4544513	X	X	NA
rs6109637	X	X	NA
rs382003	X	X	NA
rs360539	X	X	NA
rs73079703	X	X	NA
rs8183164	X	X	NA
rs6131414	X	X	NA
rs7272107	X	X	NA

rs73079713	X	X	NA
rs6134734	X	X	NA
rs6109634	X	X	NA
rs3848748	X	X	NA
rs3848749	X	X	NA
rs6131417	X	X	NA
rs6134740	X	X	NA
rs6134741	X	X	NA
rs59131252	X	X	NA
rs7160525	Serum metabolite concentrations in chronic kidney disease	Mean platelet (thrombocyte) volume (P=3.2825e-29); Red blood cell (erythrocyte) distribution width (P=5.963e-14); Platelet count (P=5.5601e-13); High light scatter reticulocyte percentage (P=1.2661e-12); High light scatter reticulocyte count (P=6.6865e-11); Immature reticulocyte fraction (P=1.0614e-10); Reticulocyte percentage (P=1.9119e-08)	SGPP1
rs17101394	SL levels	Mean platelet (thrombocyte) volume (P=5.1305e-29); Red blood cell (erythrocyte) distribution width (P=4.9311e-14); Platelet count (P=6.397e-13); High light scatter reticulocyte percentage (P=3.2699e-12); High light scatter reticulocyte count (P=1.1051e-10); Immature reticulocyte fraction (P=2.0931e-10); Reticulocyte percentage (P=1.8191e-08)	SGPP1
rs8008068	Red cell distribution width	Mean platelet (thrombocyte) volume (P=4.7089e-29); Red blood cell (erythrocyte) distribution width (P=4.8152e-14); Platelet count (P=6.2582e-13); High light scatter reticulocyte percentage (P=1.866e-12); Immature reticulocyte fraction (P=5.3687e-11); High light scatter reticulocyte count (P=6.6258e-11); Reticulocyte percentage (P=2.0003e-08)	SGPP1
rs8008070	Serum metabolite ratios in chronic kidney disease	Mean platelet (thrombocyte) volume (P=2.8823e-29); Red blood cell (erythrocyte) distribution width (P=4.8987e-14); Platelet count (P=5.7716e-13); High light scatter reticulocyte percentage (P=1.7608e-12); Immature reticulocyte fraction (P=4.8545e-11); High light scatter reticulocyte count (P=6.1781e-11); Reticulocyte percentage (P=1.9361e-08)	SGPP1
rs8012828	X	Mean platelet (thrombocyte) volume (P=4.6195e-29); Red blood cell (erythrocyte) distribution width (P=5.5573e-14); Platelet count (P=7.3757e-13); High light scatter reticulocyte percentage (P=3.4938e-12); High light scatter reticulocyte count (P=1.2452e-10); Immature reticulocyte fraction (P=2.1356e-10); Reticulocyte percentage (P=1.6812e-	SGPP1

		08)	
rs34609767	X	Mean platelet (thrombocyte) volume (P=3.1253e-29); Red blood cell (erythrocyte) distribution width (P=4.292e-14); Platelet count (P=5.6134e-13); High light scatter reticulocyte percentage (P=1.75e-12); High light scatter reticulocyte count (P=6.6063e-11); Immature reticulocyte fraction (P=6.8857e-11); Reticulocyte percentage (P=1.6786e-08)	SGPP1
rs4902243	Blood metabolite levels	Mean platelet (thrombocyte) volume (P=4.2606e-29); Red blood cell (erythrocyte) distribution width (P=2.0997e-14); Platelet count (P=6.7994e-13); High light scatter reticulocyte percentage (P=2.5783e-12); High light scatter reticulocyte count (P=9.5096e-11); Immature reticulocyte fraction (P=1.0739e-10); Reticulocyte percentage (P=2.0268e-08)	SGPP1
rs7157785	SL levels, blood metabolites, glycerophospholipids, total cholesterol [Hicks, shin, draisma, other]	Mean platelet (thrombocyte) volume (P=1.2992e-28); Red blood cell (erythrocyte) distribution width (P=3.5732e-13); Platelet count (P=9.5427e-13); High light scatter reticulocyte percentage (P=1.0509e-12); Immature reticulocyte fraction (P=5.289e-11); High light scatter reticulocyte count (P=5.6669e-11); Reticulocyte percentage (P=1.6633e-08)	SGPP1
rs34817779	X	Mean platelet (thrombocyte) volume (P=3.6761e-29); Red blood cell (erythrocyte) distribution width (P=3.6016e-14); Platelet count (P=4.2642e-13); High light scatter reticulocyte percentage (P=1.5459e-12); High light scatter reticulocyte count (P=6.0295e-11); Immature reticulocyte fraction (P=8.1684e-11); Reticulocyte percentage (P=1.5888e-08)	SGPP1
rs35372182	X	Mean platelet (thrombocyte) volume (P=4.1731e-29); Red blood cell (erythrocyte) distribution width (P=4.1885e-14); Platelet count (P=4.8832e-13); High light scatter reticulocyte percentage (P=2.0206e-12); High light scatter reticulocyte count (P=7.8142e-11); Immature reticulocyte fraction (P=8.981e-11); Reticulocyte percentage (P=1.7224e-08)	SGPP1
rs12897637	red blood cell distribution width	Mean platelet (thrombocyte) volume (P=9.1577e-29); Red blood cell (erythrocyte) distribution width (P=1.2123e-14); Platelet count (P=3.9706e-13); High light scatter reticulocyte percentage (P=1.2183e-12); High light scatter reticulocyte count (P=4.5639e-11); Immature reticulocyte fraction (P=9.9928e-11); Reticulocyte percentage (P=1.2307e-08)	SGPP1
rs12878001	X	Mean platelet (thrombocyte) volume (P=4.1485e-29); Red blood cell (erythrocyte) distribution width (P=1.9038e-14); Platelet count (P=2.7341e-13); High light scatter reticulocyte percentage (P=7.549e-	SGPP1

		13); High light scatter reticulocyte count (P=2.948e-11); Immature reticulocyte fraction (P=5.8129e-11); Reticulocyte percentage (P=1.1537e-08)	
rs608994	X	X	NA
rs3910136	X	X	NA
rs6041755	X	X	NA
rs3848754	X	X	NA
rs3848755	X	X	NA
rs13037956	X	X	NA
rs6074538	X	X	NA
rs6078866	X	X	NA
rs6074539	X	X	NA
rs6940658	X	X	upstream to CD83
rs4333409	X	X	upstream to CD83
rs2039310	X	X	upstream to CD83
rs9382948	X	X	upstream to CD83
rs6910045	X	X	upstream to CD83
rs9367828	X	X	upstream to CD83
rs9370735	X	X	upstream to CD83
rs6940973	X	X	upstream to CD83
rs1537152	X	X	upstream to CD83
rs1537151	X	X	upstream to CD83
rs9396477	X	X	upstream to CD83
rs12208698	X	X	upstream to CD83
rs12190393	X	X	upstream to CD83
rs12212956	X	X	upstream to CD83
rs12207359	X	X	upstream to CD83
rs75762794	X	X	upstream to CD83
rs2876349	X	X	upstream to CD83
rs12213267	X	X	upstream to CD83
rs115366574	X	X	upstream to CD83
rs79263173	X	X	upstream to CD83
rs4653568	X	Mean platelet (thrombocyte) volume (P=4.7652e-12)	DEGS1
rs4654000	X	Mean platelet (thrombocyte) volume (P=4.8955e-12)	DEGS1

rs9793489	X	Mean platelet (thrombocyte) volume (P=6.0912e-12)	DEGS1
rs2011117	X	Mean platelet (thrombocyte) volume (P=6.3718e-12)	DEGS1
rs908801	X	Mean platelet (thrombocyte) volume (P=6.3548e-12)	DEGS1
rs6681673	X	Mean platelet (thrombocyte) volume (P=7.1624e-12)	DEGS1
rs4654003	X	Mean platelet (thrombocyte) volume (P=6.6526e-12)	DEGS1
rs6682292	X	Mean platelet (thrombocyte) volume (P=7.1866e-12)	DEGS1
rs12038372	X	Mean platelet (thrombocyte) volume (P=3.3948e-18); Neutrophil count (P=1.1357e-09); White blood cell (leukocyte) count (P=3.1852e-09)).	DEGS1
rs6426143	X	Mean platelet (thrombocyte) volume (P=1.9455e-16); White blood cell (leukocyte) count (P=4.4764e-10); Neutrophil count (P=1.7302e-09); Red blood cell (erythrocyte) count (P=2.3883e-09)).	DEGS1
rs6682551	X	Mean platelet (thrombocyte) volume (P=1.409e-16); White blood cell (leukocyte) count (P=2.7024e-10); Neutrophil count (P=4.8023e-09); Red blood cell (erythrocyte) count (P=1.2395e-09)).	DEGS1
rs10799505	X	Mean platelet (thrombocyte) volume (P=1.513e-16); White blood cell (leukocyte) count (P=4.7039e-10); Red blood cell (erythrocyte) count (P=5.487e-10); Neutrophil count (P=6.0446e-10)).	DEGS1
rs10916403	X	Mean platelet (thrombocyte) volume (P=1.6163e-16); White blood cell (leukocyte) count (P=4.6079e-10); Red blood cell (erythrocyte) count (P=5.9676e-10); Neutrophil count (P=6.0795e-10); Mean sphered cell volume (P=9.3161e-08).	DEGS1
rs55664906	X	Mean platelet (thrombocyte) volume (P=5.3355e-16); Red blood cell (erythrocyte) count (P=5.5676e-10); White blood cell (leukocyte) count (P=1.6824e-09); Neutrophil count (P=2.3745e-09); Mean sphered cell volume (P=6.4397e-08).	DEGS1
rs12076788	X	Mean platelet (thrombocyte) volume (P=2.1908e-16); White blood cell (leukocyte) count (P=7.174e-10); Red blood cell (erythrocyte) count (P=1.3034e-09); Neutrophil count (P=1.6361e-09).	DEGS1
rs4653563	X	Mean platelet (thrombocyte) volume (P=9.2972e-17); White blood cell (leukocyte) count (P=4.6256e-10); Neutrophil count (P=1.5481e-09); Red blood cell (erythrocyte) count (P=2.0336e-09).	DEGS1
rs4653564	X	Mean platelet (thrombocyte) volume (P=8.4965e-17); White blood cell (leukocyte) count (P=4.9323e-10); Neutrophil count (P=1.5762e-09); Red blood cell (erythrocyte) count (P=1.8868e-09).	DEGS1
rs6426139	X	Mean platelet (thrombocyte) volume (P=8.978e-17); White blood cell (leukocyte) count (P=4.5862e-10); Neutrophil count (P=1.5587e-09); Red blood cell (erythrocyte) count (P=2.0271e-09).	DEGS1

rs10753454	X	Mean platelet (thrombocyte) volume (P=1.0592e-16); White blood cell (leukocyte) count (P=4.7645e-10); Neutrophil count (P=1.5892e-09); Red blood cell (erythrocyte) count (P=1.8429e-09).	DEGS1
rs7542293	X	Mean platelet (thrombocyte) volume (P=1.0092e-16); White blood cell (leukocyte) count (P=4.5228e-10); Neutrophil count (P=1.5183e-09); Red blood cell (erythrocyte) count (P=1.8065e-09).	DEGS1
rs7542474	X	Mean platelet (thrombocyte) volume (P=9.8213e-17); White blood cell (leukocyte) count (P=4.5324e-10); Neutrophil count (P=1.5489e-09); Red blood cell (erythrocyte) count (P=1.8863e-09).	DEGS1
rs6698041	X	Mean platelet (thrombocyte) volume (P=1.0606e-16); White blood cell (leukocyte) count (P=4.7114e-10); Neutrophil count (P=1.5776e-09); Red blood cell (erythrocyte) count (P=1.8254e-09).	DEGS1
rs7519433	X	Mean platelet (thrombocyte) volume (P=9.8438e-17); White blood cell (leukocyte) count (P=3.8904e-10); Neutrophil count (P=1.2884e-09); Red blood cell (erythrocyte) count (P=2.125e-09).	DEGS1
rs12730611	X	Mean platelet (thrombocyte) volume (P=1.0549e-16); White blood cell (leukocyte) count (P=4.5558e-10); Neutrophil count (P=1.5474e-09); Red blood cell (erythrocyte) count (P=1.7702e-09).	DEGS1
rs2014782	X	Mean platelet (thrombocyte) volume (P=1.037e-16); White blood cell (leukocyte) count (P=3.6376e-10); Neutrophil count (P=7.8676e-10); Red blood cell (erythrocyte) count (P=1.4462e-09).	DEGS1
rs869945	X	Mean platelet (thrombocyte) volume (P=1.0322e-16); White blood cell (leukocyte) count (P=3.5835e-10); Neutrophil count (P=7.7447e-10); Red blood cell (erythrocyte) count (P=1.4451e-09).	DEGS1
rs6691405	X	Mean platelet (thrombocyte) volume (P=8.4246e-17); White blood cell (leukocyte) count (P=4.2562e-10); Neutrophil count (P=1.469e-09); Red blood cell (erythrocyte) count (P=2.0154e-09).	DEGS1
rs6685783	X	Mean platelet (thrombocyte) volume (P=9.2798e-17); White blood cell (leukocyte) count (P=4.0593e-10); Neutrophil count (P=1.4225e-09); Red blood cell (erythrocyte) count (P=2.0518e-09).	DEGS1
rs10916355	X	Mean platelet (thrombocyte) volume (P=8.8465e-17); White blood cell (leukocyte) count (P=3.6897e-10); Neutrophil count (P=7.8979e-10); Red blood cell (erythrocyte) count (P=1.3939e-09).	DEGS1
rs66506223	X	Mean platelet (thrombocyte) volume (P=8.4007e-17); White blood cell (leukocyte) count (P=4.1628e-10); Neutrophil count (P=1.4926e-09); Red blood cell (erythrocyte) count (P=2.1213e-09).	DEGS1
rs61827699	X	Mean platelet (thrombocyte) volume (P=8.882e-17); White blood cell	DEGS1

		(leukocyte) count (P=3.9042e-10); Neutrophil count (P=1.3987e-09); Red blood cell (erythrocyte) count (P=2.1437e-09).	
rs10916367	X	Mean platelet (thrombocyte) volume (P=5.0835e-17); White blood cell (leukocyte) count (P=3.555e-10); Neutrophil count (P=1.21e-09); Red blood cell (erythrocyte) count (P=1.8198e-09).	DEGS1
rs4653986	X	Mean platelet (thrombocyte) volume (P=4.6912e-17); White blood cell (leukocyte) count (P=3.513e-10); Neutrophil count (P=1.1833e-09); Red blood cell (erythrocyte) count (P=1.8124e-09).	DEGS1
rs7518839	X	Mean platelet (thrombocyte) volume (P=4.9221e-17); White blood cell (leukocyte) count (P=3.4999e-10); Neutrophil count (P=1.1542e-09); Red blood cell (erythrocyte) count (P=1.8229e-09).	DEGS1
rs13374070	X	Mean platelet (thrombocyte) volume (P=4.8718e-17); White blood cell (leukocyte) count (P=2.794e-10); Neutrophil count (P=6.354e-10); Red blood cell (erythrocyte) count (P=1.0861e-09).	DEGS1
rs55736782	X	Mean platelet (thrombocyte) volume (P=4.8498e-17); White blood cell (leukocyte) count (P=2.7262e-10); Neutrophil count (P=6.1833e-10); Red blood cell (erythrocyte) count (P=1.0713e-09).	DEGS1
rs997297	X	Mean platelet (thrombocyte) volume (P=3.7928e-17); White blood cell (leukocyte) count (P=3.6642e-10); Neutrophil count (P=1.3338e-09); Red blood cell (erythrocyte) count (P=1.8344e-09).	DEGS1
rs997296	X	Mean platelet (thrombocyte) volume (P=3.1302e-17); White blood cell (leukocyte) count (P=3.2187e-10); Neutrophil count (P=1.1007e-09); Red blood cell (erythrocyte) count (P=1.5943e-09).	DEGS1
rs7531891	X	Mean platelet (thrombocyte) volume (P=6.0753e-17); White blood cell (leukocyte) count (P=2.5459e-10); Neutrophil count (P=5.6997e-10); Red blood cell (erythrocyte) count (P=9.9016e-10).	DEGS1
rs1492694	X	Mean platelet (thrombocyte) volume (P=5.3205e-17); White blood cell (leukocyte) count (P=2.5298e-10); Neutrophil count (P=5.7417e-10); Red blood cell (erythrocyte) count (P=1.0494e-09).	DEGS1
rs4653991	X	Mean platelet (thrombocyte) volume (P=5.0426e-17); White blood cell (leukocyte) count (P=3.204e-10); Neutrophil count (P=1.0394e-09); Red blood cell (erythrocyte) count (P=1.7487e-09).	DEGS1
rs7546235	X	Mean platelet (thrombocyte) volume (P=5.2812e-17); White blood cell (leukocyte) count (P=2.4989e-10); Neutrophil count (P=5.7025e-10); Red blood cell (erythrocyte) count (P=1.0533e-09).	DEGS1
rs10916371	X	Mean platelet (thrombocyte) volume (P=4.6957e-17); White blood cell (leukocyte) count (P=3.03e-10); Neutrophil count (P=9.7973e-10);	DEGS1

		Red blood cell (erythrocyte) count (P=1.7318e-09).	
rs7524705	X	Mean platelet (thrombocyte) volume (P=4.7556e-17); White blood cell (leukocyte) count (P=2.4762e-10); Neutrophil count (P=5.7017e-10); Red blood cell (erythrocyte) count (P=1.0728e-09).	DEGS1
rs1826421	X	Mean platelet (thrombocyte) volume (P=4.5592e-17); White blood cell (leukocyte) count (P=3.3766e-10); Neutrophil count (P=1.1157e-09); Red blood cell (erythrocyte) count (P=1.733e-09).	DEGS1
rs12563153	X	Mean platelet (thrombocyte) volume (P=4.8579e-17); White blood cell (leukocyte) count (P=2.6386e-10); Neutrophil count (P=6.0969e-10); Red blood cell (erythrocyte) count (P=1.0513e-09).	DEGS1
rs8328	X	Mean platelet (thrombocyte) volume (P=5.1273e-17); White blood cell (leukocyte) count (P=3.3169e-10); Neutrophil count (P=1.1078e-09); Red blood cell (erythrocyte) count (P=1.7271e-09).	DEGS1
rs7526252	X	Mean platelet (thrombocyte) volume (P=4.3542e-17); White blood cell (leukocyte) count (P=3.5044e-10); Neutrophil count (P=1.2329e-09); Red blood cell (erythrocyte) count (P=1.7174e-09).	DEGS1

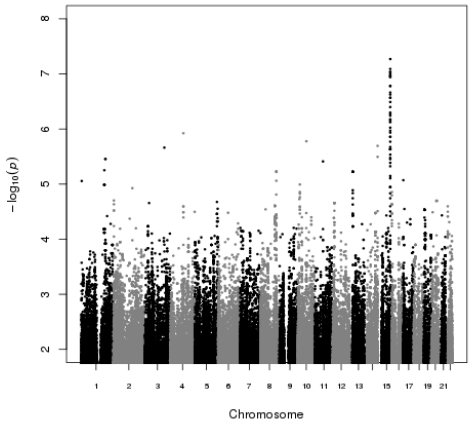
Table 0.13: Results from the 2SMR analysis

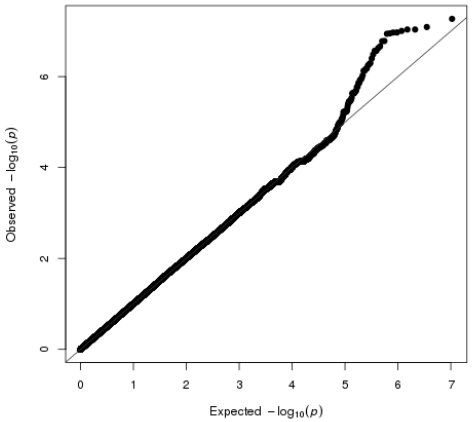
The outcomes depicted are the GWAS of coronary heart disease, type-2 diabetes, and blood cell traits. The ID for the outcome is the ID used in the 2SMR software. The exposure depicts the lipid traits assessed via their GWAS results. All analyses were completed via Wald ratio.

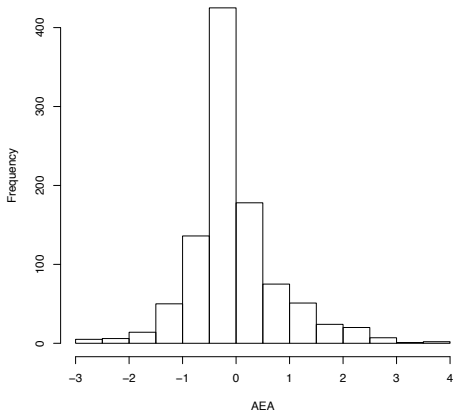
ID	Outcome	Exposure	beta	SE	P	Padj
7	Coronary heart disease	N22S19	-0.02	0.026	4.89E-01	3.47E+01
23	Type 2 diabetes	N22S19	0.03	0.034	4.29E-01	3.04E+01
7	Coronary heart disease	N24S16	0.00	0.037	9.77E-01	6.94E+01
1008	Platelet count	N24S16	-1.82	1.699	2.84E-01	2.02E+01
275	Red blood cell count	N24S16	0.01	0.009	4.78E-01	3.39E+01
1247	Mean platelet volume	N24S16	-0.08	0.014	1.15E-08	8.13E-07
1248	Eosinophil percentage of white cells	N24S16	-0.02	0.013	8.68E-02	6.16E+00
1249	Red blood cell count	N24S16	0.04	0.013	1.05E-03	7.45E-02
1250	Mean corpuscular volume	N24S16	-0.03	0.013	3.53E-02	2.51E+00
1251	Platelet count	N24S16	0.06	0.014	4.72E-05	3.35E-03
1252	Hematocrit	N24S16	0.03	0.013	3.97E-02	2.82E+00
1253	Mean corpuscular hemoglobin concentration	N24S16	0.00	0.013	7.98E-01	5.67E+01
1254	Eosinophil counts	N24S16	-0.01	0.013	4.15E-01	2.95E+01
1255	Plateletcrit	N24S16	0.02	0.014	1.95E-01	1.39E+01
1256	Granulocyte percentage of myeloid white cells	N24S16	0.02	0.013	1.54E-01	1.09E+01
1257	Monocyte percentage of white cells	N24S16	-0.03	0.013	3.34E-02	2.37E+00
1258	White blood cell count	N24S16	0.04	0.014	7.64E-03	5.42E-01
1259	High light scatter reticulocyte count	N24S16	-0.06	0.014	1.22E-05	8.68E-04
1260	High light scatter reticulocyte percentage of red cells	N24S16	-0.07	0.014	1.02E-06	7.22E-05
1261	Sum neutrophil eosinophil counts	N24S16	0.02	0.014	7.28E-02	5.17E+00
1262	Granulocyte count	N24S16	0.03	0.014	5.97E-02	4.24E+00
1263	Hemoglobin concentration	N24S16	0.03	0.013	5.50E-02	3.91E+00
1264	Platelet distribution width	N24S16	-0.02	0.014	2.66E-01	1.89E+01
1265	Eosinophil percentage of granulocytes	N24S16	-0.02	0.013	1.46E-01	1.03E+01
1266	White blood cell count (basophil)	N24S16	0.01	0.013	6.00E-01	4.26E+01
1267	Reticulocyte fraction of red cells	N24S16	-0.06	0.014	6.20E-06	4.40E-04
1268	Sum basophil neutrophil counts	N24S16	0.03	0.014	5.35E-02	3.80E+00
1269	Red cell distribution width	N24S16	0.08	0.013	2.28E-10	1.62E-08
1270	Reticulocyte count	N24S16	-0.05	0.014	2.53E-04	1.80E-02
1271	Neutrophil percentage of granulocytes	N24S16	0.02	0.013	2.11E-01	1.50E+01
1272	Sum eosinophil basophil counts	N24S16	0.00	0.013	7.37E-01	5.23E+01
1273	Monocyte count	N24S16	0.00	0.013	9.16E-01	6.50E+01
1274	Myeloid white cell count	N24S16	0.03	0.014	5.66E-02	4.02E+00
1275	Lymphocyte counts	N24S16	0.05	0.014	3.58E-04	2.54E-02

1276	Immature fraction of reticulocytes	N24S16	-0.05	0.013	9.78E-05	6.94E-03
1277	Neutrophil count	N24S16	0.02	0.013	6.39E-02	4.54E+00
7	Coronary heart disease	N24S19ratio	0.00	0.033	9.55E-01	6.78E+01
1008	Platelet count	N24S19ratio	-3.14	1.535	4.07E-02	2.89E+00
275	Red blood cell count	N24S19ratio	-0.03	0.009	1.12E-03	7.93E-02
1247	Mean platelet volume	N24S19ratio	0.03	0.013	7.93E-03	5.63E-01
1248	Eosinophil percentage of white cells	N24S19ratio	0.00	0.013	7.01E-01	4.98E+01
1249	Red blood cell count	N24S19ratio	-0.02	0.013	5.18E-02	3.68E+00
1250	Mean corpuscular volume	N24S19ratio	0.04	0.012	2.47E-03	1.75E-01
1251	Platelet count	N24S19ratio	-0.02	0.013	1.69E-01	1.20E+01
1252	Hematocrit	N24S19ratio	0.00	0.012	9.62E-01	6.83E+01
1253	Mean corpuscular hemoglobin concentration	N24S19ratio	-0.01	0.012	4.92E-01	3.49E+01
1254	Eosinophil counts	N24S19ratio	-0.02	0.013	2.17E-01	1.54E+01
1255	Plateletcrit	N24S19ratio	0.00	0.013	7.44E-01	5.28E+01
1256	Granulocyte percentage of myeloid white cells	N24S19ratio	-0.02	0.013	2.28E-01	1.62E+01
1257	Monocyte percentage of white cells	N24S19ratio	0.01	0.013	3.44E-01	2.44E+01
1258	White blood cell count	N24S19ratio	-0.03	0.013	8.41E-03	5.97E-01
1259	High light scatter reticulocyte count	N24S19ratio	0.00	0.013	8.25E-01	5.86E+01
1260	High light scatter reticulocyte percentage of red cells	N24S19ratio	0.01	0.013	5.78E-01	4.11E+01
1261	Sum neutrophil eosinophil counts	N24S19ratio	-0.04	0.013	4.25E-03	3.02E-01
1262	Granulocyte count	N24S19ratio	-0.04	0.013	4.37E-03	3.10E-01
1263	Hemoglobin concentration	N24S19ratio	-0.01	0.013	6.75E-01	4.80E+01
1264	Platelet distribution width	N24S19ratio	0.01	0.013	3.11E-01	2.21E+01
1265	Eosinophil percentage of granulocytes	N24S19ratio	0.00	0.013	9.99E-01	7.09E+01
1266	White blood cell count (basophil)	N24S19ratio	-0.03	0.012	1.18E-02	8.41E-01
1267	Reticulocyte fraction of red cells	N24S19ratio	0.00	0.013	9.36E-01	6.65E+01
1268	Sum basophil neutrophil counts	N24S19ratio	-0.03	0.013	6.83E-03	4.85E-01
1269	Red cell distribution width	N24S19ratio	-0.02	0.013	1.52E-01	1.08E+01
1270	Reticulocyte count	N24S19ratio	0.00	0.013	7.29E-01	5.18E+01
1271	Neutrophil percentage of granulocytes	N24S19ratio	0.00	0.013	8.17E-01	5.80E+01
1272	Sum eosinophil basophil counts	N24S19ratio	-0.02	0.013	7.03E-02	4.99E+00
1273	Monocyte count	N24S19ratio	-0.02	0.013	2.00E-01	1.42E+01
1274	Myeloid white cell count	N24S19ratio	-0.04	0.013	3.71E-03	2.64E-01
1275	Lymphocyte counts	N24S19ratio	-0.01	0.013	3.43E-01	2.44E+01
1276	Immature fraction of reticulocytes	N24S19ratio	0.01	0.013	3.72E-01	2.64E+01
1277	Neutrophil count	N24S19ratio	-0.03	0.013	5.85E-03	4.15E-01
7	Coronary heart disease	PEA	-0.05	0.038	1.77E-01	1.25E+01

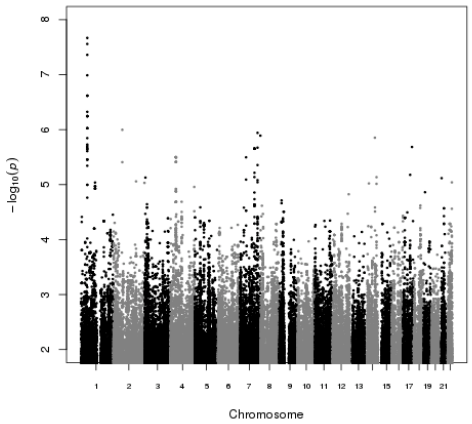
AEA

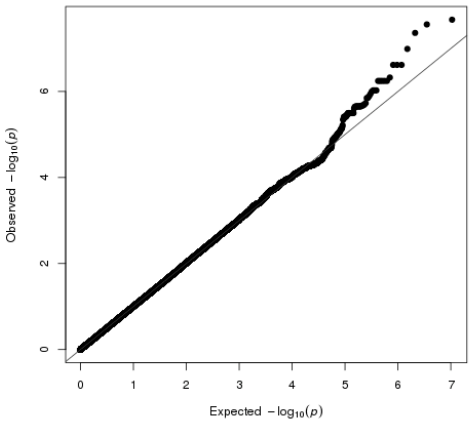

−log10(p)
Chromosome

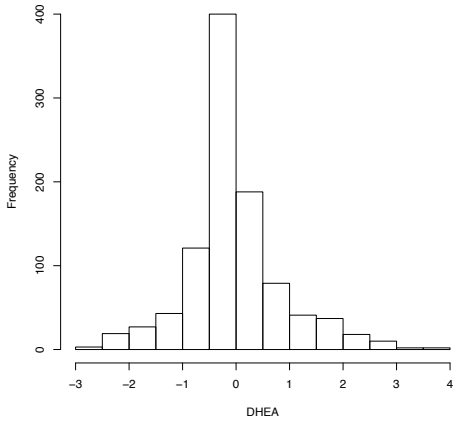

Observed −log10(p)
Expected −log10(p)


Frequency
AEA

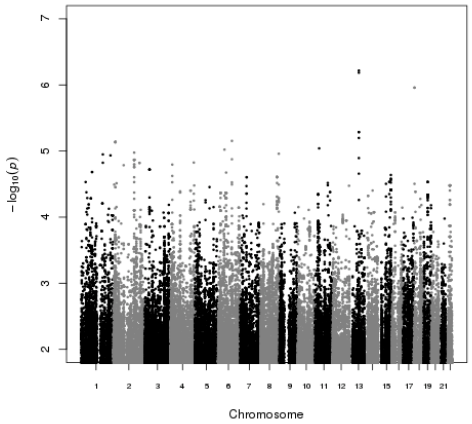
DHEA

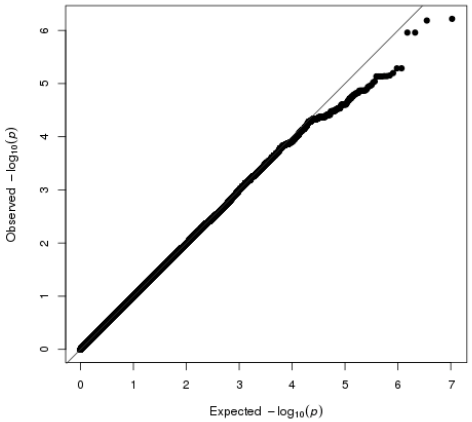

−log10(p)
Chromosome

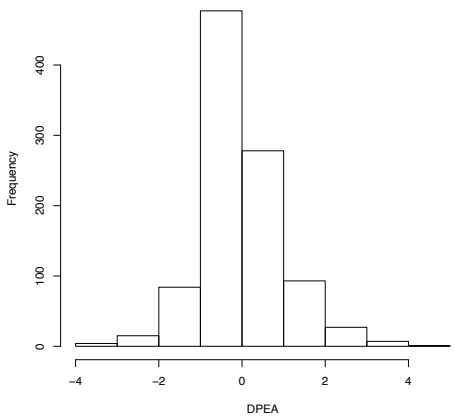

Observed −log10(p)
Expected −log10(p)


Frequency
DHEA

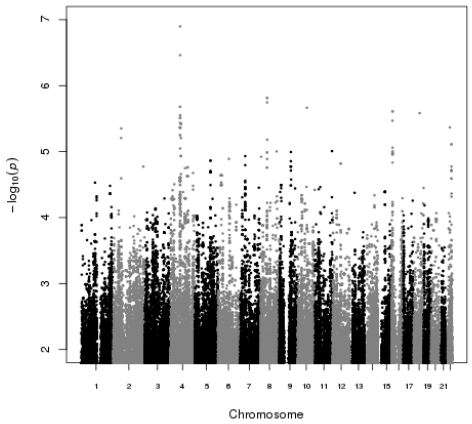
DPEA

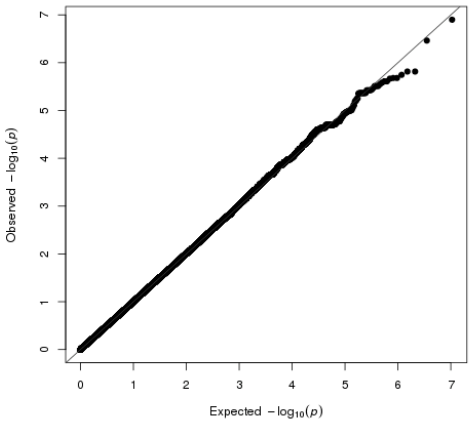

−log10(p)
Chromosome

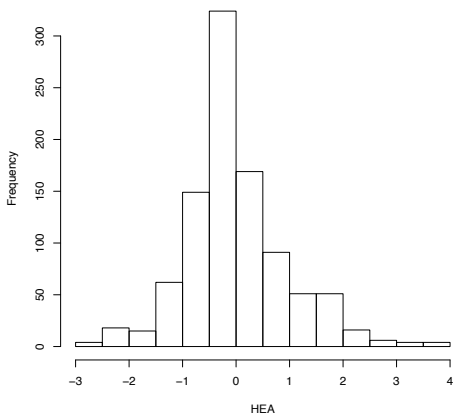

Observed −log10(p)
Expected −log10(p)


Frequency
DPEA

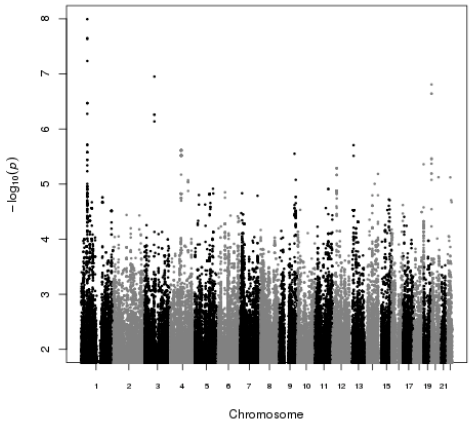
HEA

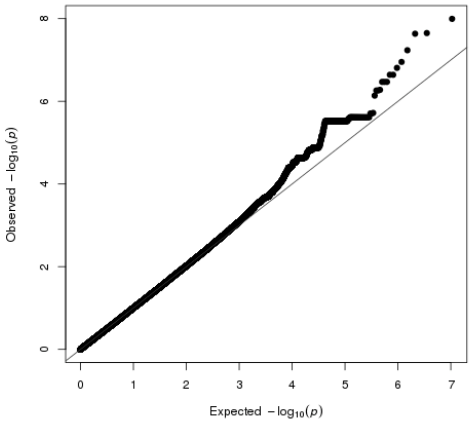

−log10(p)
Chromosome

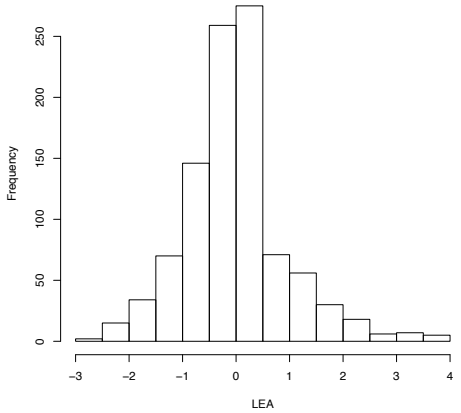

Observed −log10(p)
Expected −log10(p)


Frequency
HEA

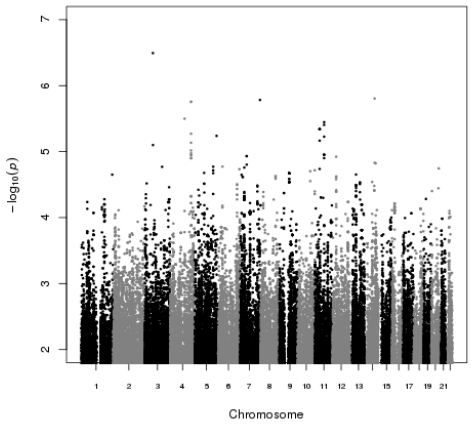
LEA

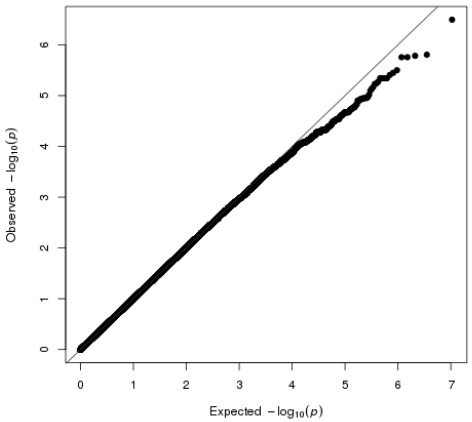

−log10(p)
Chromosome

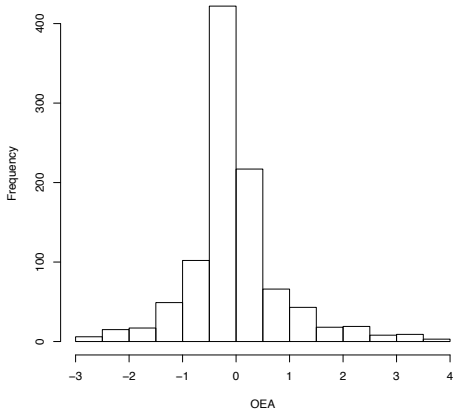

Observed −log10(p)
Expected −log10(p)


Frequency
LEA

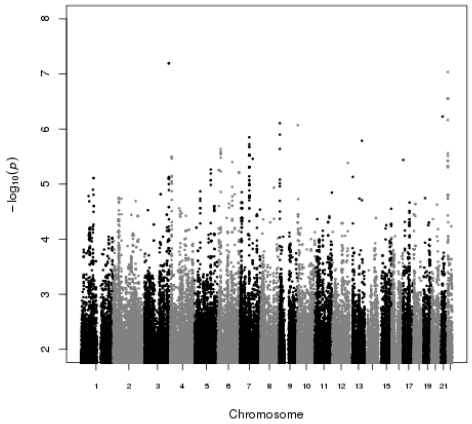
OEA

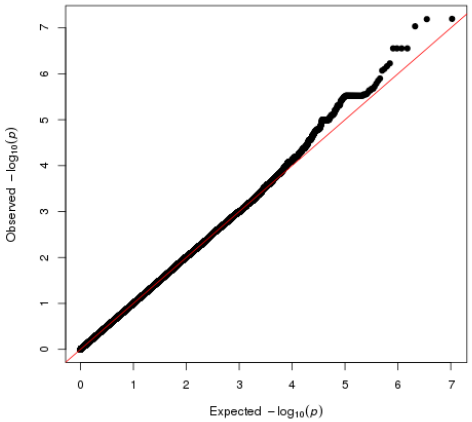

−log10(p)
Chromosome

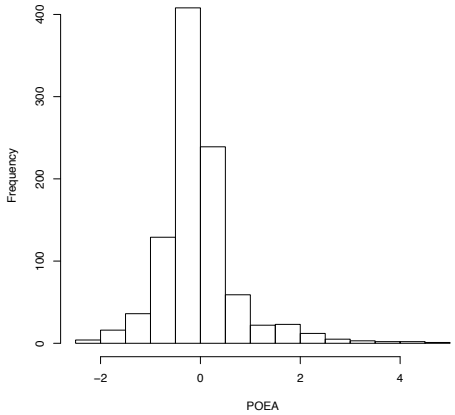

Observed −log10(p)
Expected −log10(p)


Frequency
OEA

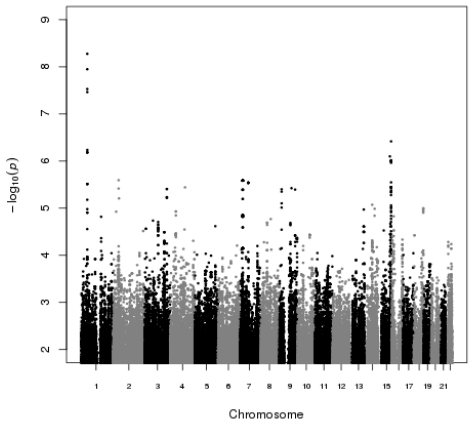
POEA

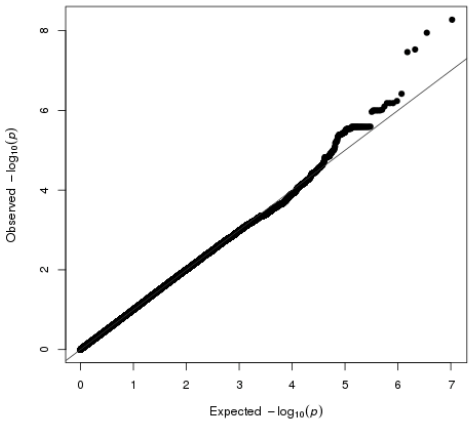

−log10(p)
Chromosome

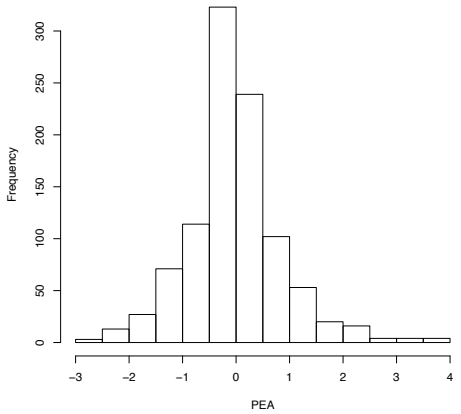

Observed −log10(p)
Expected −log10(p)


Frequency
POEA

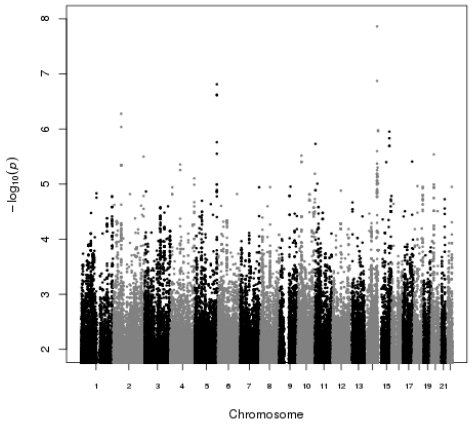
PEA

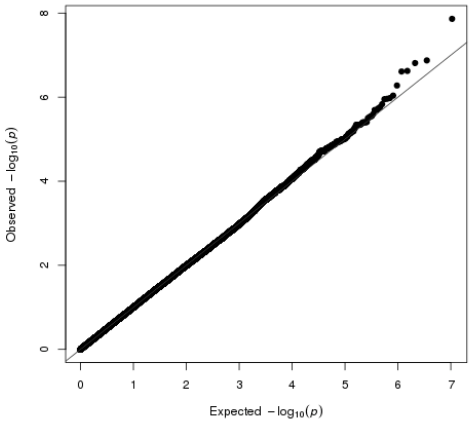

−log10(p)
Chromosome

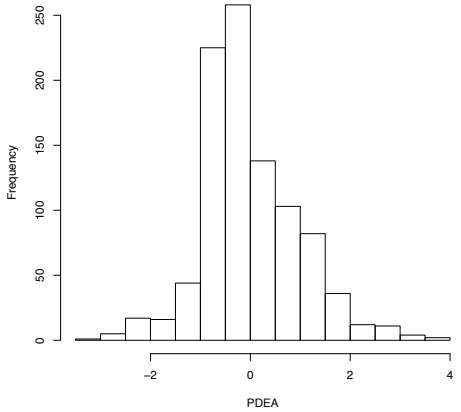

Observed −log10(p)
Expected −log10(p)


Frequency
PEA

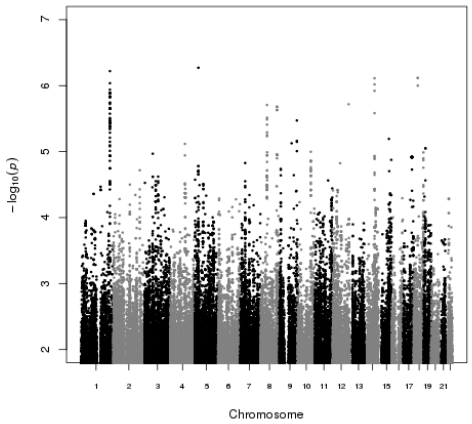
PDEA

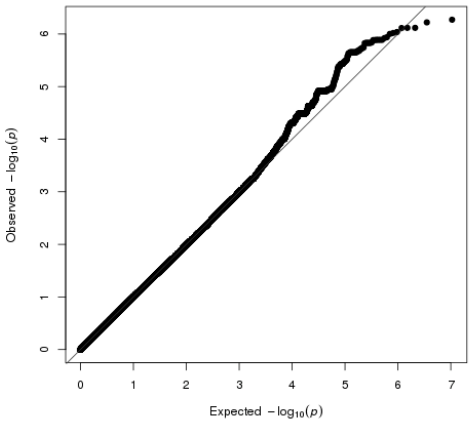

−log10(p)
Chromosome

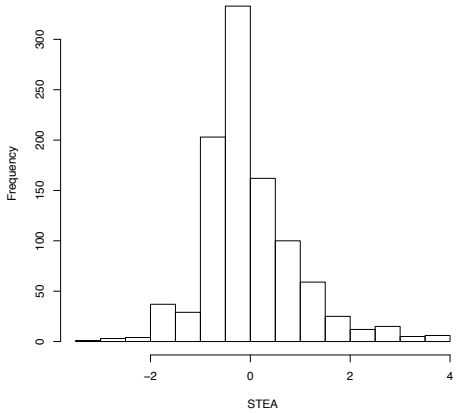

Observed −log10(p)
Expected −log10(p)


Frequency
PDEA

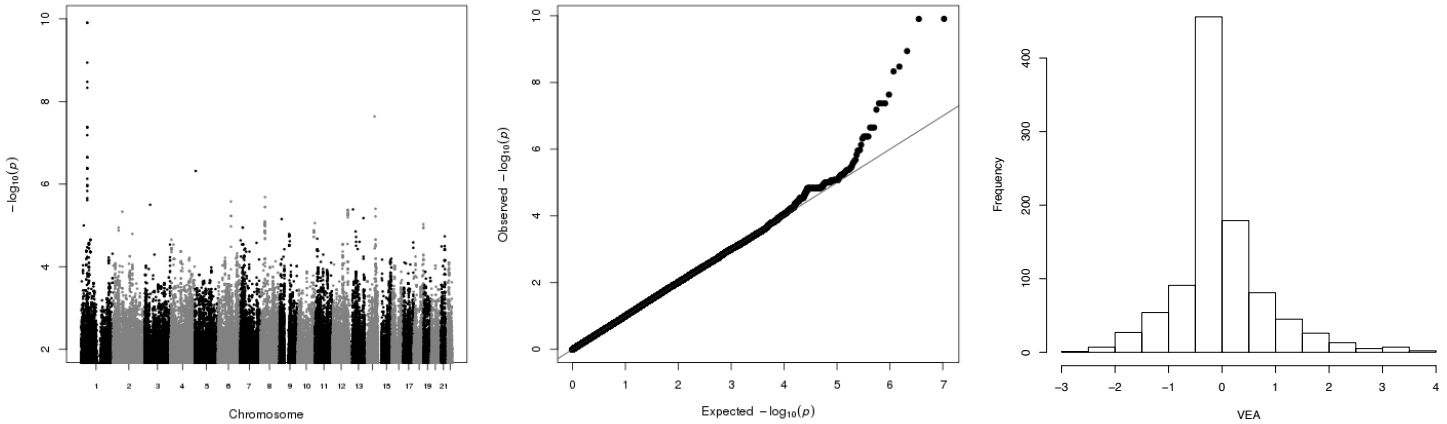
STEA


−log10(p)
Chromosome


Observed −log10(p)
Expected −log10(p)

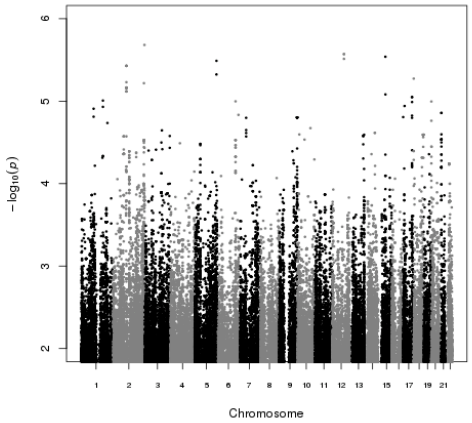

Frequency
STEA

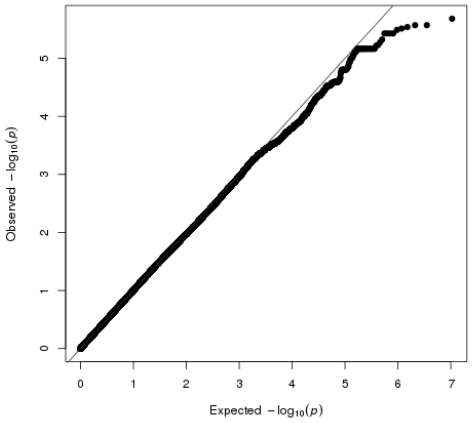
VEA

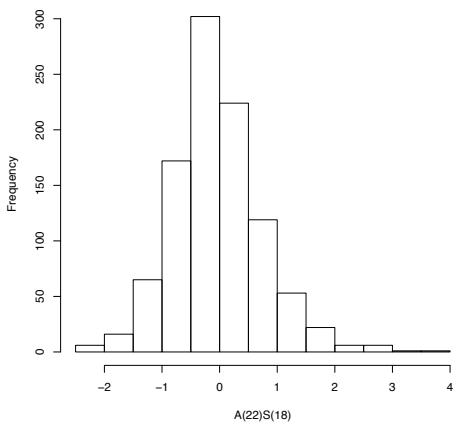


Figures 0-2: Manhattan plots, Quantile-Quantile plots, and trait distributions of the GWAS for N-acylethanolamine species.

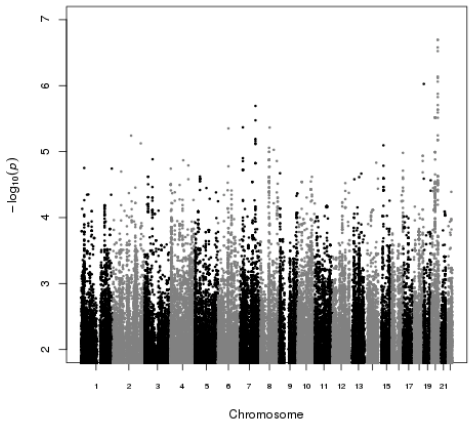
A(22)S(18)

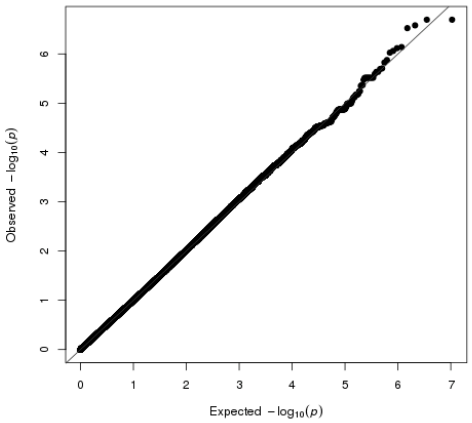


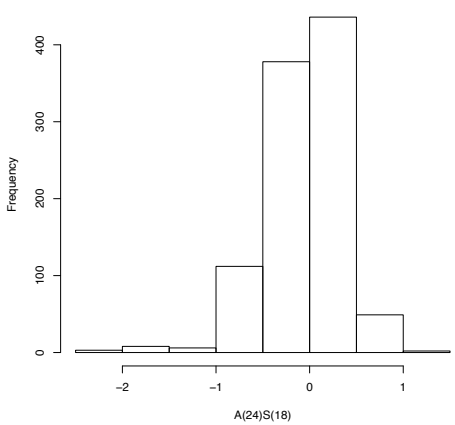




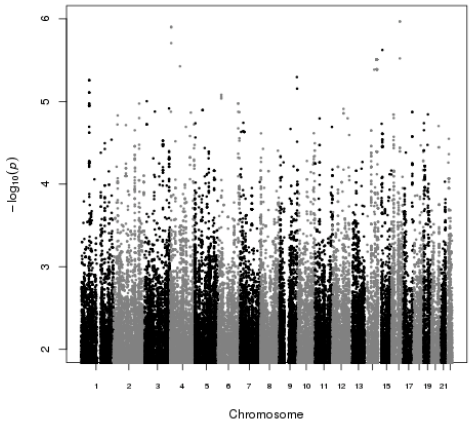
A(24)S(18)

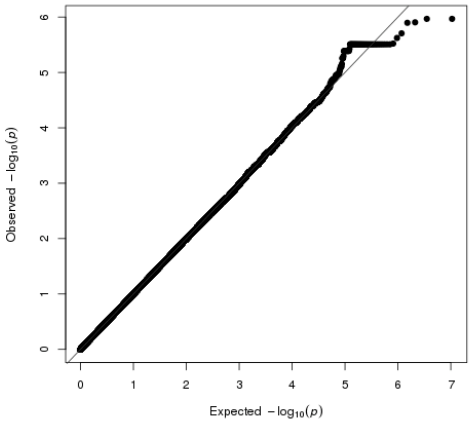


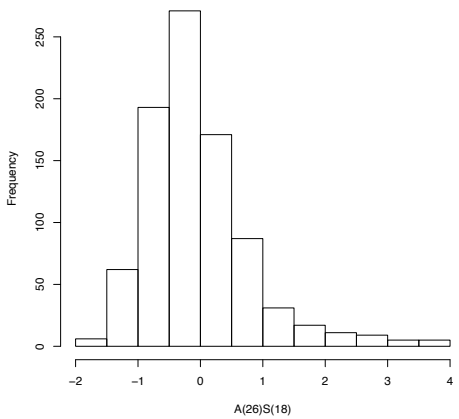




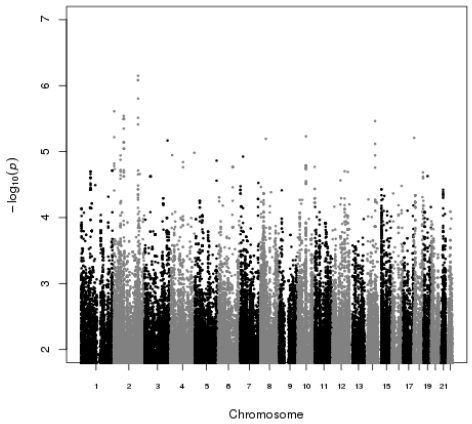
A(26)S(18)

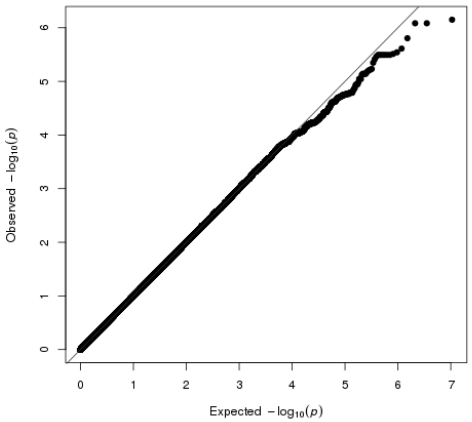


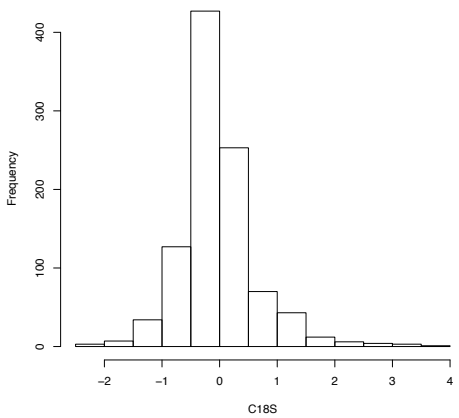




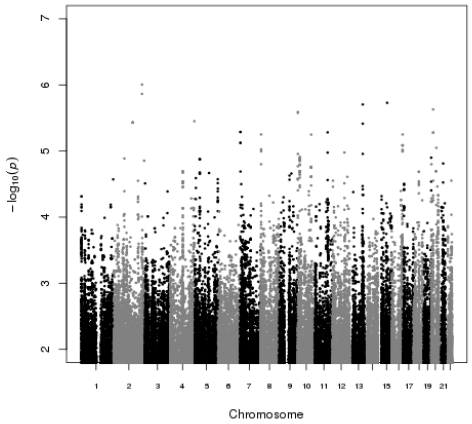
C18DS

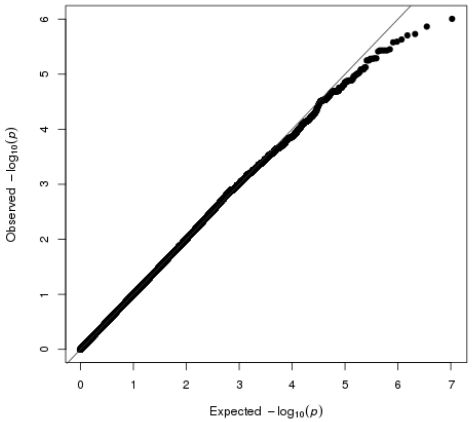


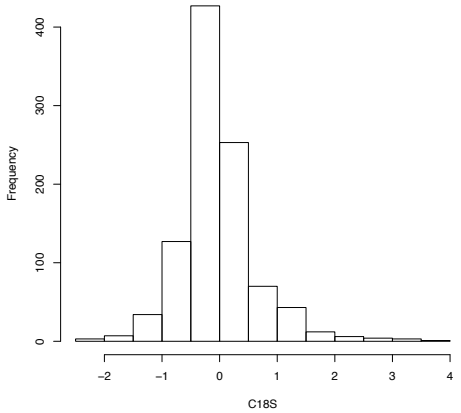




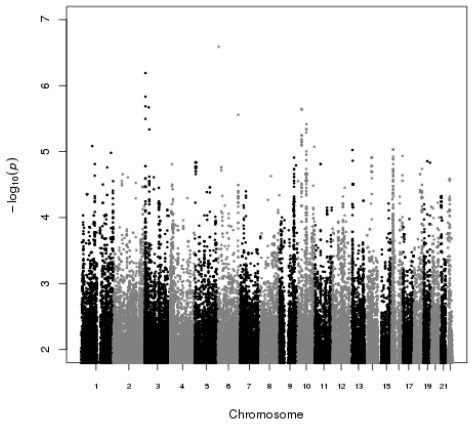
C18S

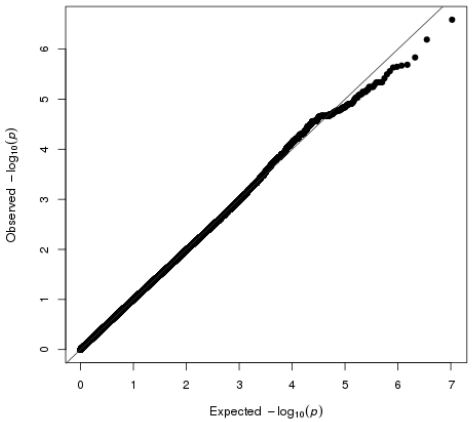

−log10(p)
Chromosome

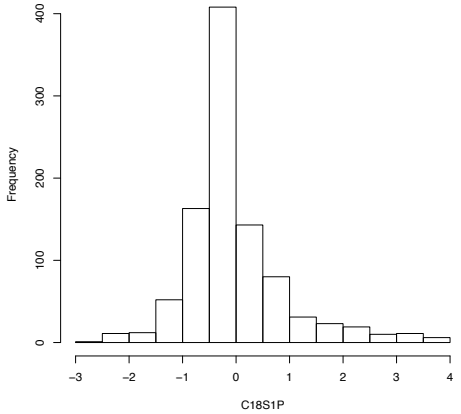

Observed −log10(p)
Expected −log10(p)


Frequency
C18S

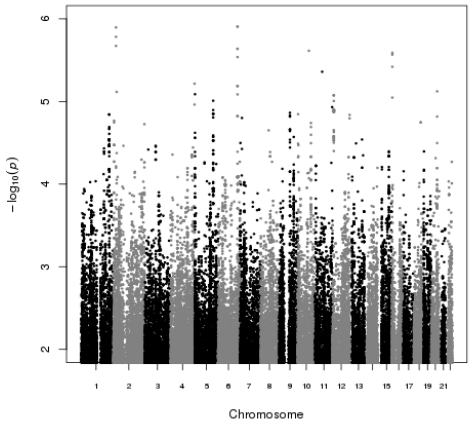
C18S1P

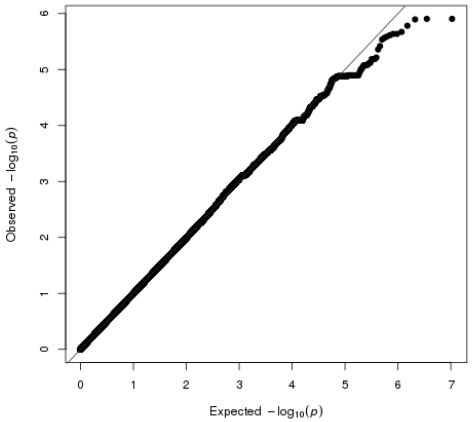

−log10(p)
Chromosome

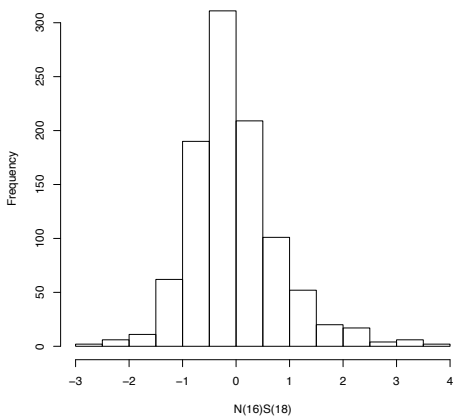

Observed −log10(p)
Expected −log10(p)


Frequency
C18S1P

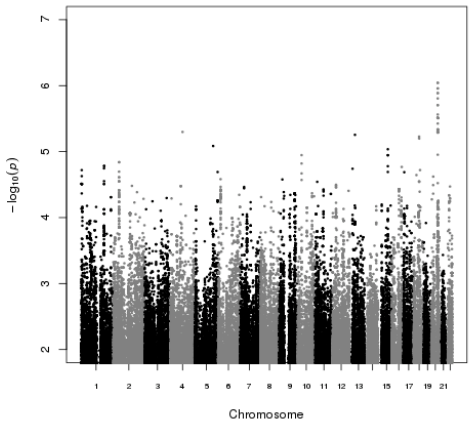
N(16)S(18)

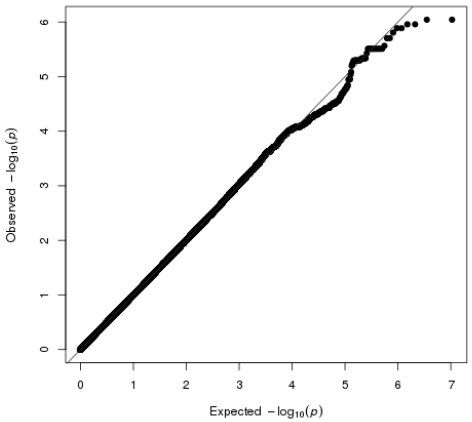


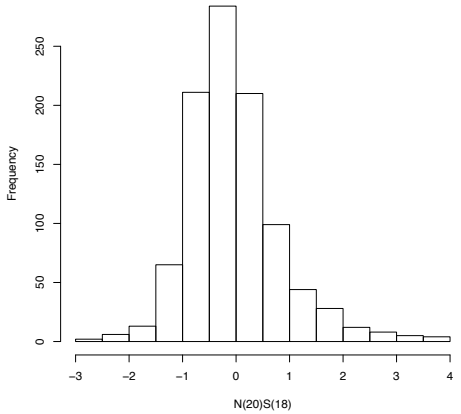




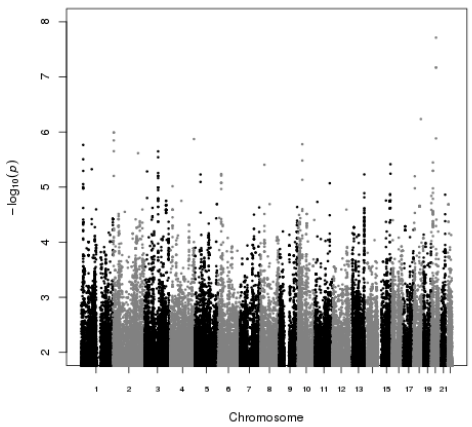
N(20)S(18)

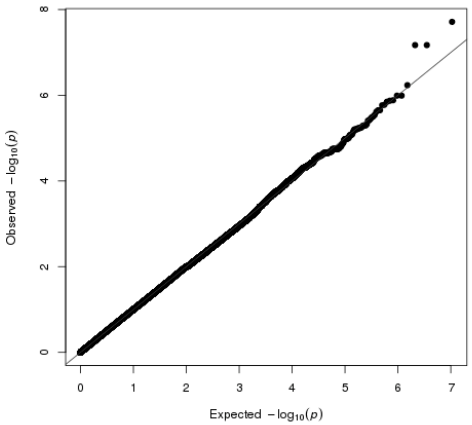


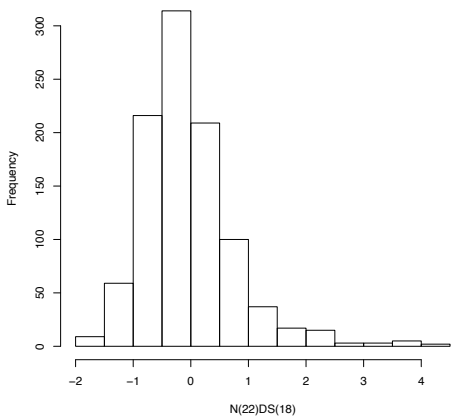




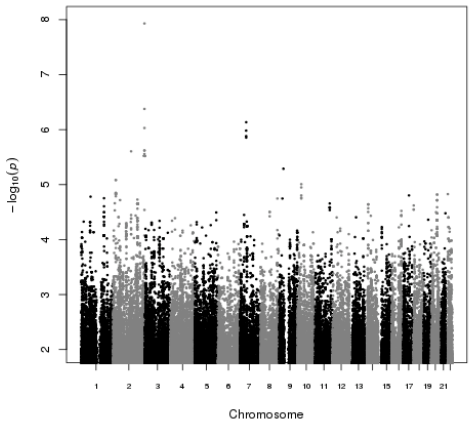
N(22)DS(18)

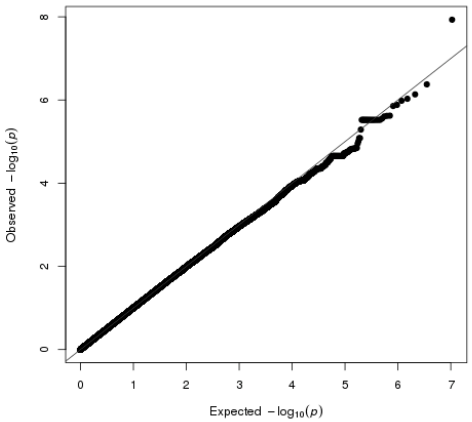


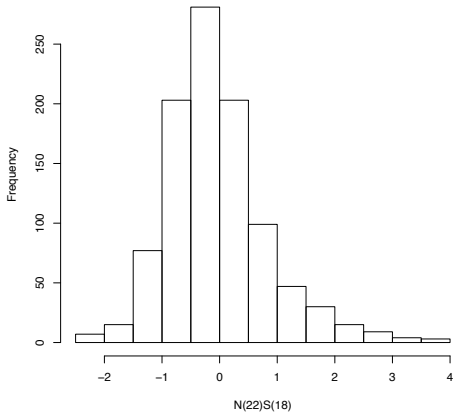




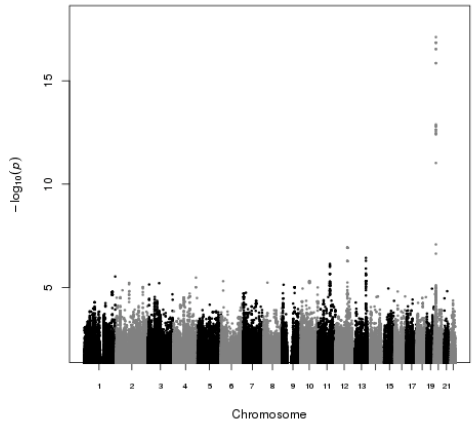
N(22)S(18)

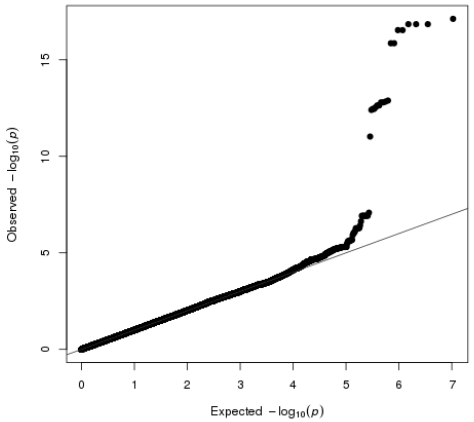


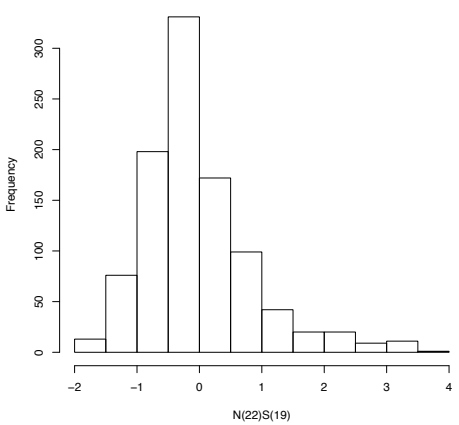




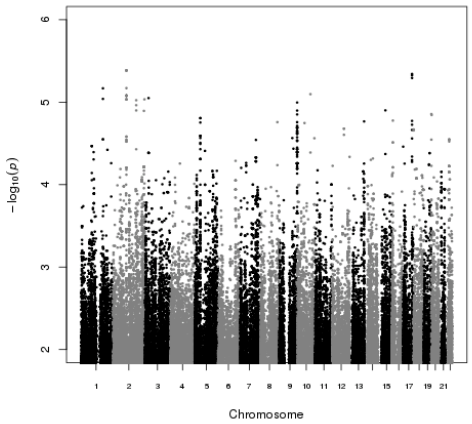
N(22)S(19)

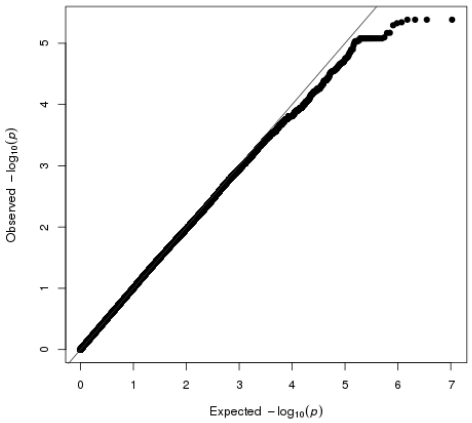


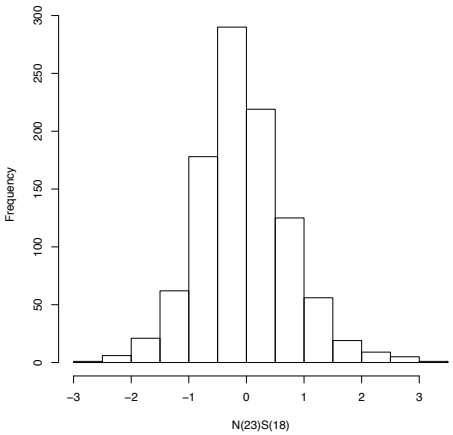




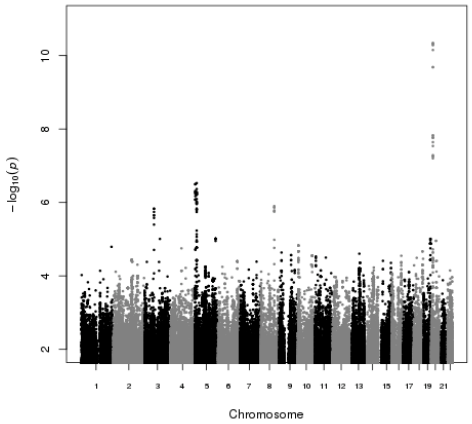
N(23)S(18)

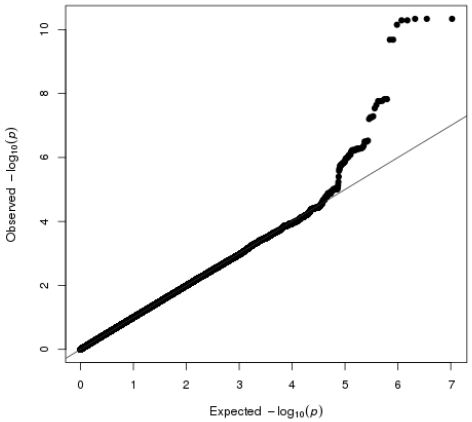


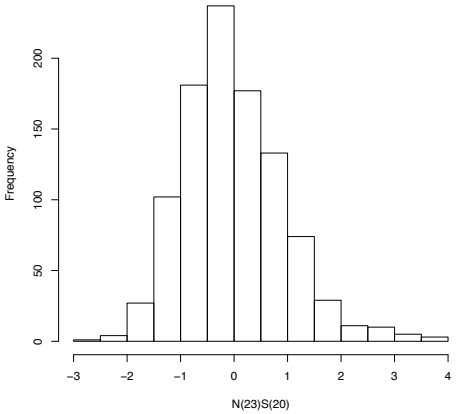




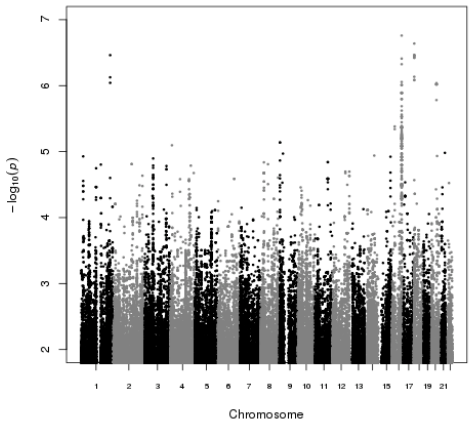
N(23)S(20)

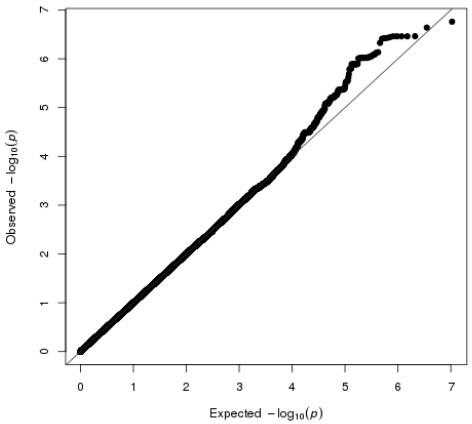


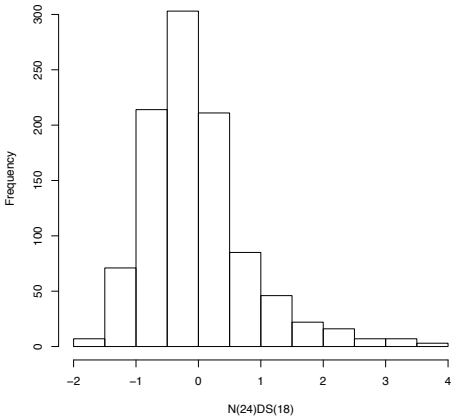




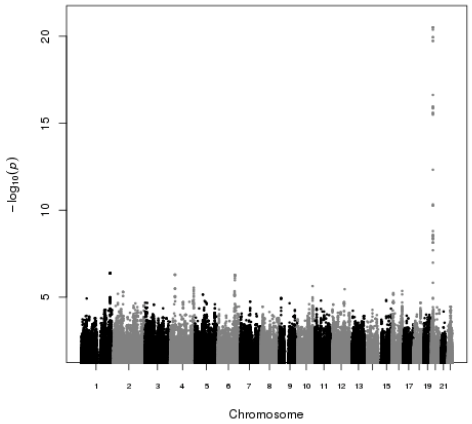
N(24)DS(18)

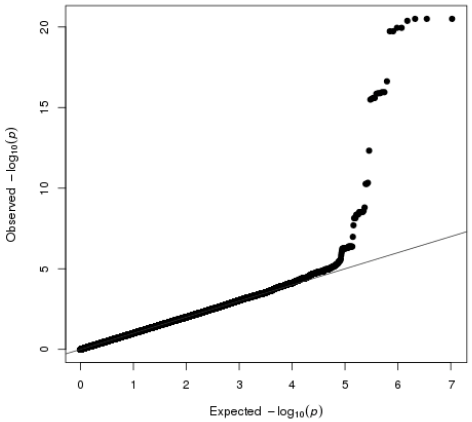


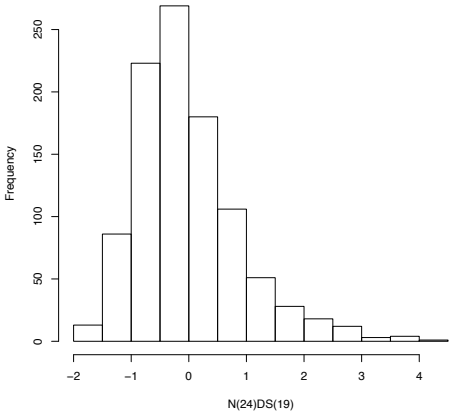




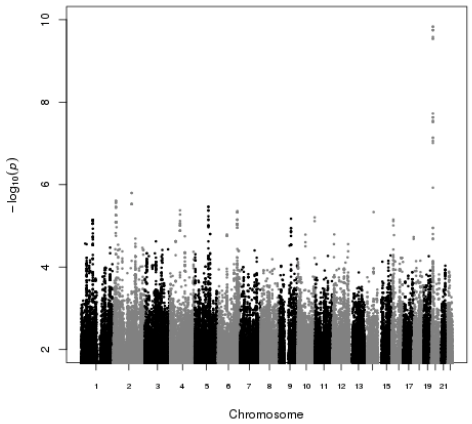
N(24)DS(19)

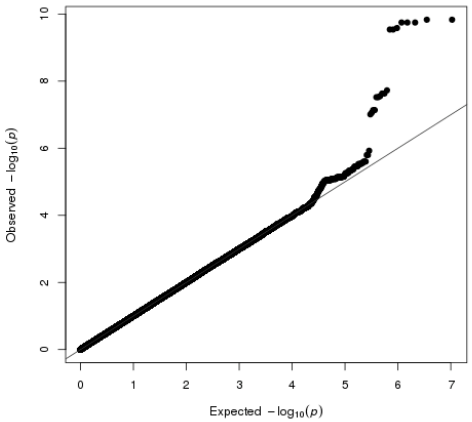


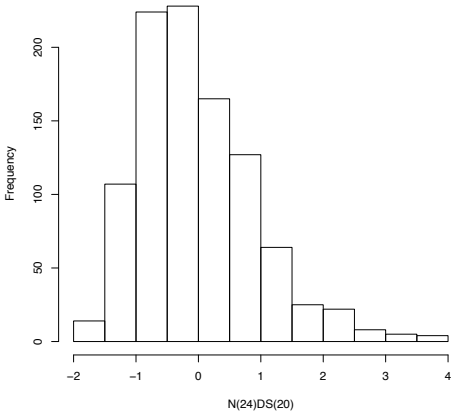




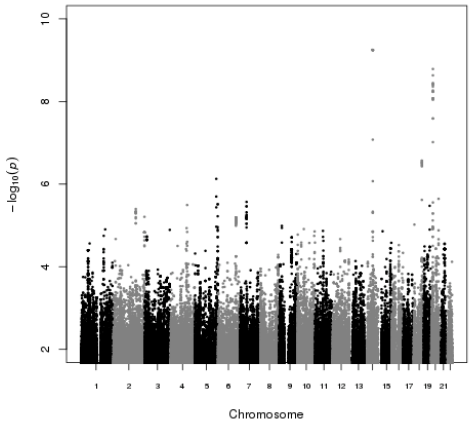
N(24)DS(20)

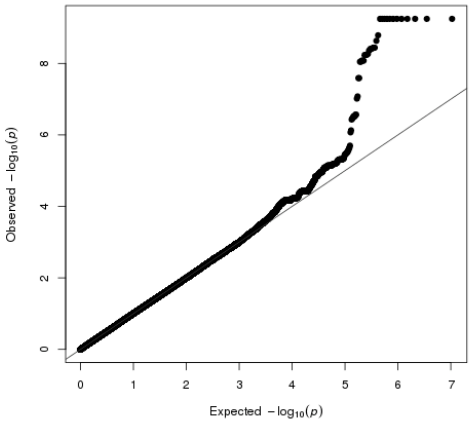


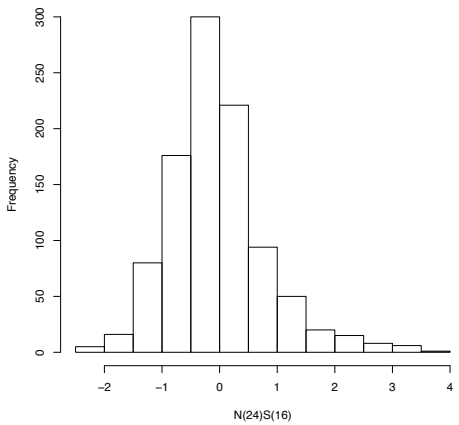




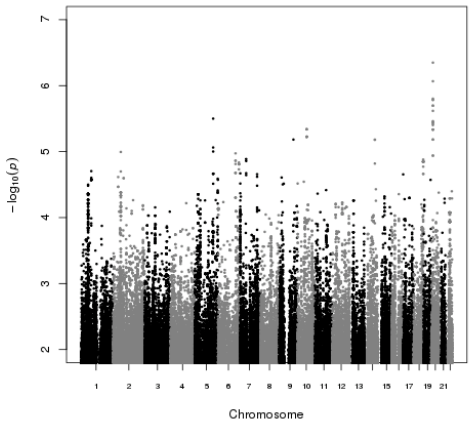
N(24)S(16)

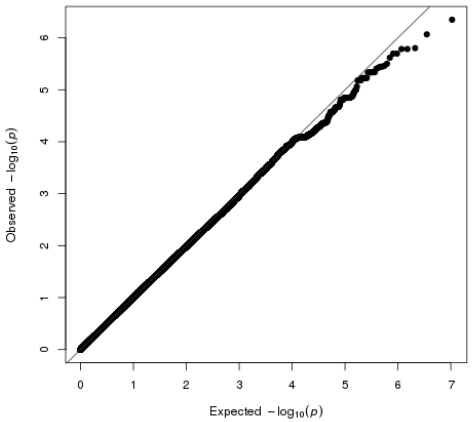


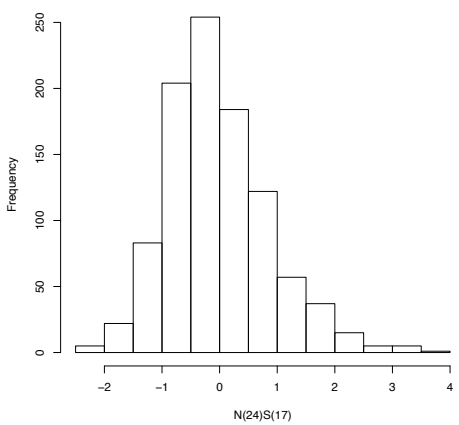




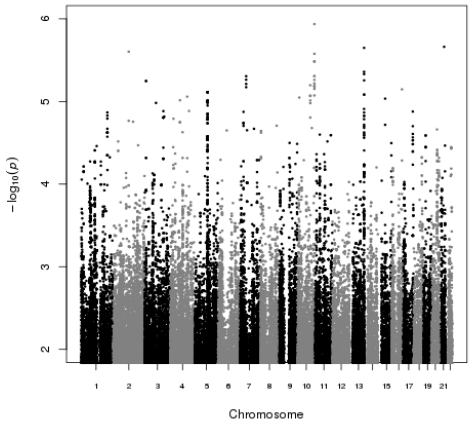
N(24)S(17)

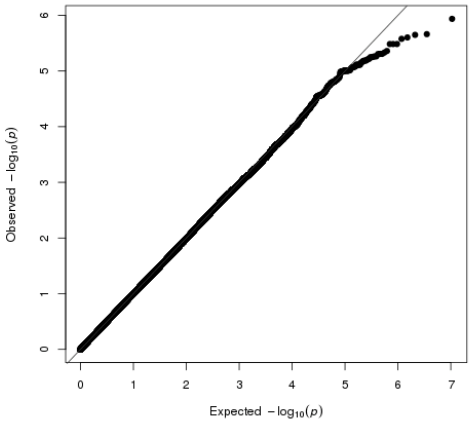


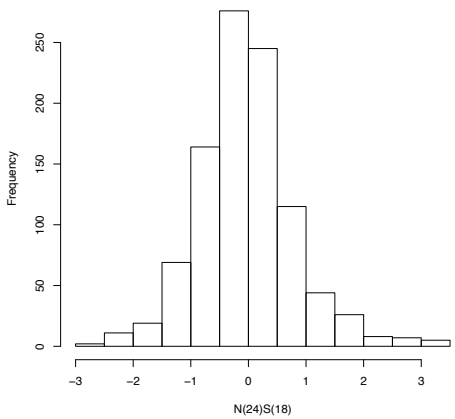




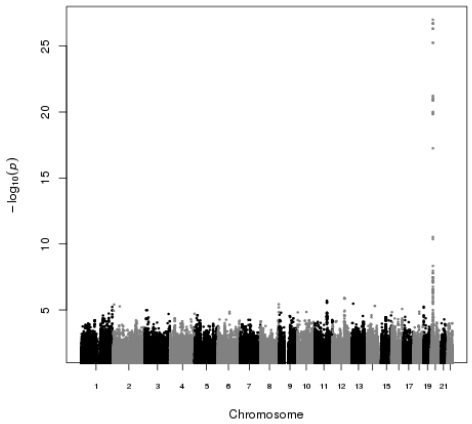
N(24)S(18)

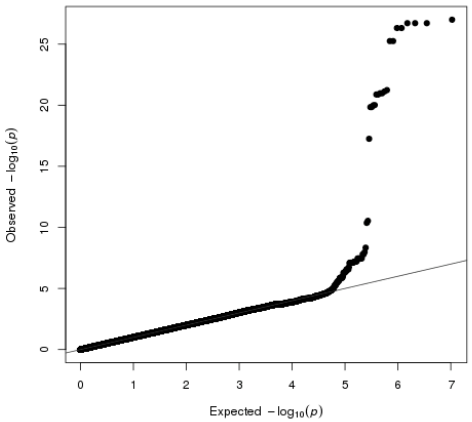


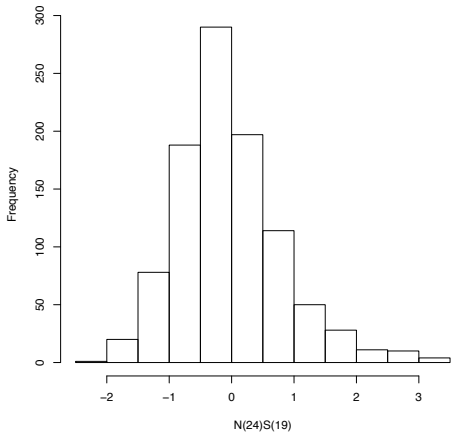




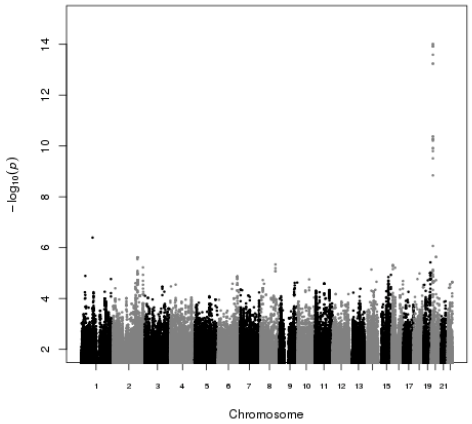
N(24)S(19)

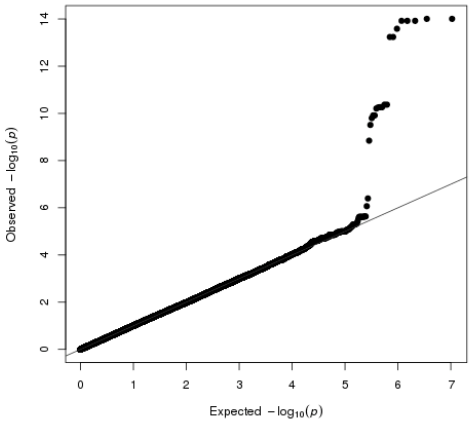


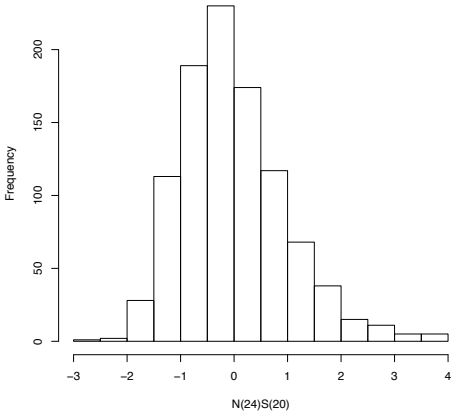




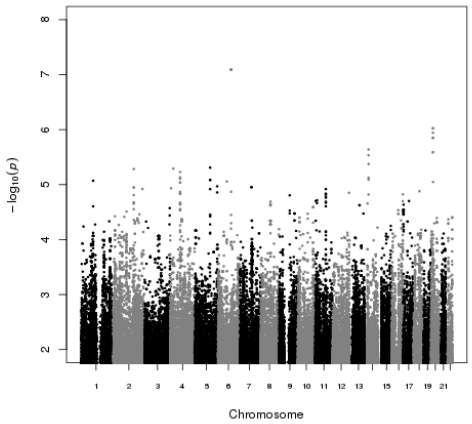
N(24)S(20)

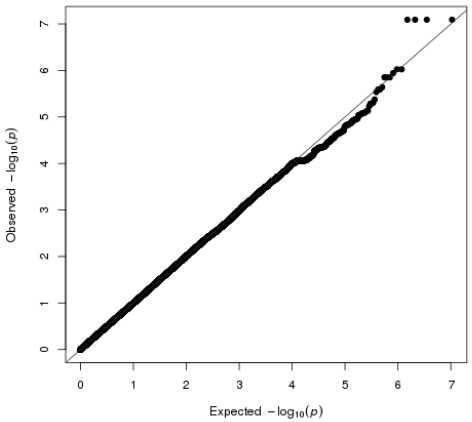


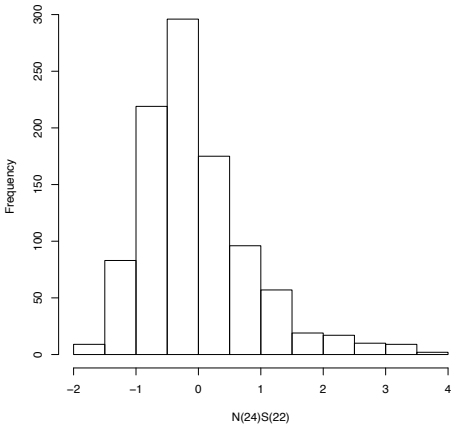




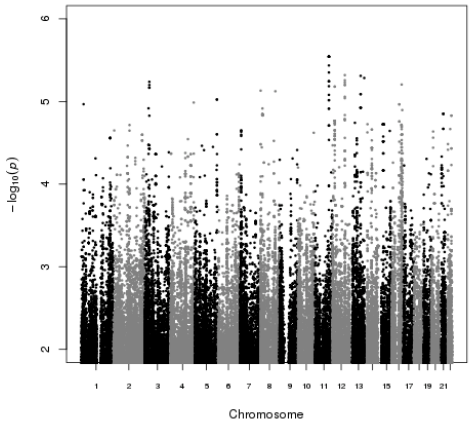
N(24)S(22)

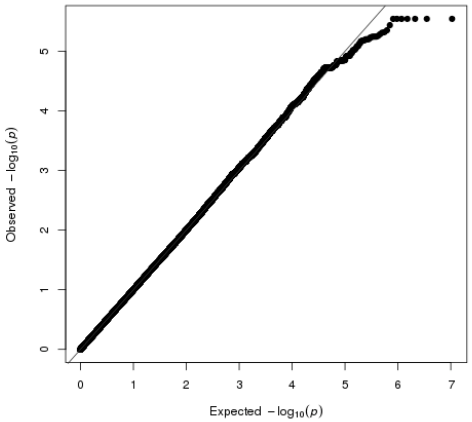


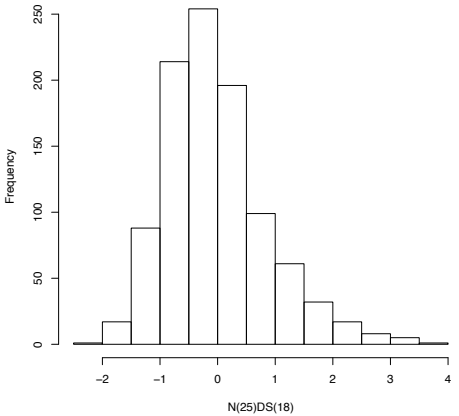




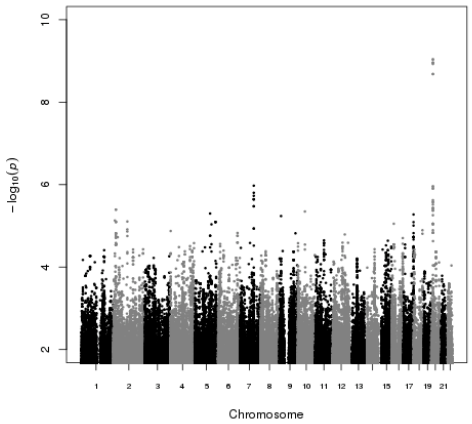
N(25)DS(18)

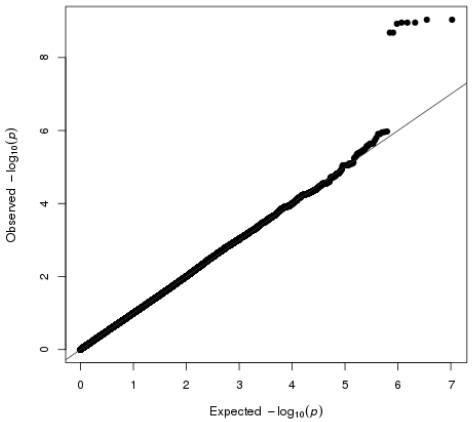


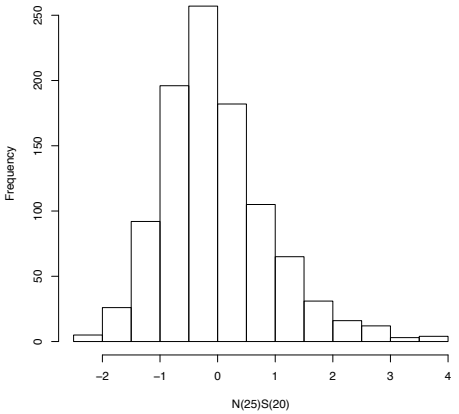




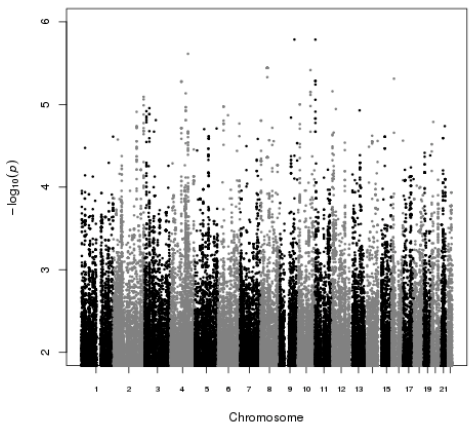
N(25)S(20)

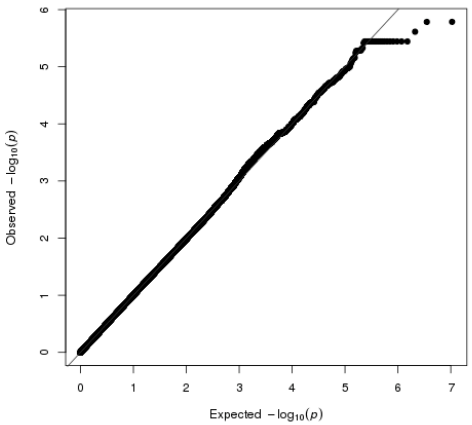


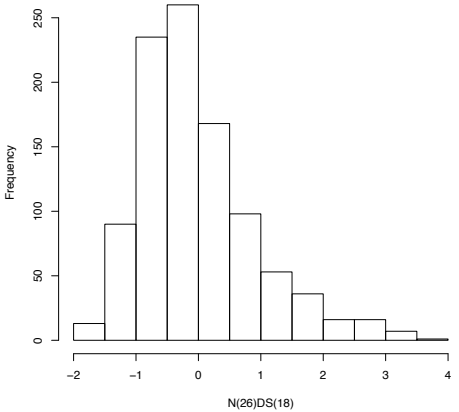




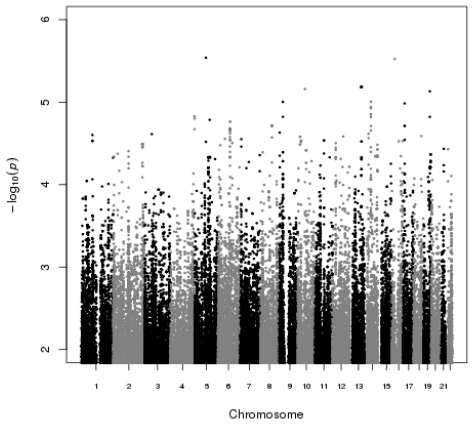
N(26)DS(18)

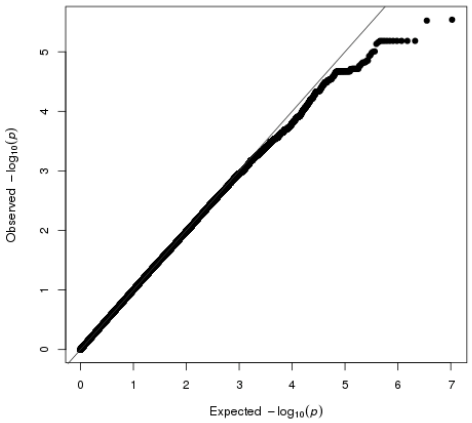


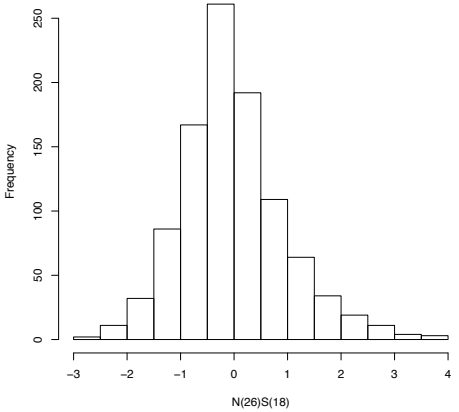




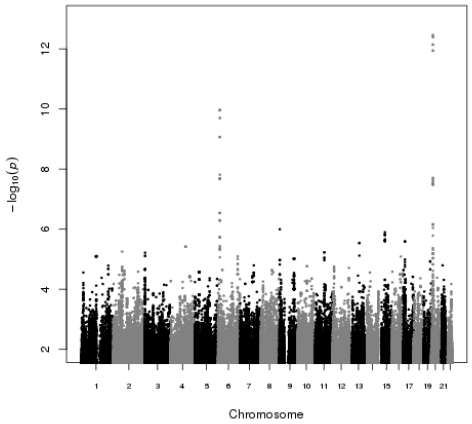
N(26)S(18)

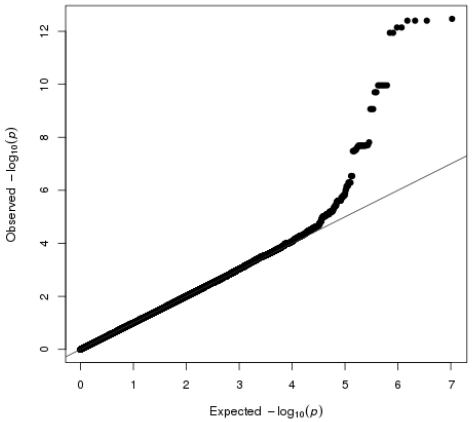


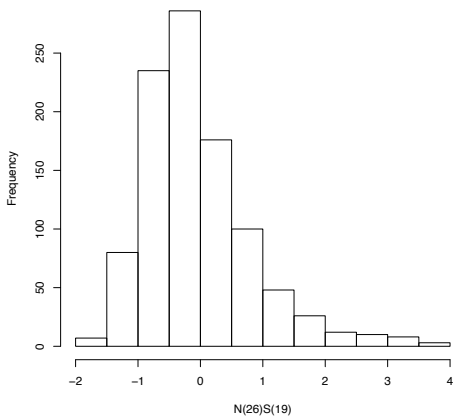




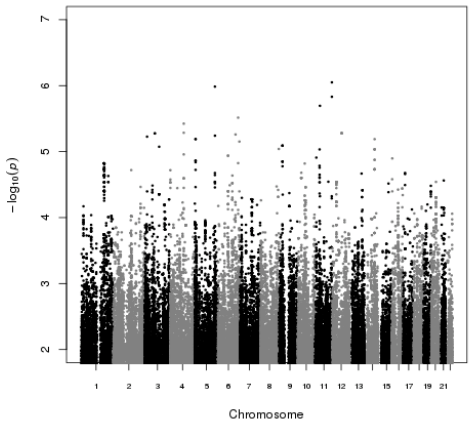
N(26)S(19)

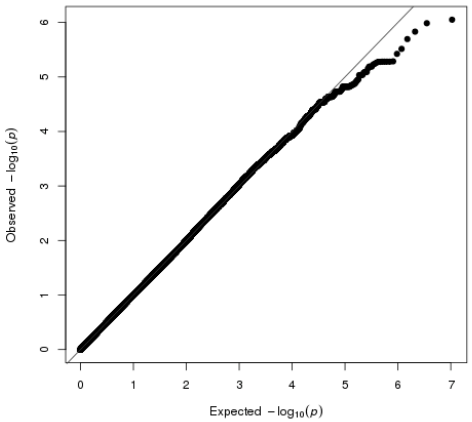


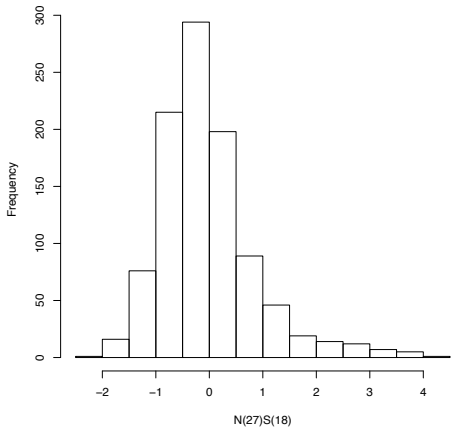




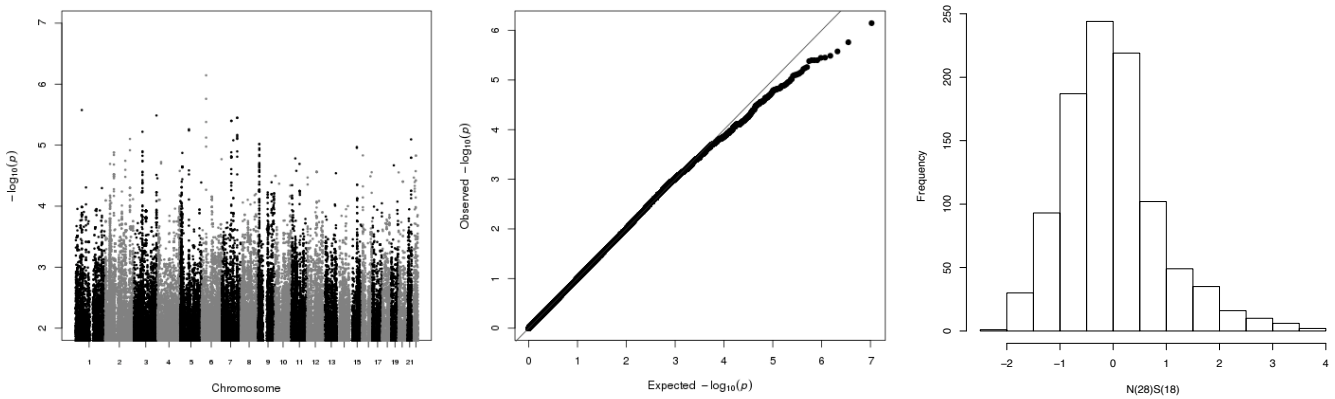
N(27)S(18)







N(28)S(18)



N(29)S(18)

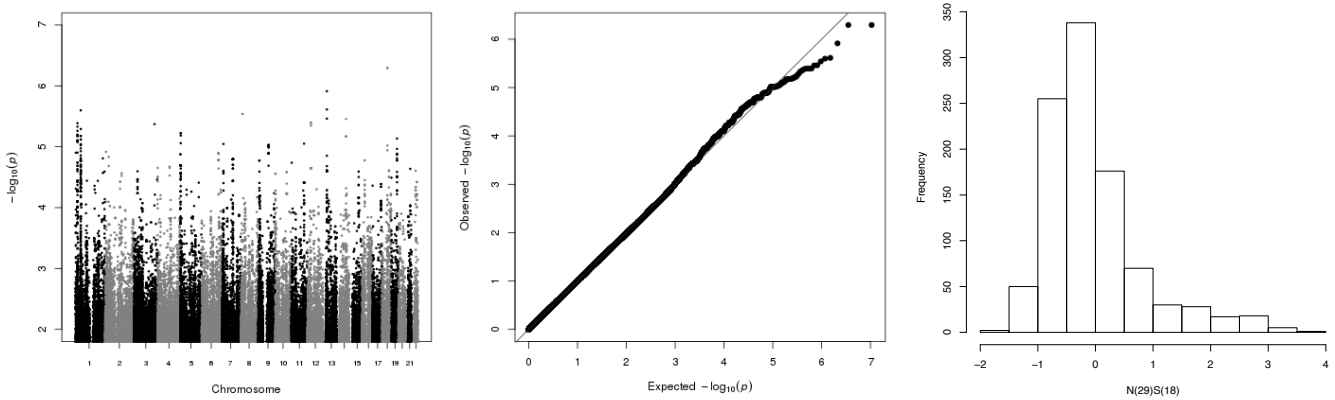
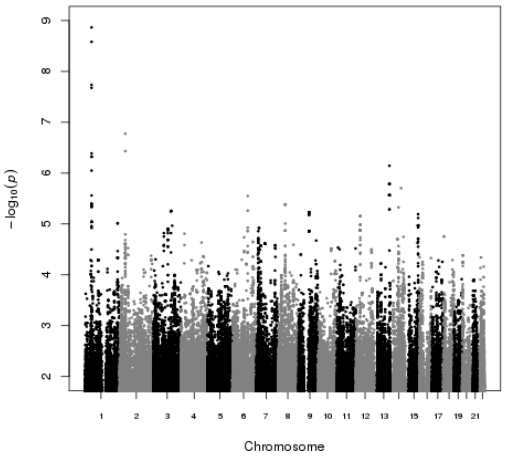
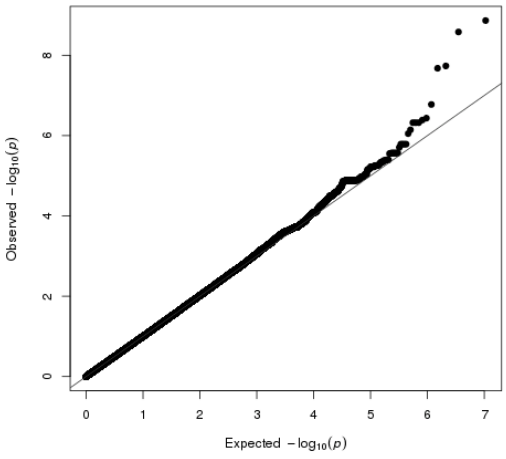


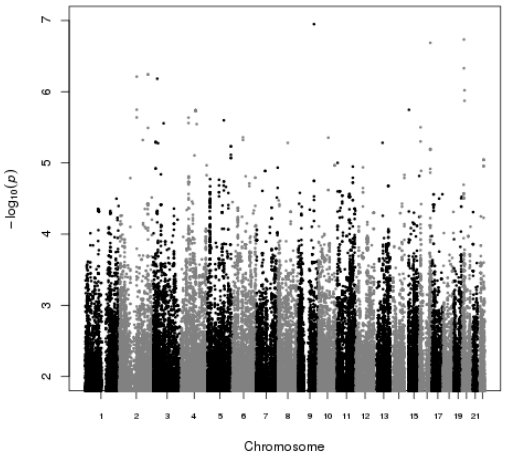
Figure 0-3: Manhattan plots, Quantile-Quantile plots, and trait distributions of the GWAS for ceramides and related species.

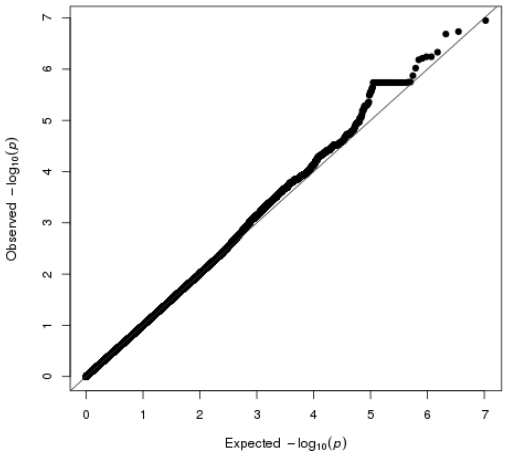
SumEA



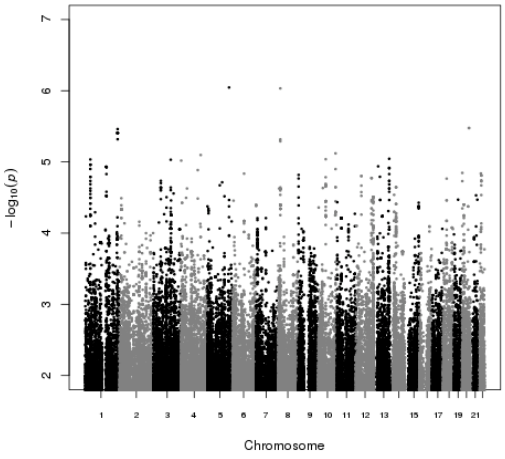


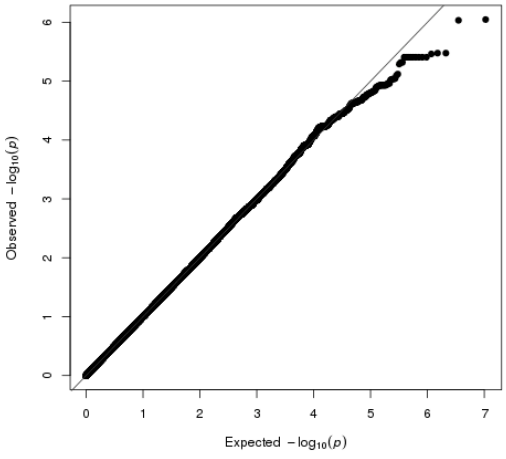
ratio16to24



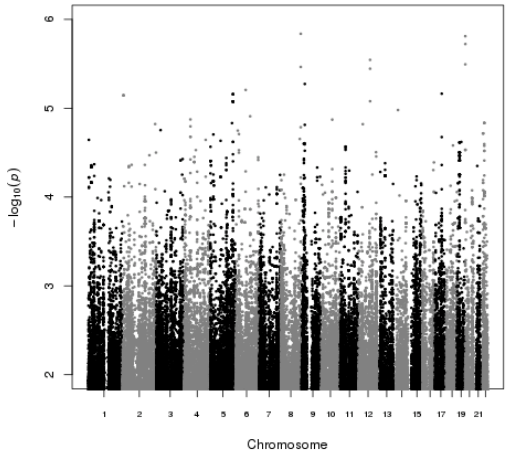


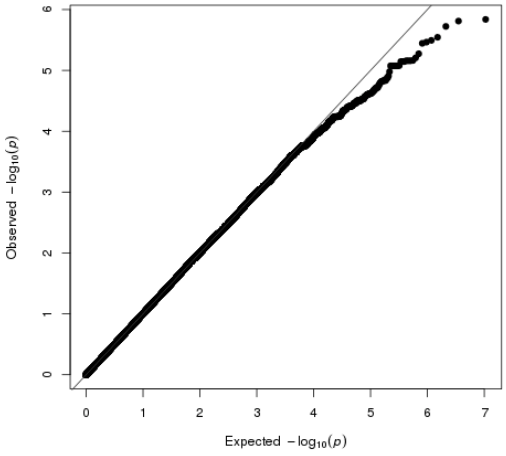
ratio22to24



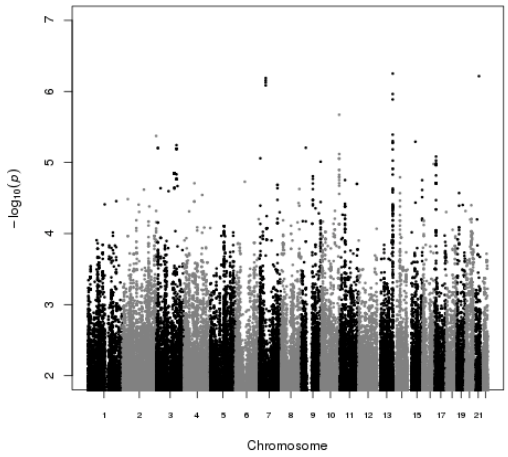


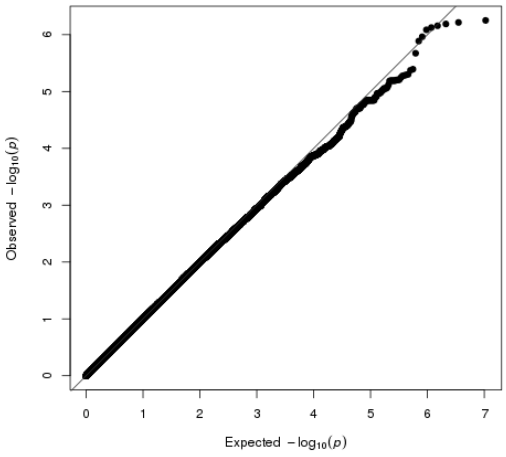
ratio20to24



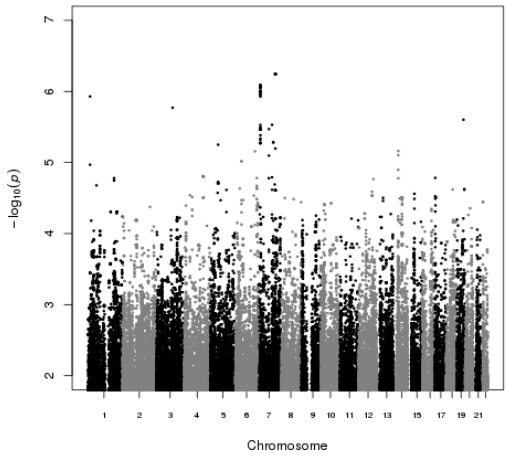


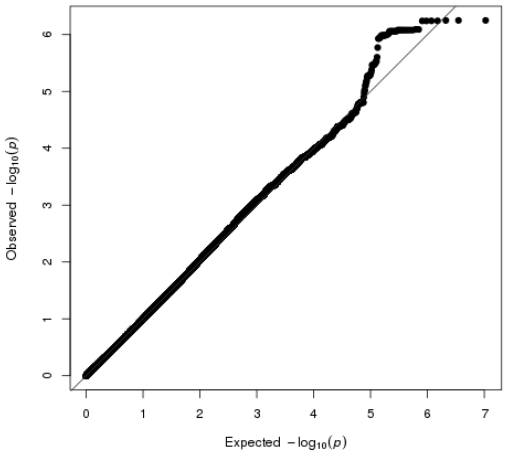
biomcers



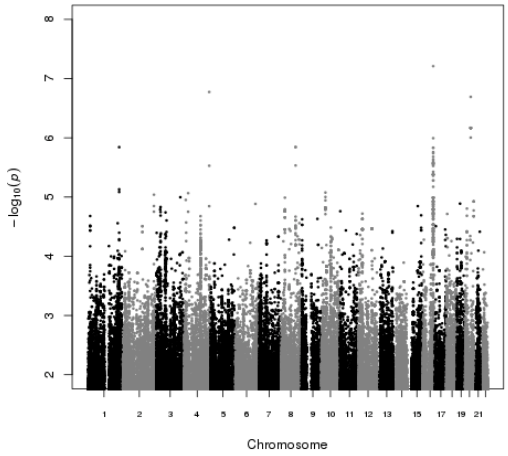


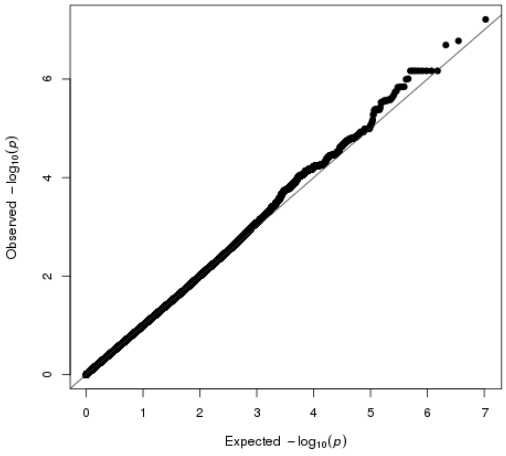
ns_sum



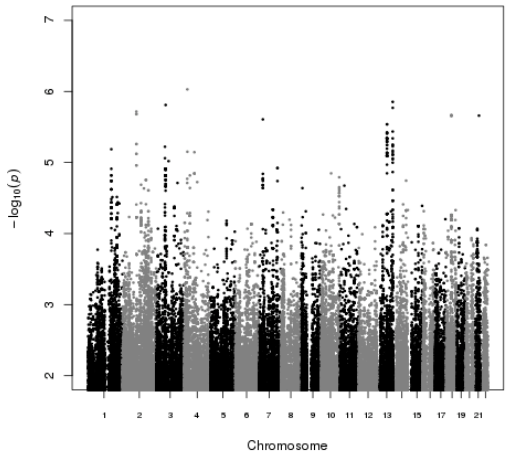


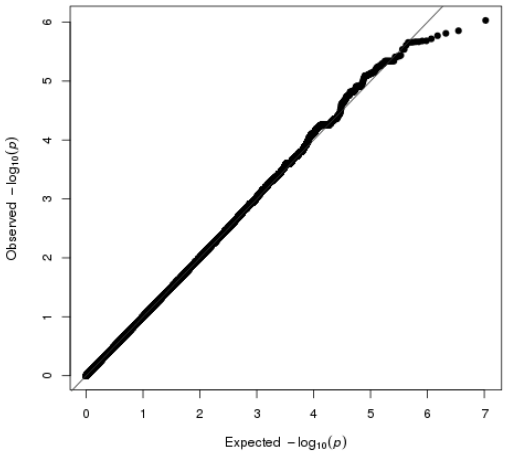
nds_sum



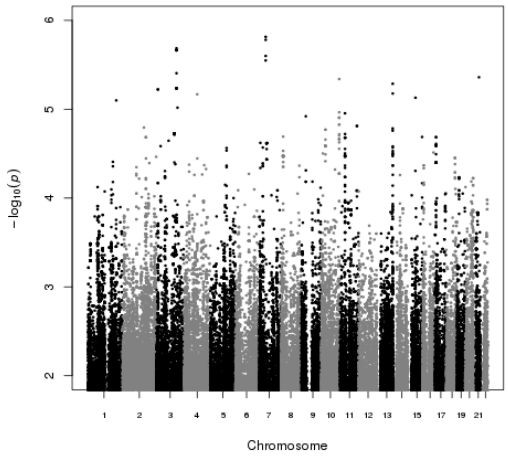


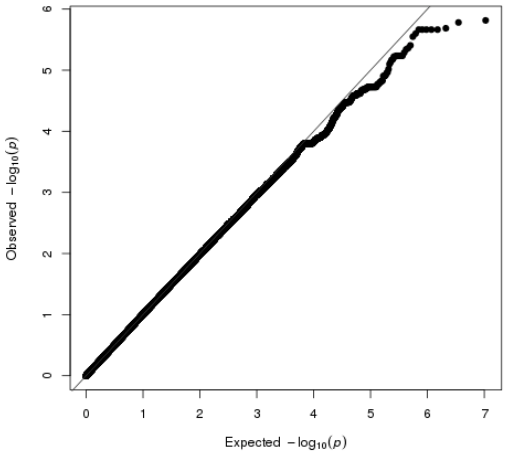
s18_sum



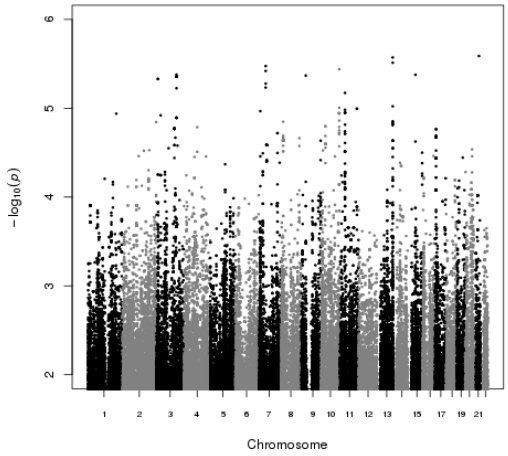


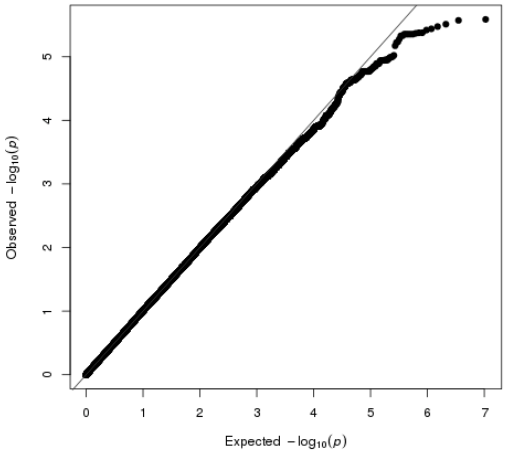
N_s18sum



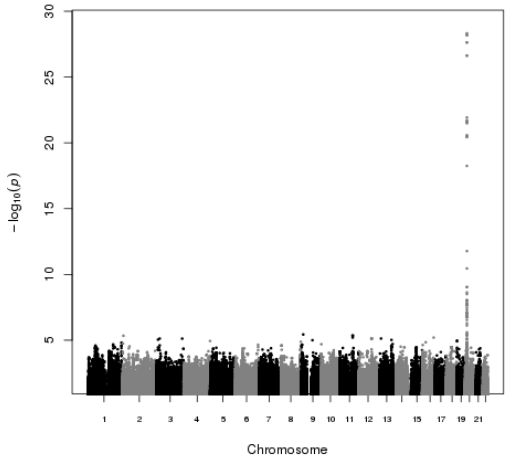


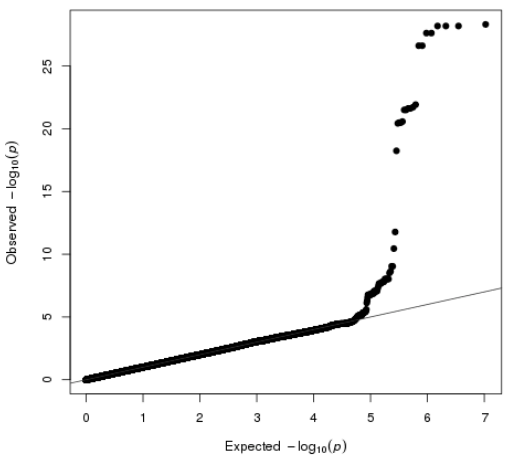
allxs18



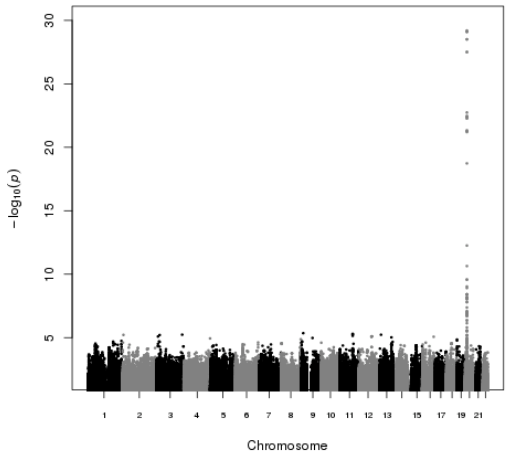


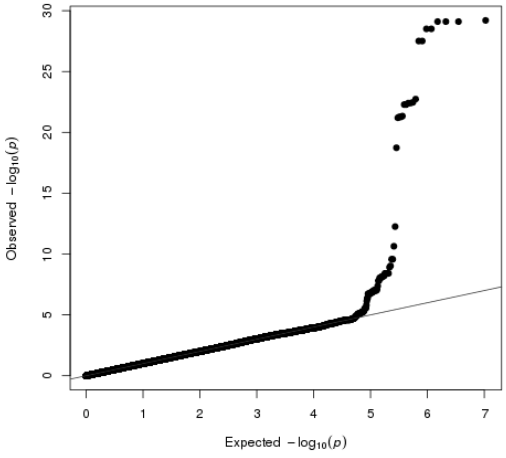
s19_sum



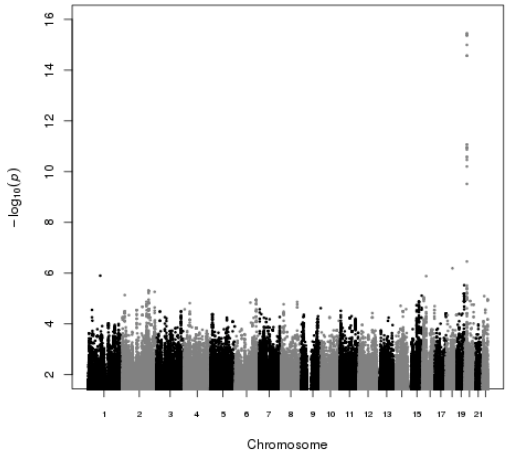


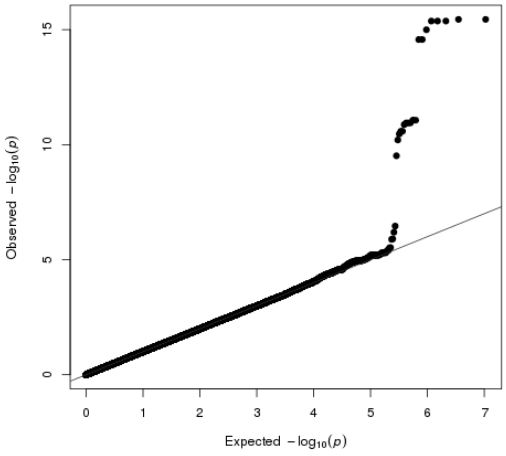
alls19



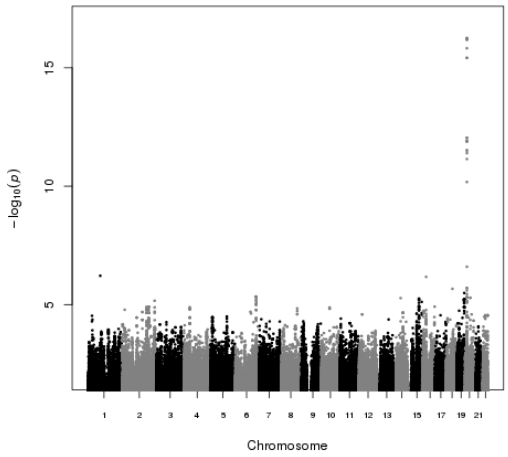


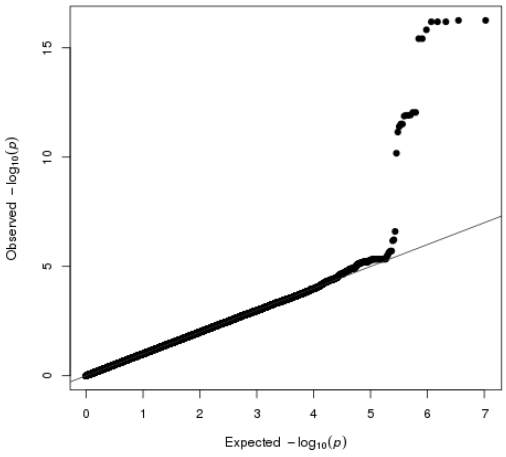
s20_sum



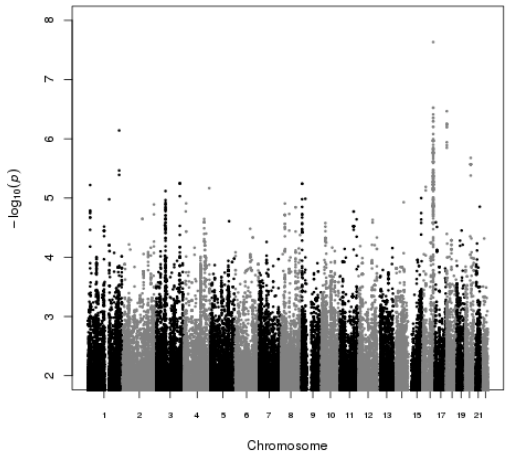


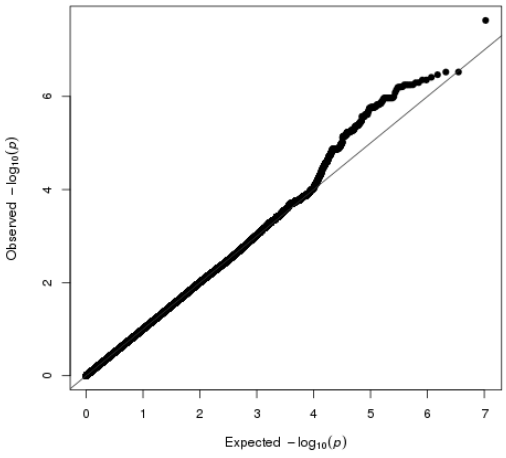
alls20



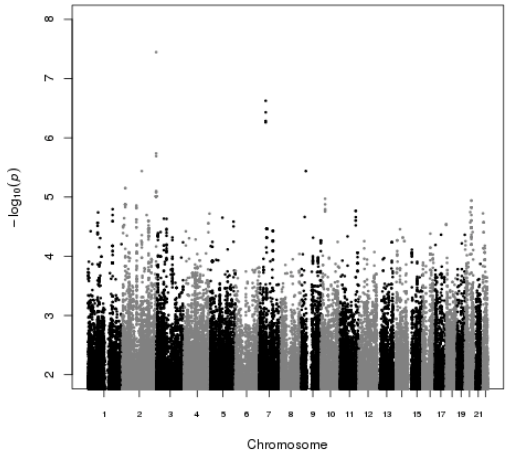


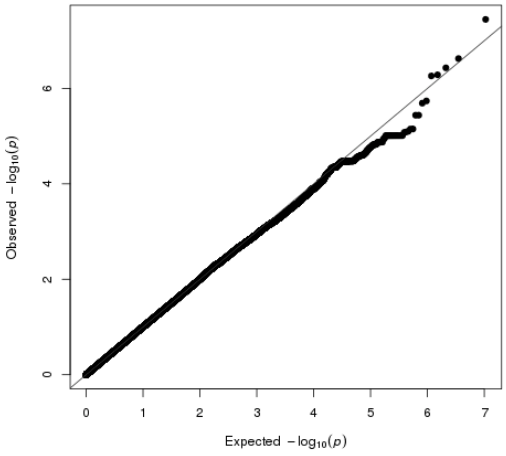
ds18_sum



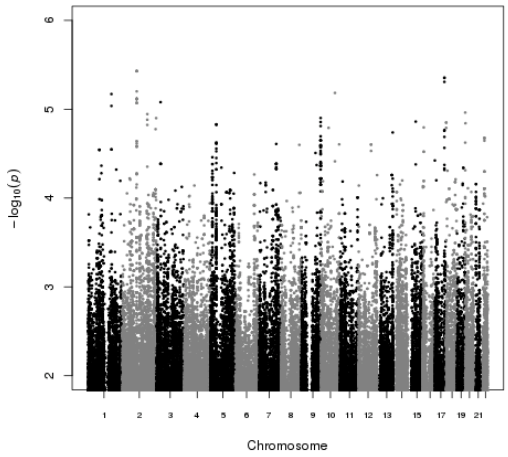


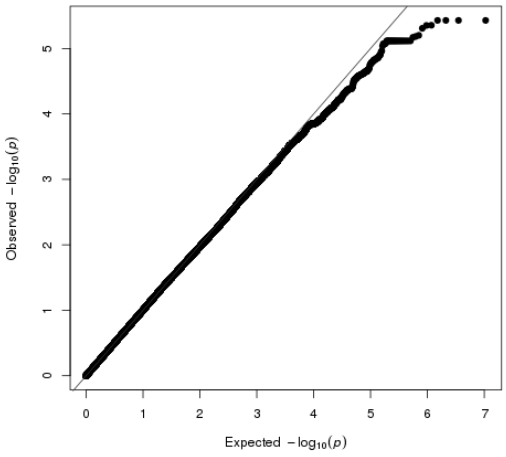
n22_sum



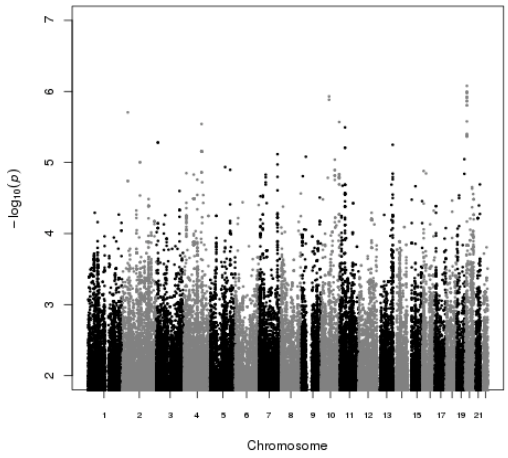


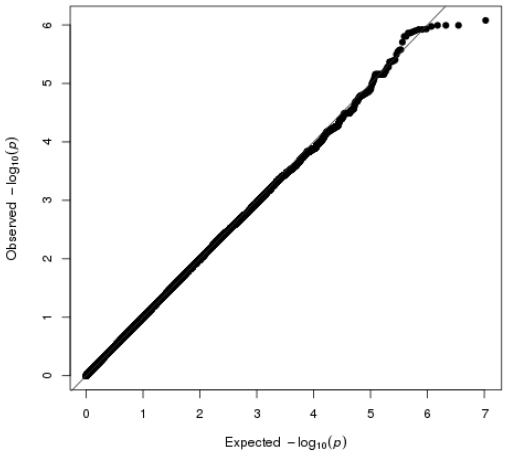
n23_sum



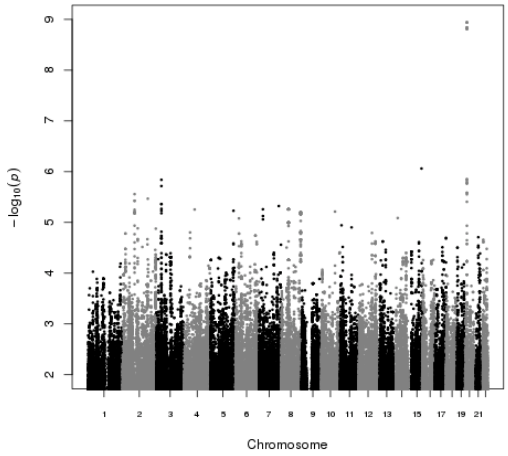


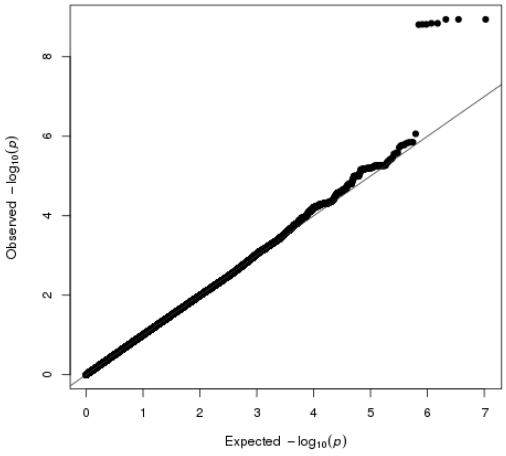
n24_sum



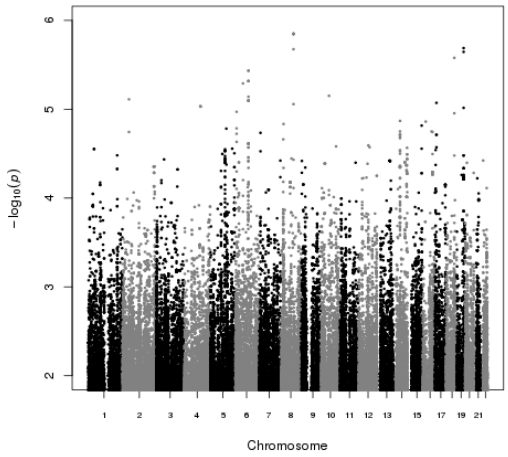


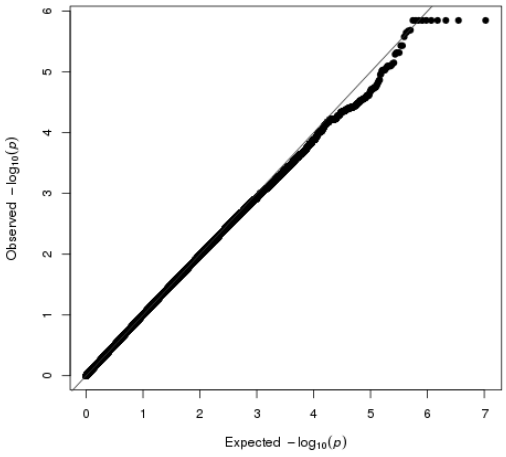
n25_sum



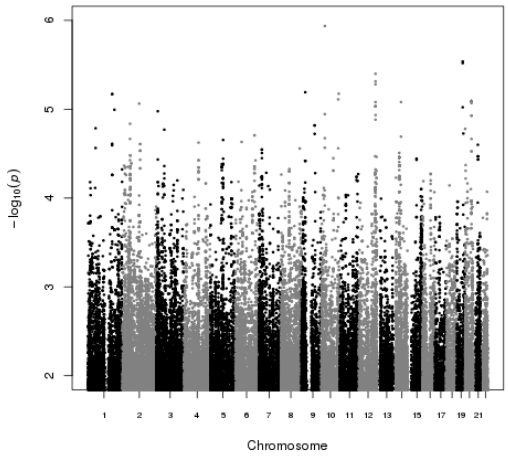


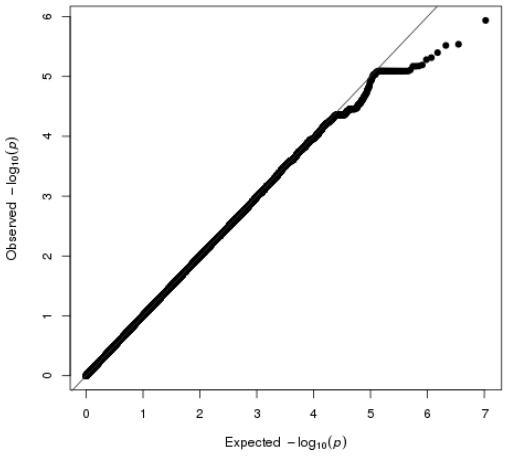
n26_sum



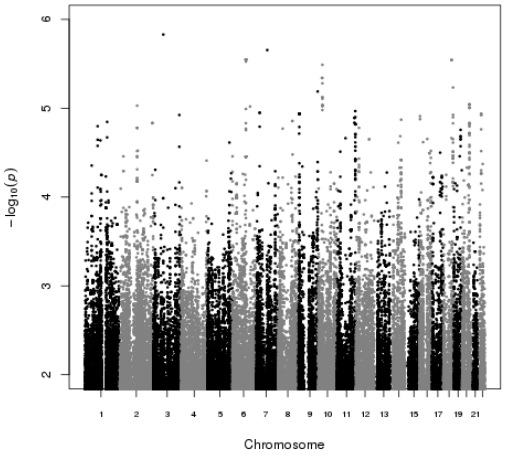


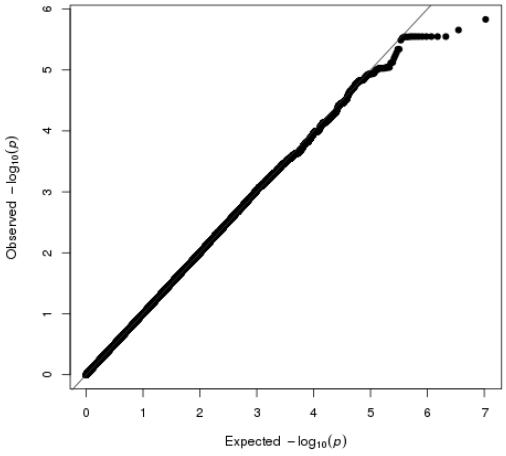
n22ratio



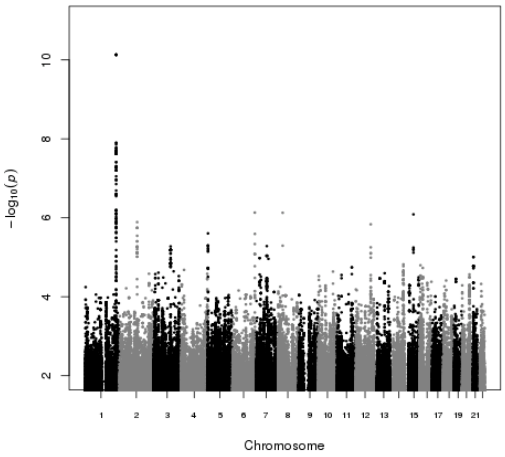


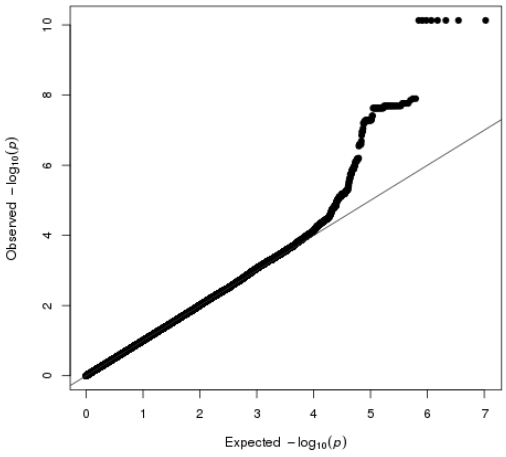
n24ratio



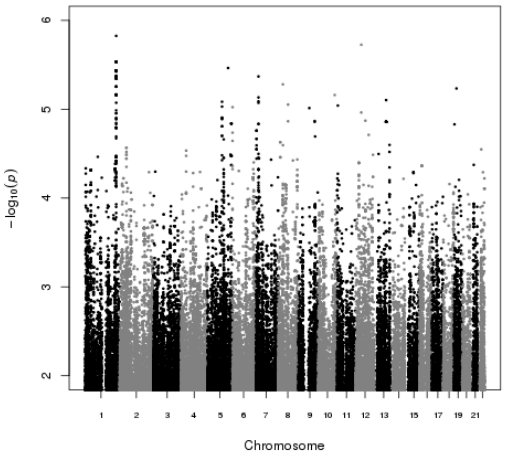


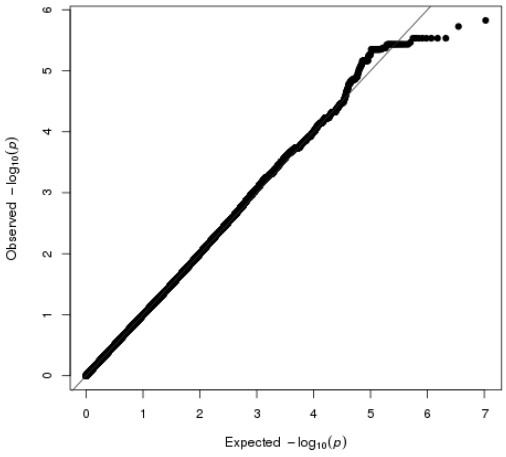
n24s19ratio



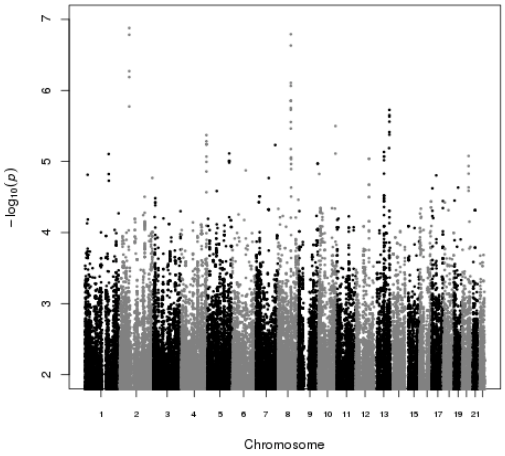


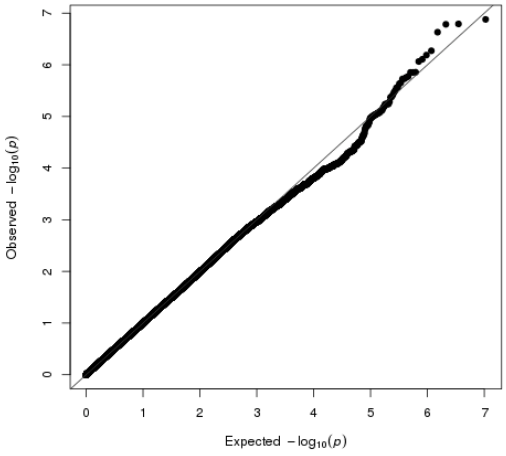
n24s20ratio



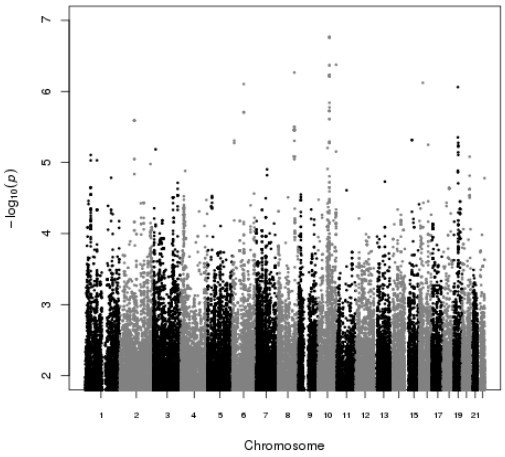


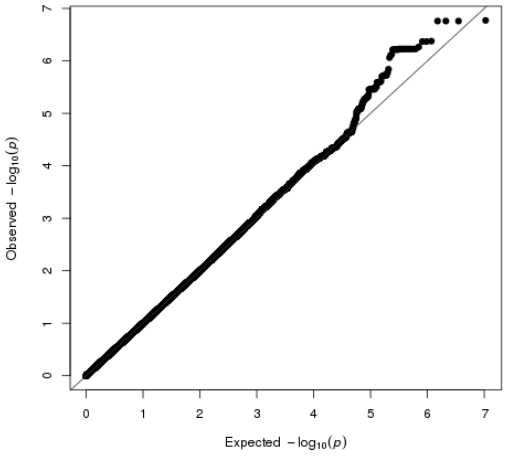
n26ratio



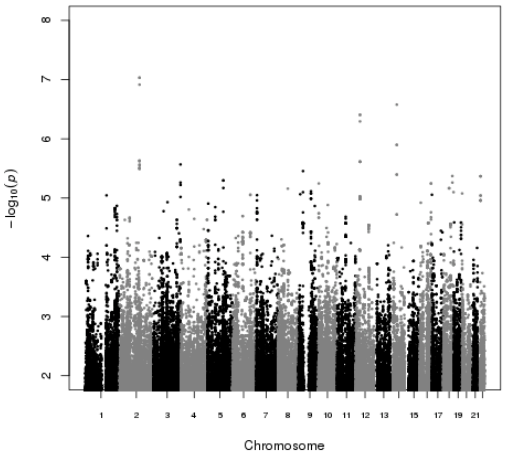


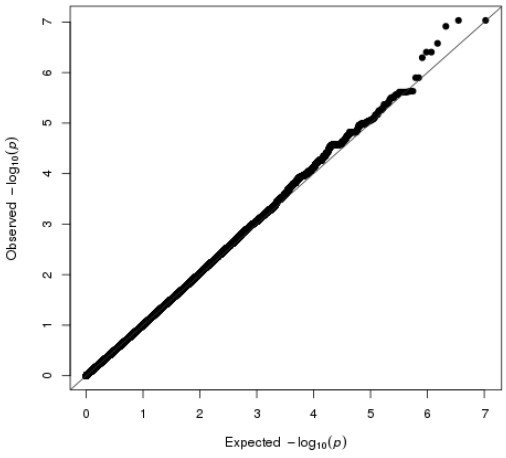
c18s1psratio



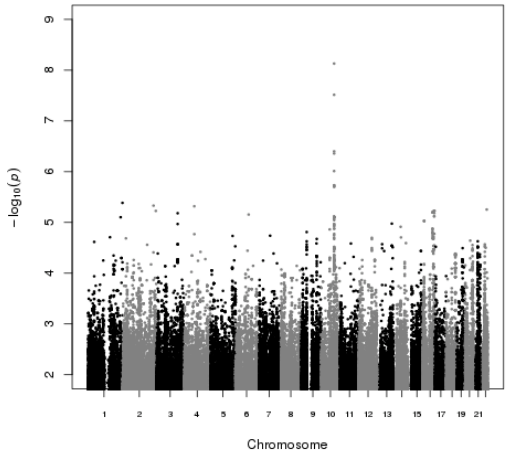


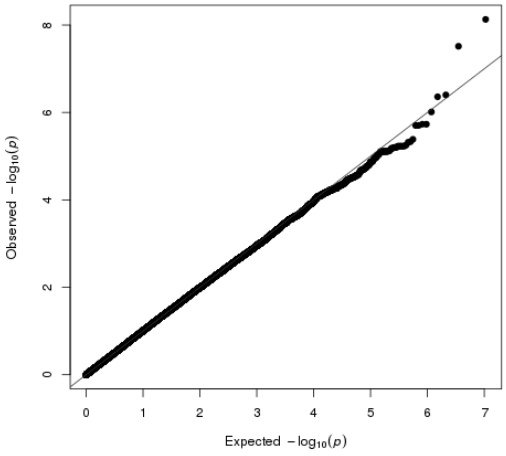
ndssumc18dsratio



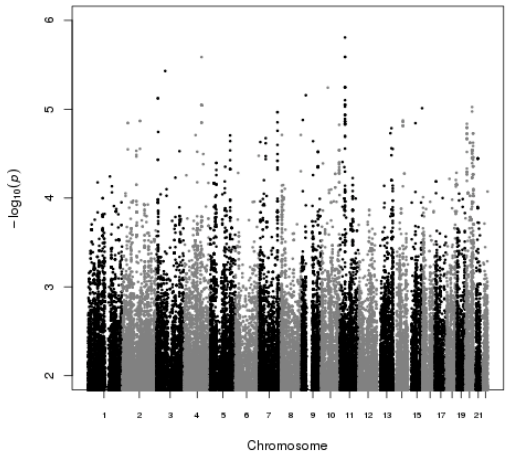


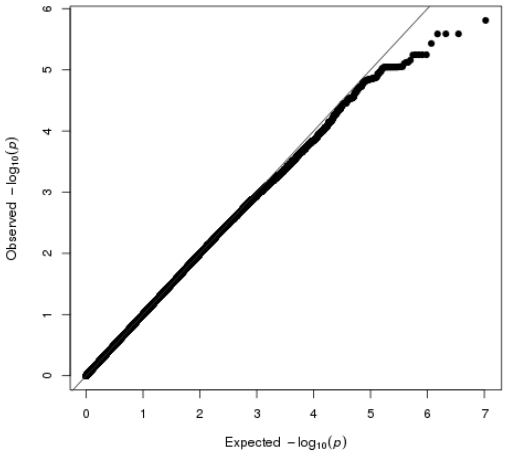
c18snsratio



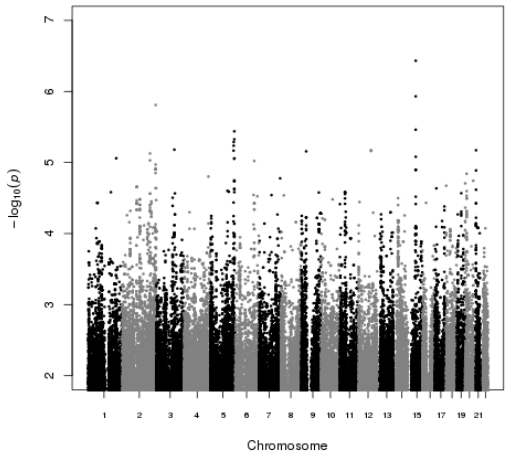


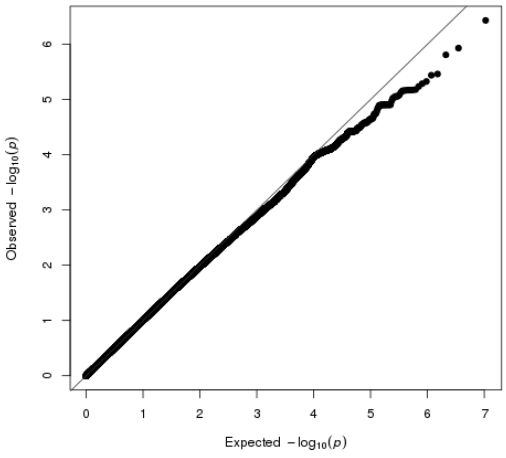
totalsphingo



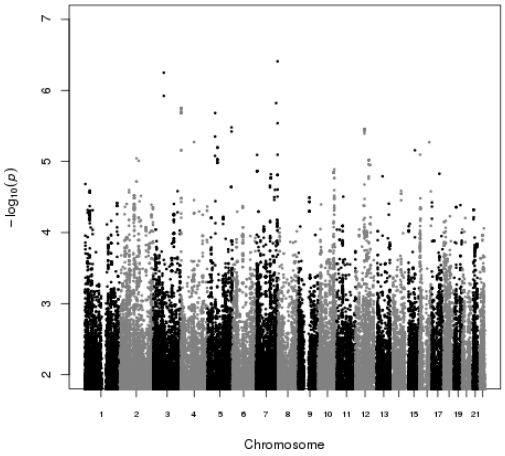


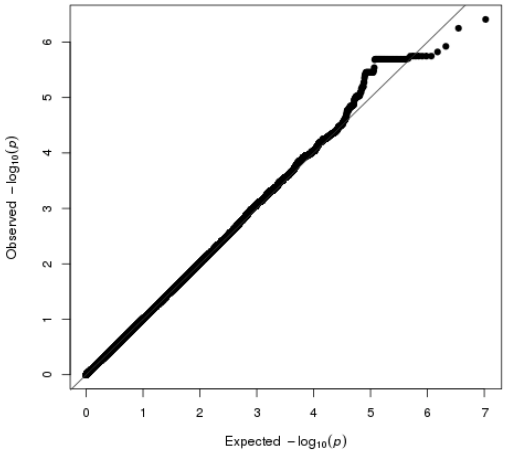
assum



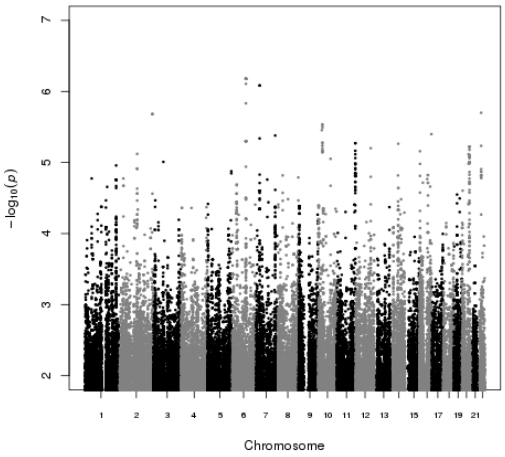


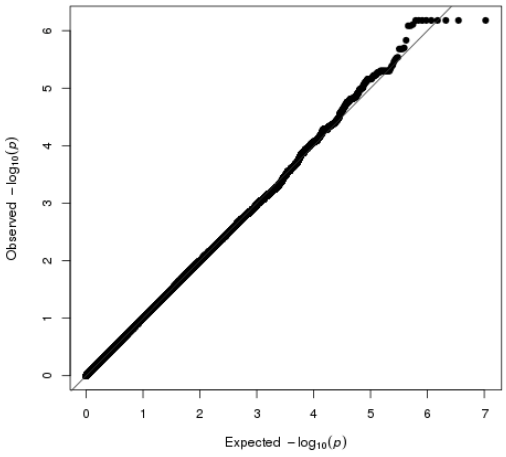
assumc18dsratio



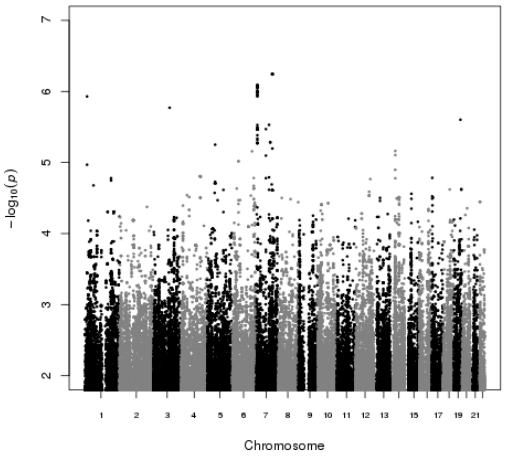


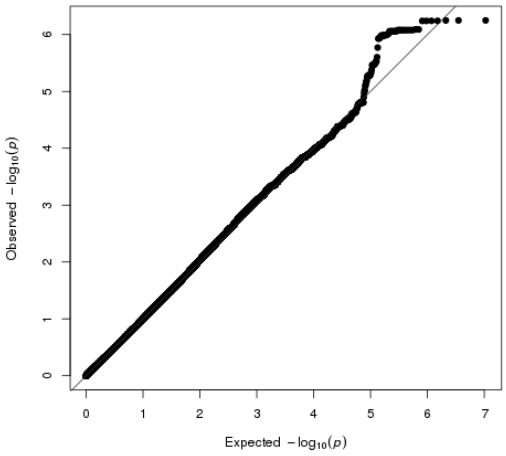
nssumndssumratio



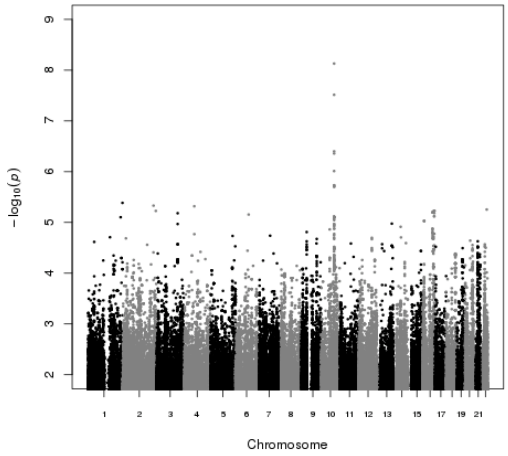


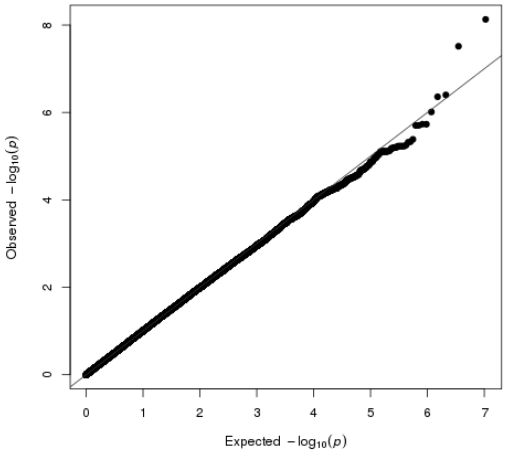
nssum_c18sratio



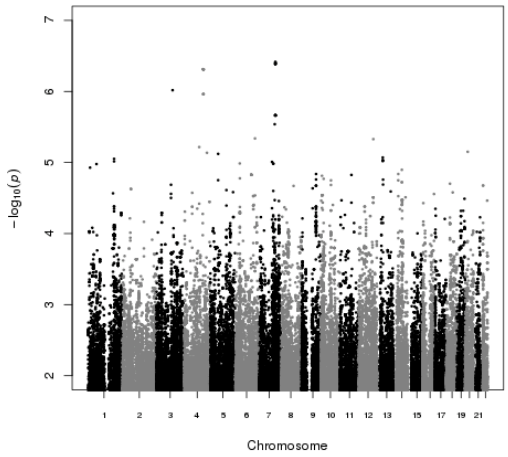


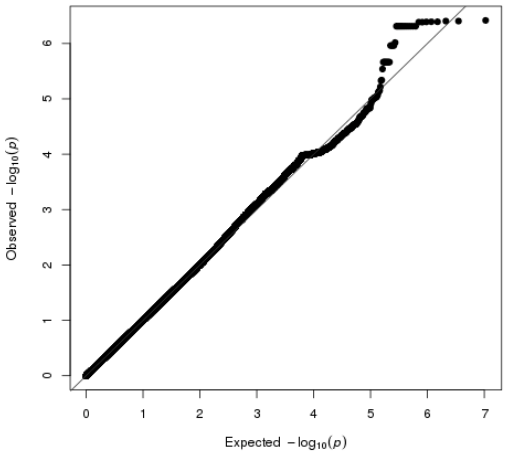
c18s_nssumratio



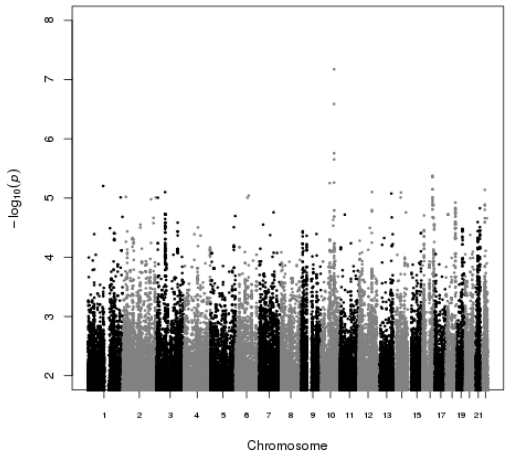


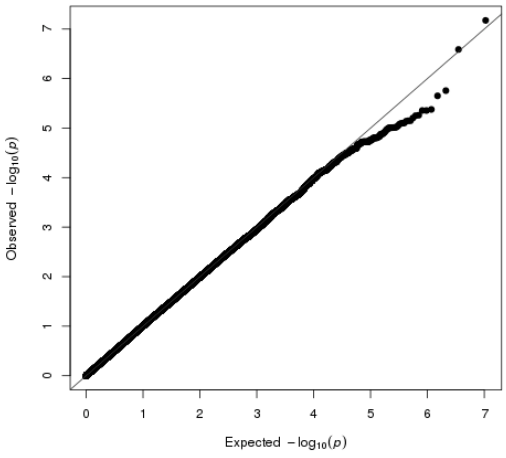
ns18sumc18sratio



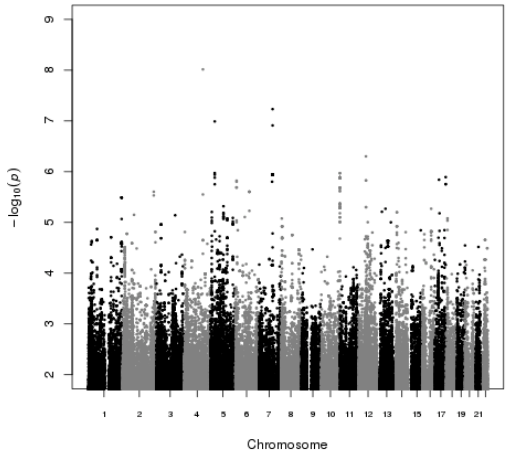


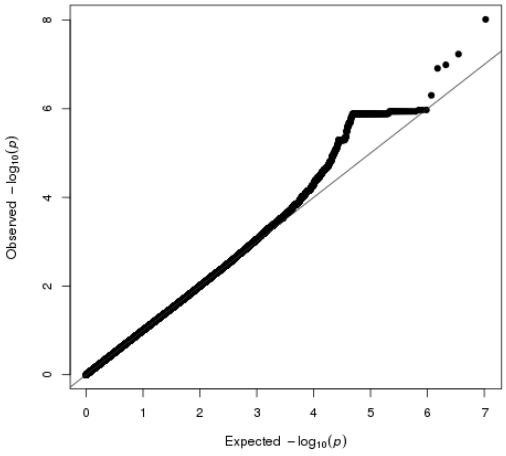
c18sns18sumratio



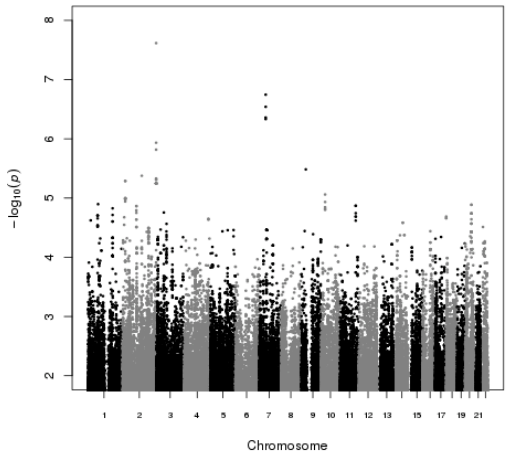


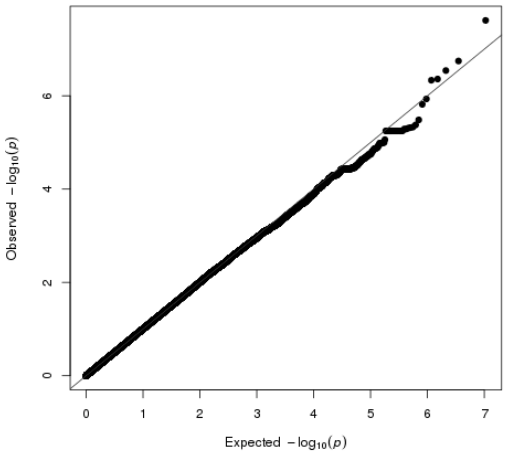
c18sc18s1pratio



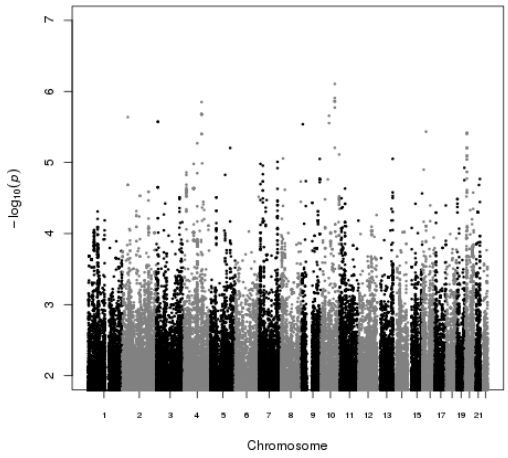


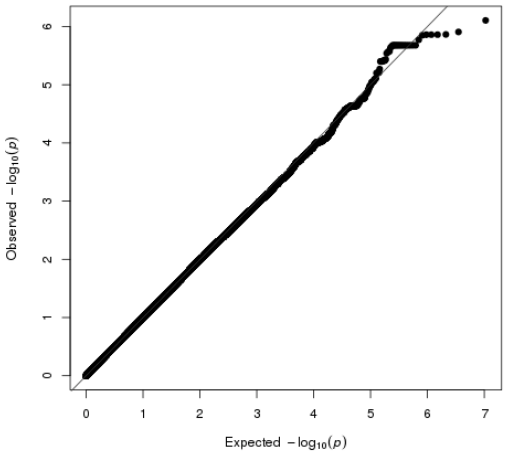
sumn22s



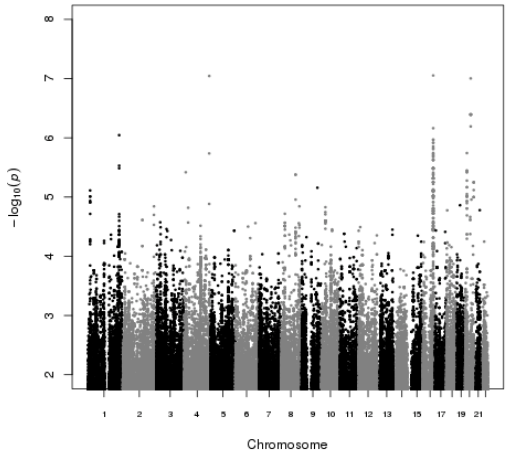


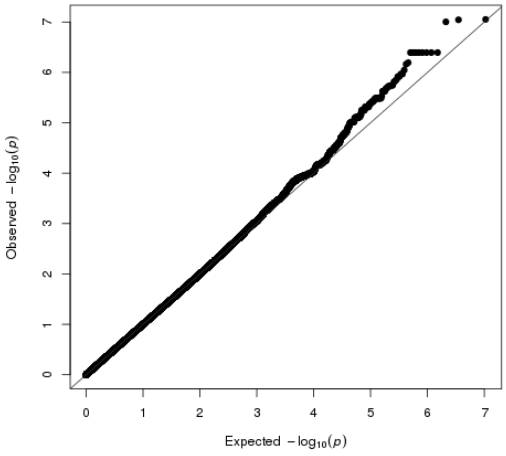
sumn24s



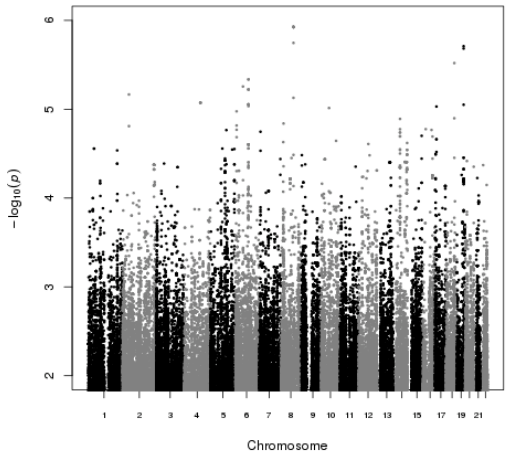


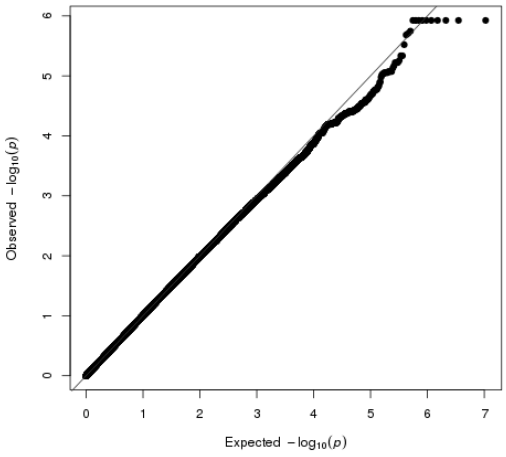
sumn24ds



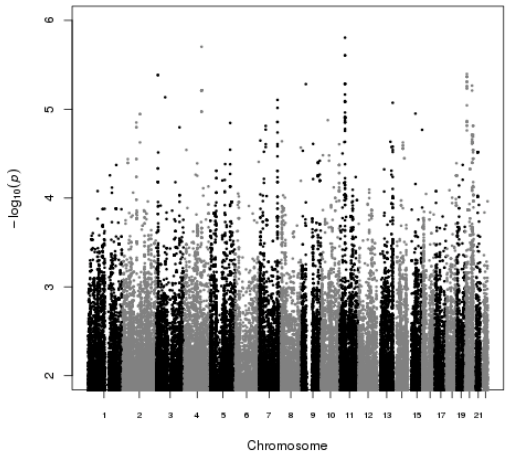


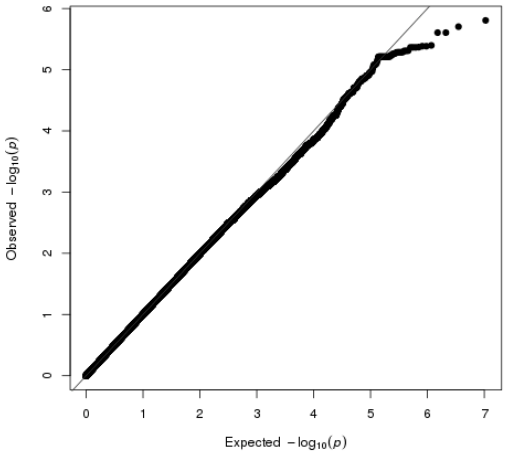
sumn26s





sumcer





sumc18

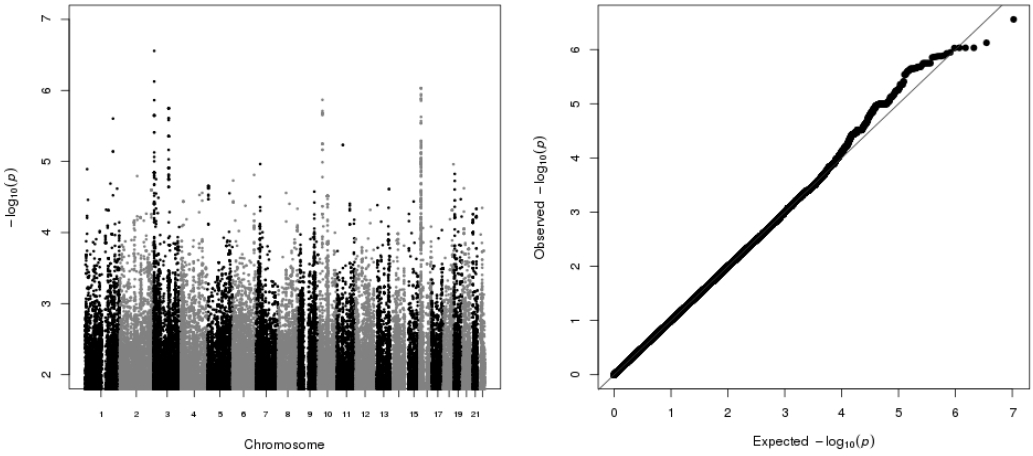


Figure 0-4: Manhattan plots and Quantile-Quantile plots of the GWAS results for the calculated traits.

Table 0.14: Identification of outliers for removal for each lipid species of the range-finding study

The table depicts the Class of lipids, the species, and the number, rstudent value, unadjusted P-value, and Bonferroni corrected P-value for the identification and removal of the most extreme outliers from analyses. Only those samples that were significantly identified as an outlier (P<0.05) were removed from analyses.

Class	Lipid	Number	rstudent	P-value	Bonferroni P-value
NAE	DHEA	113	4.453727	1.40E-05	0.0028611
NAE					
NAE	AEA	32	5.993992	9.41E-09	1.92E-06
NAE					
NAE	DPEA	32	5.157208	5.99E-07	0.0001221
NAE		113	4.277881	2.92E-05	0.0059501
NAE					
NAE	HEA	32	4.333818	2.31E-05	0.0047215
NAE		195	4.029376	7.93E-05	0.016177
NAE					
NAE	OEA	146	4.783083	3.35E-06	0.00068238
NAE		32	3.828973	1.72E-04	0.03511
NAE					
NAE	PEA	32	5.023871	1.12E-06	0.00022843
NAE					
NAE	STEA	149	4.95096	1.56E-06	0.00031836
NAE					
NAE	LEA	126	3.820032	0.00017803	0.036317
NAE					
NAE	VEA	146	5.420887	1.71E-07	3.48E-05
NAE		167*	5.180706	5.40E-07	1.10E-04
NAE		32	4.605262	7.34E-06	1.50E-03
NAE					
NAE	sumEA	32	4.755358	3.77E-06	0.00076978
NAE					
NAE	POEA	167*	6.974368	4.34E-11	8.85E-09
NAE		146	6.799501	1.17E-10	2.39E-08
NAE		203	4.639455	6.29E-06	1.28E-03
NAE					
NAE	PDEA	168	3.226553	0.0014639	0.29864
NAE					
Eico	AA	49	5.193918	5.03E-07	0.00010265
Eico		82*	5.080991	8.56E-07	0.00017455
Eico		123	4.620792	6.82E-06	0.001392
Eico					

Eico	aLA	95	8.755711	9.44E-16	1.93E-13
Eico					
Eico	cox	106	7.613305	1.02E-12	2.08E-10
Eico		95	4.156801	4.78E-05	9.74E-03
Eico					
Eico	cyp450	157	6.629648	3.03E-10	6.19E-08
Eico		101	5.063579	9.28E-07	1.89E-04
Eico		130	4.254283	3.21E-05	6.56E-03
Eico		2	3.922318	1.20E-04	2.45E-02
Eico					
Eico	DHA	37	3.650467	0.00033412	0.067827
Eico					
Eico	DHET1112	9	5.673304	4.86E-08	9.92E-06
Eico		105	4.231108	3.54E-05	7.22E-03
Eico		134	3.839346	1.65E-04	3.38E-02
Eico					
Eico	DHET1415	105	4.709617	4.66E-06	0.00095113
Eico					
Eico	DiHDPA1920	67	3.804883	0.00018875	0.038316
Eico					
Eico	DiHOME910	101	5.692506	4.39E-08	8.95E-06
Eico		173	4.471573	1.30E-05	2.65E-03
Eico		130	4.410658	1.68E-05	3.42E-03
Eico		18	3.832492	1.70E-04	3.46E-02
Eico					
Eico	DiHOME1213	157	14.49129	5.81E-33	1.19E-30
Eico					
Eico	EPA	49	5.669348	4.96E-08	1.01E-05
Eico		82*	4.350934	2.16E-05	4.41E-03
Eico		99	3.789142	2.00E-04	4.08E-02
Eico					
Eico	epdi9	13	66.94585	3.87E-129	7.08E-127
Eico					
Eico	epdi13	17	47.67645	2.11E-104	3.88E-102
Eico					
Eico	EpOME910	101	6.663009	3.20E-10	5.85E-08
Eico		96	5.398931	2.11E-07	3.86E-05
Eico					
Eico	EpOME1213	2	10.062031	3.86E-19	7.11E-17
Eico		183	3.809988	1.91E-04	3.52E-02
Eico					
Eico	HDHA4	100	7.178552	1.44E-11	2.90E-09
Eico		37	5.555737	8.98E-08	1.81E-05
Eico		38	4.72343	4.43E-06	8.90E-04

Eico					
Eico	HETE5	82*	7.661778	9.79E-13	1.85E-10
Eico		123	6.39513	1.26E-09	2.39E-07
Eico		49	5.437267	1.69E-07	3.19E-05
Eico					
Eico	HETE11	123	5.80069	2.53E-08	5.17E-06
Eico		49	4.378654	1.92E-05	3.92E-03
Eico		82*	4.098751	6.03E-05	1.23E-02
Eico					
Eico	HETE12	99	6.378547	1.26E-09	2.57E-07
Eico		115	4.275536	2.98E-05	6.07E-03
Eico					
Eico	HETE15	49	6.547733	5.42E-10	1.04E-07
Eico		82*	4.513213	1.12E-05	2.16E-03
Eico					
Eico	HODE9	106	7.862346	2.34E-13	4.77E-11
Eico		95	4.133853	5.26E-05	1.07E-02
Eico					
Eico	HODE13	106	9.517376	5.78E-18	1.18E-15
Eico		92	7.103379	2.06E-11	4.21E-09
Eico					
Eico	HOTrE9	95	7.592517	1.91E-12	3.34E-10
Eico					
Eico	HOTrE13	95	8.889299	5.78E-16	1.07E-13
Eico					
Eico	LA	106	5.924117	1.34E-08	2.74E-06
Eico		92	4.735083	4.13E-06	8.42E-04
Eico					
Eico	lox5	95	7.667824	7.46E-13	1.52E-10
Eico					
Eico	lox15	106	9.029047	1.43E-16	2.91E-14
Eico		92	6.883153	7.30E-11	1.49E-08
Eico					
Eico	omega3	95	6.749279	1.57E-10	3.19E-08
Eico					
Eico	omega6	106	5.91919	1.38E-08	2.81E-06
Eico		92	4.727009	4.28E-06	8.73E-04
Eico					
Eico	oxho9	12	7.441561	2.84E-12	5.80E-10
Eico		18	6.367167	1.28E-09	2.62E-07
Eico		15	5.852906	1.94E-08	3.96E-06
Eico		20	5.761652	3.09E-08	6.31E-06
Eico		8	4.14817	4.94E-05	1.01E-02
Eico					

Eico	oxho13	15	9.184508	5.18E-17	1.06E-14
Eico		12	5.728766	3.65E-08	7.45E-06
Eico		20	3.954145	1.06E-04	2.17E-02
Eico					
Eico	OxoODE9	8	4.695322	4.94E-06	0.0010071
Eico		204	4.671731	5.48E-06	0.0011171
Eico					
Eico	OxoODE13	9	3.683829	0.0002957	0.060324
Eico					
Eico	she	157	7.881584	2.11E-13	4.31E-11
Eico		101	4.153257	4.87E-05	9.94E-03
Eico		130	4.128947	5.37E-05	1.10E-02
Eico					
Eico	SumEicos	106	5.895757	1.57E-08	3.19E-06
Eico		92	4.781912	3.36E-06	6.86E-04
Eico		95	3.922203	1.21E-04	2.46E-02
Eico					
Eico	TransEKODE	32	4.727912	4.29E-06	0.00087038
Eico		189	4.675644	5.40E-06	0.0010959
Eico					
CER	A22_S18	60	8.656051	1.58E-15	3.22E-13
CER		64	5.688507	4.48E-08	9.14E-06
CER		58	5.198818	4.92E-07	1.00E-04
CER					
CER	A24_S18	59	8.146888	3.91E-14	7.98E-12
CER		60	7.52758	1.70E-12	3.48E-10
CER		58	7.498809	2.02E-12	4.13E-10
CER		64	6.794945	1.20E-10	2.45E-08
CER		177	3.779684	2.07E-04	4.22E-02
CER					
CER	A26_S18	34	5.762851	3.23E-08	6.49E-06
CER		60	5.220381	4.59E-07	9.24E-05
CER		94	4.296602	2.75E-05	5.52E-03
CER					
CER	as_sum	60	9.030136	1.42E-16	2.89E-14
CER		58	6.524544	5.43E-10	1.11E-07
CER		64	6.485804	6.71E-10	1.37E-07
CER		59	5.67556	4.78E-08	9.75E-06
CER		177	3.970467	9.99E-05	2.04E-02
CER					
CER	biomcers	60	6.928722	5.70E-11	1.16E-08
CER		58	5.230321	4.25E-07	8.67E-05
CER		177	3.906182	1.28E-04	2.62E-02
CER					

CER	C18_DS	78	4.200178	4.07E-05	0.0083055
CER					
CER	C18_S	32	4.44287	1.50E-05	0.0029729
CER		19	4.278885	2.97E-05	0.0058763
CER		130	3.733217	2.49E-04	0.049374
CER					
CER	C18_S1P	77	6.054838	7.99E-09	1.47E-06
CER		60	4.431393	1.62E-05	2.99E-03
CER		171	4.304708	2.74E-05	5.04E-03
CER					
CER	N16_S18	58	3.563702	0.00045976	0.093791
CER					
CER	N20_S18	60	9.0125	1.68E-16	3.42E-14
CER		58	5.902963	1.52E-08	3.10E-06
CER		1	3.92512	1.19E-04	2.44E-02
CER		50	3.825783	1.74E-04	3.56E-02
CER					
CER	N22_DS18	58	6.449133	8.21E-10	1.67E-07
CER		60	5.781026	2.80E-08	5.72E-06
CER		64	5.318393	2.78E-07	5.66E-05
CER					
CER	N22_S18	60	6.826619	1.01E-10	2.05E-08
CER		58	4.147079	4.97E-05	1.01E-02
CER		64	4.069338	6.77E-05	1.38E-02
CER					
CER	N22_S19	1	7.73786	4.89E-13	9.98E-11
CER		150	4.615419	7.00E-06	1.43E-03
CER					
CER	N23_S18	60	8.379402	9.13E-15	1.86E-12
CER		64	5.532739	9.75E-08	1.99E-05
CER		58	4.940394	1.64E-06	3.34E-04
CER		177	4.289701	2.78E-05	5.67E-03
CER					
CER	N23_S20	64	5.046055	1.01E-06	0.00020536
CER		58	4.121739	5.50E-05	0.011213
CER					
CER	N24_DS18	64	5.760576	3.13E-08	6.38E-06
CER		58	5.542016	9.36E-08	1.91E-05
CER		60	4.126394	5.41E-05	1.10E-02
CER					
CER	N24_DS19	96	6.340222	1.50E-09	3.05E-07
CER		24	4.167042	4.59E-05	9.36E-03
CER		91	4.163514	4.66E-05	9.50E-03
CER		156	3.950573	1.08E-04	2.20E-02

CER					
CER	N24_DS20	96	5.585782	7.57E-08	1.55E-05
CER		109	4.220522	3.70E-05	7.55E-03
CER					
CER	N24_S16	59	7.174706	1.37E-11	2.78E-09
CER		177	4.108184	5.80E-05	1.18E-02
CER		60	4.05672	7.12E-05	1.45E-02
CER		58	4.039198	7.63E-05	1.56E-02
CER					
CER	N24_S17	177	4.947842	1.58E-06	0.00032294
CER		60	4.586524	7.92E-06	0.0016155
CER		59	4.350871	2.16E-05	0.0043981
CER		58	4.128499	5.35E-05	0.010914
CER					
CER	N24_S18	60	6.837485	9.56E-11	1.95E-08
CER		58	5.220251	4.46E-07	9.10E-05
CER		177	3.960127	1.04E-04	2.12E-02
CER					
CER	N24_S19	177	4.250184	3.27E-05	0.0066666
CER		150	4.210365	3.84E-05	0.0078427
CER		91	4.064666	6.90E-05	0.014074
CER					
CER	N24_S20	109	4.040493	7.60E-05	0.01551
CER		64	3.795002	1.96E-04	0.039909
CER					
CER	N24_S22	149	11.47745	9.92E-24	2.02E-21
CER		37	4.541111	9.68E-06	1.98E-03
CER					
CER	N25_DS18	64	5.362419	2.24E-07	4.58E-05
CER		24	4.183438	4.29E-05	8.75E-03
CER		58	4.122028	5.49E-05	1.12E-02
CER		96	3.77879	2.08E-04	4.24E-02
CER					
CER	N25_S20	78	3.399188	0.00081572	0.16641
CER					
CER	N26_DS18	96	4.302506	2.65E-05	0.0053983
CER		91	4.03145	7.89E-05	0.016099
CER					
CER	N26_S18	64	4.498228	1.16E-05	0.0023677
CER					
CER	N26_S19	78	4.97965	1.37E-06	0.00028008
CER					
CER	N27_S18	78	5.257008	3.79E-07	7.74E-05
CER		148	4.654446	5.96E-06	1.22E-03

CER		52	4.354275	2.14E-05	4.37E-03
CER		76	4.060851	7.05E-05	1.44E-02
CER					
CER	N28_S18	144	4.205128	3.94E-05	0.0080427
CER					
CER	N29_S18	148	10.403365	1.49E-20	3.03E-18
CER		144	4.803868	3.04E-06	6.20E-04
CER		78	3.894967	1.34E-04	2.73E-02
CER					
CER	c18_sum	77	6.084248	5.82E-09	1.19E-06
CER		60	4.570771	8.48E-06	1.73E-03
CER		171	4.17628	4.41E-05	9.00E-03
CER					
CER	c18dsratio	90	7.030803	3.42E-11	6.76E-09
CER		166	5.44889	1.53E-07	3.02E-05
CER					
CER	c18ratio	59	4.75072	3.85E-06	0.00078582
CER		*164	4.147427	4.96E-05	0.010116
CER					
CER	c18s1psratio	47	4.509896	1.14E-05	0.0022568
CER		171	4.051718	7.43E-05	0.014717
CER		51	3.739257	2.45E-04	0.048488
CER					
CER	c18snsratio	19	5.267846	3.57E-07	7.28E-05
CER					
CER	ds18_sum	64	5.900125	1.53E-08	3.12E-06
CER		58	5.678917	4.73E-08	9.64E-06
CER		60	4.010188	8.56E-05	1.75E-02
CER					
CER	dssumc18dsratio	24	10.62085	3.35E-21	6.84E-19
CER		33	4.877352	2.18E-06	4.45E-04
CER		25	4.269105	3.02E-05	6.17E-03
CER		59	3.814592	1.81E-04	3.70E-02
CER					
CER	n22_sum	60	6.174857	3.60E-09	7.35E-07
CER		1	4.422693	1.60E-05	3.25E-03
CER		58	4.305577	2.60E-05	5.31E-03
CER		64	3.962996	1.03E-04	2.10E-02
CER					
CER	n22ratio	199	5.647716	5.72E-08	1.17E-05
CER		146	4.105989	5.93E-05	1.21E-02
CER					
CER	n23_sum	60	8.329289	1.25E-14	2.55E-12
CER		64	5.603705	6.85E-08	1.40E-05

CER		58	4.991986	1.29E-06	2.64E-04
CER		177	4.229994	3.55E-05	7.24E-03
CER					
CER	n24_sum	60	5.428836	1.63E-07	3.32E-05
CER		58	5.012949	1.17E-06	2.39E-04
CER		177	4.425382	1.58E-05	3.22E-03
CER					
CER	n24ratio	72	3.670758	0.0003106	0.063363
CER					
CER	n24s19ratio	50	5.227203	4.45E-07	9.08E-05
CER		171	4.912986	1.91E-06	3.89E-04
CER		146	4.61441	7.17E-06	1.46E-03
CER					
CER	n24s20ratio	146	5.565031	8.40E-08	1.71E-05
CER					
CER	n25_sum	64	5.103913	7.75E-07	0.00015805
CER		58	4.412015	1.68E-05	0.0034201
CER		96	4.2362	3.47E-05	0.0070872
CER					
CER	n26_sum	78	4.226897	3.60E-05	0.0073466
CER		64	4.17316	4.48E-05	0.0091363
CER					
CER	n26ratio	199	3.372925	0.00089455	0.18249
CER					
CER	nds_sum	58	5.085399	8.41E-07	0.00017165
CER		64	4.972671	1.42E-06	0.0002892
CER					
CER	ns_sum	60	6.272552	2.14E-09	4.36E-07
CER		58	5.226205	4.32E-07	8.81E-05
CER		177	4.630552	6.54E-06	1.33E-03
CER					
CER	ratio16to24	*164	5.204546	4.87E-07	9.93E-05
CER					
CER	ratio20to24	129	5.445924	1.51E-07	3.08E-05
CER		*164	4.202233	3.99E-05	8.14E-03
CER					
CER	ratio22to24	100	2.807197	0.0054946	NA
CER					
CER	s18_sum	60	7.600865	1.10E-12	2.24E-10
CER		58	5.484899	1.23E-07	2.52E-05
CER		177	4.179277	4.36E-05	8.90E-03
CER		64	3.808392	1.86E-04	3.79E-02
CER					
CER	s19_sum	150	4.304106	2.63E-05	0.0053628

CER		177	4.161292	4.71E-05	0.0096023
CER		91	3.73803	2.42E-04	0.04944
CER					
CER	s20_sum	64	3.726243	0.00025417	0.05185
CER					
CER	totalsphingo	60	6.404015	1.05E-09	2.14E-07
CER		58	5.497878	1.16E-07	2.36E-05
CER		177	4.424993	1.58E-05	3.22E-03
CER		64	3.811586	1.84E-04	3.74E-02

Table 0.15: Identification of outliers for removal for each lipid species of the full cohort study

The table depicts the Lipid trait, the number, rstudent value, unadjusted P-value, and Bonferroni corrected P-value for the identification and removal of the most extreme outliers from analyses. Only those samples that were significantly identified as an outlier (P<0.05) were removed from analyses.

Lipid	Individual	RStudent	Pvalue	Bonferoni Pvalue
DHEA	973	8.178076	8.63E-16	8.74E-13
	949	7.483035	1.58E-13	1.60E-10
	969	5.192718	2.51E-07	2.54E-04
	888	4.683303	3.21E-06	3.25E-03
	113	4.6364	4.01E-06	4.06E-03
	974	4.110583	4.27E-05	4.32E-02
AEA	928	16.210379	1.07E-52	1.08E-49
	922	8.742394	9.39E-18	9.53E-15
	32	6.111868	1.41E-09	1.43E-06
	843	4.560216	5.74E-06	5.82E-03
DPEA	1006	5.223783	2.13E-07	2.15E-04
	32	5.003779	6.64E-07	6.69E-04
	756	4.920546	1.01E-06	1.02E-03
	880	4.124205	4.03E-05	4.06E-02
HEA	772	5.428204	7.18E-08	7.07E-05
	841	5.390656	8.80E-08	8.67E-05
	32	4.377574	1.33E-05	1.31E-02
	629	4.241506	2.43E-05	2.39E-02
OEA	146	8.248661	4.96E-16	5.03E-13
	32	6.649706	4.80E-11	4.88E-08
	126	5.992332	2.87E-09	2.92E-06
	149	5.026049	5.92E-07	6.01E-04
	113	4.803185	1.80E-06	1.83E-03
PEA	928	9.463171	2.02E-20	2.06E-17
	126	5.404467	8.11E-08	8.24E-05
	32	5.330963	1.20E-07	1.22E-04
	809	5.090266	4.26E-07	4.33E-04
	113	5.03386	5.69E-07	5.78E-04
	922	4.785163	1.96E-06	1.99E-03
STEA	32	5.784562	9.69E-09	9.85E-06

	863	5.093812	4.19E-07	4.25E-04
	149	4.872837	1.28E-06	1.30E-03
	880	4.705168	2.89E-06	2.94E-03
	772	4.500125	7.58E-06	7.70E-03
LEA	126	5.252803	1.83E-07	1.86E-04
	27	4.73287	2.53E-06	2.57E-03
	146	4.475313	8.50E-06	8.63E-03
	782	4.42944	1.05E-05	1.06E-02
	924	4.375384	1.34E-05	0.013595
VEA	146	9.446231	2.36E-20	2.40E-17
	32	7.69751	3.30E-14	3.35E-11
	126	6.345617	3.34E-10	3.39E-07
	113	5.91857	4.45E-09	4.52E-06
	880	4.410961	1.14E-05	1.16E-02
POEA	828	8.279985	4.00E-16	3.95E-13
	924	7.629439	5.57E-14	5.50E-11
	843	6.775211	2.14E-11	2.11E-08
	756	6.467363	1.57E-10	1.55E-07
	146	6.409285	2.27E-10	2.24E-07
	888	6.125711	1.31E-09	1.29E-06
	893	5.553794	3.60E-08	3.55E-05
	875	5.382228	9.20E-08	9.09E-05
	904	4.833421	1.56E-06	1.54E-03
PDEA	826	5.438733	6.80E-08	0.00006623
	828	4.370793	1.37E-05	0.013364
	756	4.281065	2.05E-05	0.019923
A22_S18	60	13.112621	2.17E-36	2.21E-33
	64	9.204592	1.92E-19	1.95E-16
	58	8.509143	6.26E-17	6.36E-14
	394	7.25542	7.97E-13	8.10E-10
	59	5.713524	1.46E-08	1.48E-05
	177	5.364729	1.01E-07	1.02E-04
A24_S18	59	16.795554	4.96E-56	5.04E-53
	60	15.59472	2.67E-49	2.72E-46
	58	15.538653	5.42E-49	5.51E-46
	64	14.159718	1.17E-41	1.18E-38
	177	8.052402	2.27E-15	2.30E-12

A26_S18	985	8.066152	2.35E-15	2.09E-12
	936	7.958606	5.32E-15	4.73E-12
	430	5.329285	1.25E-07	1.11E-04
	34	4.97553	7.82E-07	6.96E-04
	858	4.596046	4.93E-06	4.39E-03
	735	4.371094	1.38E-05	1.23E-02
	983	4.211502	2.80E-05	2.49E-02
C18_DS	901	5.695406	1.62E-08	1.64E-05
	342	5.658823	1.99E-08	2.02E-05
	768	4.761344	2.21E-06	2.24E-03
	258	4.50976	7.25E-06	7.37E-03
	582	4.318108	1.73E-05	1.76E-02
	244	4.24673	2.37E-05	2.41E-02
	893	4.10866	4.30E-05	4.37E-02
C18_S	565	30.627764	8.20E-146	8.29E-143
	342	8.526808	5.50E-17	5.56E-14
	258	5.156948	3.03E-07	3.06E-04
C18_S1P	857	5.505455	4.70E-08	4.66E-05
	263	5.259578	1.77E-07	1.76E-04
	1006	5.188479	2.57E-07	2.55E-04
	391	4.673958	3.36E-06	3.34E-03
	260	4.452973	9.44E-06	9.36E-03
	393	4.376419	1.34E-05	1.32E-02
	427	4.119136	4.12E-05	4.09E-02
N16_S18	763	12.805192	6.86E-35	6.97E-32
	350	7.082976	2.65E-12	2.69E-09
	734	6.296996	4.52E-10	4.60E-07
	766	5.190095	2.54E-07	2.58E-04
	558	4.956519	8.42E-07	8.55E-04
	741	4.424397	1.07E-05	1.09E-02
N20_S18	60	8.763867	7.88E-18	8.00E-15
	394	7.156579	1.59E-12	1.62E-09
	58	6.25996	5.69E-10	5.78E-07
	283	5.33115	1.20E-07	1.22E-04
	1	4.469721	8.72E-06	8.86E-03
	877	4.466829	8.84E-06	8.98E-03
	350	4.15059	3.60E-05	3.65E-02
	50	4.10078	4.45E-05	4.52E-02

N22_DS18	58	9.517787	1.25E-20	1.27E-17
	60	8.6451	2.08E-17	2.11E-14
	64	8.040071	2.50E-15	2.53E-12
	394	5.330182	1.21E-07	1.23E-04
	344	4.670101	3.42E-06	3.47E-03
	396	4.56479	5.61E-06	5.70E-03
	876	4.552104	5.96E-06	6.05E-03
	37	4.247332	2.36E-05	2.40E-02
	682	4.154474	3.54E-05	3.59E-02
	96	4.1495	3.61E-05	3.67E-02
N22_S18	60	8.554276	4.37E-17	4.44E-14
	394	7.867722	9.28E-15	9.43E-12
	64	5.566128	3.34E-08	3.39E-05
	58	5.476077	5.49E-08	5.58E-05
	1	4.955627	8.45E-07	8.59E-04
	37	4.458855	9.17E-06	9.31E-03
N22_S19	1	9.821422	8.31E-22	8.45E-19
	458	7.744749	2.32E-14	2.36E-11
	150	5.754648	1.15E-08	1.17E-05
	457	5.129496	3.48E-07	3.54E-04
	455	4.797877	1.85E-06	1.87E-03
	394	4.419904	1.09E-05	1.11E-02
	454	4.261219	2.22E-05	2.26E-02
N23_S18	60	12.519079	1.57E-33	1.59E-30
	64	8.804685	5.61E-18	5.70E-15
	58	7.976814	4.05E-15	4.12E-12
	394	6.567219	8.19E-11	8.32E-08
	177	6.227793	6.93E-10	7.04E-07
	59	5.259461	1.76E-07	1.79E-04
	1	4.544188	6.18E-06	6.28E-03
N23_S20	64	6.297225	4.51E-10	4.59E-07
	58	5.102512	4.00E-07	4.07E-04
	59	4.809833	1.74E-06	1.77E-03
	60	4.510129	7.24E-06	7.35E-03
	795	4.161161	3.44E-05	3.49E-02
N24_DS18	64	9.297456	8.66E-20	8.79E-17
	58	8.741311	9.49E-18	9.64E-15
	60	7.659569	4.36E-14	4.43E-11
	830	5.68678	1.70E-08	1.72E-05

	24	5.002012	6.69E-07	6.80E-04
	156	4.409789	1.15E-05	1.16E-02
	96	4.116609	4.16E-05	4.23E-02
N24_DS19	96	8.30489	3.19E-16	3.24E-13
	24	5.734227	1.29E-08	1.31E-05
	156	5.229225	2.07E-07	2.10E-04
	87	5.131664	3.44E-07	3.50E-04
	91	5.056743	5.06E-07	5.14E-04
N24_DS20	96	7.053046	3.26E-12	3.30E-09
	109	5.312943	1.33E-07	1.35E-04
	884	4.977626	7.57E-07	7.69E-04
	411	4.877406	1.25E-06	1.27E-03
	24	4.365155	1.40E-05	1.42E-02
N24_S16	59	9.63884	4.31E-21	4.38E-18
	453	8.048693	2.34E-15	2.38E-12
	60	5.888465	5.31E-09	5.39E-06
	58	5.662059	1.95E-08	1.98E-05
	1	5.441902	6.62E-08	6.72E-05
	177	4.668348	3.45E-06	3.50E-03
	247	4.349118	1.51E-05	1.53E-02
N24_S17	60	7.102564	2.31E-12	2.35E-09
	59	6.889273	9.86E-12	1.00E-08
	177	6.640623	5.10E-11	5.18E-08
	58	6.250937	6.01E-10	6.11E-07
	266	4.415608	1.12E-05	1.13E-02
N24_S18	60	10.416271	3.35E-24	3.40E-21
	58	8.139062	1.17E-15	1.18E-12
	177	5.720215	1.40E-08	1.42E-05
	254	4.86012	1.36E-06	1.38E-03
	64	4.663721	3.52E-06	3.58E-03
	150	4.422919	1.08E-05	1.10E-02
	394	4.352945	1.48E-05	1.50E-02
	398	4.205205	2.84E-05	2.89E-02
N24_S19	455	8.657529	1.88E-17	1.91E-14
	454	8.499212	6.77E-17	6.88E-14
	458	7.029112	3.82E-12	3.88E-09
	457	5.257379	1.78E-07	1.81E-04
	150	5.022622	6.02E-07	6.12E-04

	177	4.958819	8.31E-07	8.45E-04
	444	4.699597	2.97E-06	3.01E-03
	91	4.516374	7.03E-06	7.14E-03
N24_S20	294	5.584345	3.02E-08	3.07E-05
	857	4.426743	1.06E-05	1.08E-02
	304	4.311041	1.78E-05	1.81E-02
N24_S22	149	6.927866	7.60E-12	7.72E-09
	249	6.061356	1.90E-09	1.93E-06
	847	5.790625	9.36E-09	9.51E-06
	1003	5.741197	1.24E-08	1.26E-05
	294	5.112858	3.79E-07	3.86E-04
	334	4.485784	8.10E-06	8.23E-03
	747	4.219891	2.66E-05	2.71E-02
N25_DS18	64	7.689358	3.50E-14	3.56E-11
	24	6.050915	2.03E-09	2.06E-06
	58	6.024271	2.38E-09	2.42E-06
	96	5.66634	1.90E-08	1.93E-05
	33	4.857109	1.38E-06	1.40E-03
	384	4.244633	2.39E-05	2.43E-02
N25_S20	429	5.544191	3.77E-08	3.83E-05
	210	4.921161	1.00E-06	1.02E-03
	400	4.562517	5.68E-06	5.77E-03
	849	4.525591	6.74E-06	6.85E-03
	439	4.127587	3.97E-05	4.03E-02
N26_DS18	384	6.574001	7.84E-11	7.96E-08
	607	5.773969	1.03E-08	1.05E-05
	96	4.808081	1.76E-06	1.78E-03
	347	4.422088	1.08E-05	1.10E-02
	91	4.409222	1.15E-05	1.17E-02
	33	4.196133	2.95E-05	3.00E-02
N26_S18	590	4.774732	2.07E-06	0.0020988
	286	4.380709	1.31E-05	0.013275
	307	4.280588	2.04E-05	0.020745
	846	4.181893	3.14E-05	0.031927
N26_S19	590	7.29849	5.89E-13	5.98E-10
	454	6.758838	2.35E-11	2.39E-08
	425	6.229506	6.86E-10	6.97E-07

	429	5.298505	1.43E-07	1.46E-04
	227	5.145337	3.21E-07	3.26E-04
	400	4.48456	8.14E-06	8.27E-03
	223	4.295748	1.91E-05	1.94E-02
	444	4.239652	2.44E-05	2.48E-02
N27_S18	985	7.071512	2.86E-12	2.91E-09
	936	6.236117	6.59E-10	6.70E-07
	854	5.176546	2.73E-07	2.77E-04
	735	4.934038	9.42E-07	9.57E-04
	320	4.574277	5.37E-06	5.46E-03
	237	4.14866	3.63E-05	3.68E-02
N28_S18	851	5.869122	5.93E-09	6.03E-06
	235	5.702539	1.55E-08	1.57E-05
	852	4.917569	1.02E-06	1.04E-03
	223	4.449901	9.54E-06	9.70E-03
	425	4.326985	1.66E-05	1.69E-02
	237	4.166844	3.35E-05	3.41E-02
N29_S18	985	7.555159	9.40E-14	9.54E-11
	429	7.014627	4.23E-12	4.30E-09
	417	6.460543	1.62E-10	1.65E-07
	148	6.434942	1.91E-10	1.94E-07
	851	6.112433	1.40E-09	1.42E-06
	326	5.322788	1.26E-07	1.28E-04
	237	4.832579	1.56E-06	1.58E-03
	222	4.397444	1.21E-05	1.23E-02
	915	4.086022	4.74E-05	4.81E-02
N_s18sum	596	11.309342	5.60E-28	5.59E-25
	595	8.587649	3.40E-17	3.39E-14
	599	6.191319	8.72E-10	8.71E-07
	844	5.764764	1.09E-08	1.09E-05
	831	4.613974	4.47E-06	4.46E-03
	984	4.120002	4.10E-05	4.10E-02
alls19	751	8.445887	1.07E-16	1.07E-13
	753	7.780323	1.81E-14	1.81E-11
	718	6.96797	5.86E-12	5.86E-09
	717	5.172469	2.80E-07	2.79E-04
	741	5.042652	5.46E-07	5.45E-04
	653	4.589662	5.01E-06	5.01E-03
	844	4.350592	1.50E-05	1.50E-02

	639	4.234614	2.50E-05	2.50E-02
alls20	577	5.194159	2.50E-07	0.00024942
allxs18	596	11.646084	1.79E-29	1.79E-26
	595	9.117078	4.17E-19	4.17E-16
	599	6.868861	1.14E-11	1.14E-08
	844	5.562686	3.42E-08	3.41E-05
	831	4.556485	5.85E-06	5.84E-03
assum	596	16.638577	5.23E-55	5.22E-52
	595	12.437121	4.21E-33	4.20E-30
	599	12.15163	9.02E-32	9.01E-29
	594	10.619454	5.05E-25	5.04E-22
	844	7.243082	8.80E-13	8.79E-10
	831	4.319837	1.72E-05	1.72E-02
assumc18dsratio	727	26.384232	9.30E-117	9.29E-114
	594	16.878917	2.08E-56	2.08E-53
	599	5.081078	4.48E-07	4.48E-04
	596	4.953769	8.55E-07	8.54E-04
	619	4.560272	5.74E-06	5.74E-03
	571	4.258999	2.25E-05	2.25E-02
biomcers	596	11.288594	6.90E-28	6.89E-25
	595	8.329424	2.68E-16	2.68E-13
	844	5.873911	5.80E-09	5.80E-06
	599	5.836399	7.22E-09	7.21E-06
	831	5.24867	1.87E-07	1.87E-04
c18s1psratio	616	7.86386	9.92E-15	9.66E-12
	842	7.111356	2.23E-12	2.17E-09
	621	6.632165	5.50E-11	5.35E-08
	346	5.935612	4.08E-09	3.97E-06
	412	5.366088	1.01E-07	9.81E-05
	59	5.274577	1.64E-07	1.60E-04
	596	4.947901	8.84E-07	8.61E-04
	408	4.635276	4.05E-06	3.95E-03
	875	4.624887	4.26E-06	4.15E-03
	171	4.428976	1.05E-05	1.03E-02
c18s_nssumratio	510	21.232006	1.29E-82	1.28E-79
	251	4.767862	2.14E-06	2.13E-03
	305	4.595275	4.89E-06	4.85E-03

	981	4.541512	6.28E-06	6.23E-03
	259	4.253437	2.31E-05	2.29E-02
	332	4.108359	4.32E-05	4.28E-02
c18sc18s1pratio	12	19.320433	1.29E-70	1.25E-67
	649	11.734261	7.93E-30	7.72E-27
	516	6.60998	6.33E-11	6.17E-08
	759	5.801382	8.90E-09	8.67E-06
	325	5.589256	2.96E-08	2.89E-05
	11	5.532015	4.07E-08	3.97E-05
	703	5.341556	1.15E-07	1.12E-04
	683	5.048199	5.33E-07	5.19E-04
	10	4.358794	1.45E-05	1.41E-02
c18sns18sumratio	510	17.717191	3.75E-61	3.72E-58
	251	5.922379	4.39E-09	4.35E-06
	305	4.912219	1.05E-06	1.05E-03
	259	4.307267	1.82E-05	1.80E-02
	981	4.28281	2.03E-05	2.01E-02
c18snsratio	510	21.232006	1.29E-82	1.28E-79
	251	4.767862	2.14E-06	2.13E-03
	305	4.595275	4.89E-06	4.85E-03
	981	4.541512	6.28E-06	6.23E-03
	259	4.253437	2.31E-05	2.29E-02
	332	4.108359	4.32E-05	4.28E-02
ds18_sum	599	9.285652	9.89E-20	9.88E-17
	595	8.746687	9.30E-18	9.29E-15
	596	7.423177	2.46E-13	2.46E-10
	404	5.120285	3.67E-07	3.66E-04
	203	4.811727	1.73E-06	1.73E-03
	355	4.666795	3.48E-06	3.48E-03
	672	4.229932	2.55E-05	2.55E-02
n22_sum	596	8.118744	1.39E-15	1.39E-12
	831	7.594252	7.15E-14	7.14E-11
	26	6.124015	1.31E-09	1.31E-06
	595	5.769181	1.06E-08	1.06E-05
	599	5.550784	3.65E-08	3.65E-05
	342	4.243874	2.40E-05	2.40E-02
n22ratio	883	7.287442	6.43E-13	6.42E-10
	753	6.118716	1.35E-09	1.35E-06

	271	4.540355	6.30E-06	6.30E-03
	991	4.134645	3.85E-05	3.85E-02
n23_sum	596	12.41506	5.29E-33	5.28E-30
	599	8.869389	3.36E-18	3.35E-15
	595	8.018834	2.99E-15	2.99E-12
	831	6.578435	7.68E-11	7.67E-08
	844	6.124421	1.31E-09	1.31E-06
	594	5.357374	1.05E-07	1.05E-04
	26	4.495651	7.75E-06	7.75E-03
n24_sum	596	8.628421	2.45E-17	2.45E-14
	595	7.807449	1.48E-14	1.48E-11
	844	6.00839	2.63E-09	2.63E-06
	653	4.968157	7.96E-07	7.95E-04
	599	4.760064	2.22E-06	2.22E-03
n24ratio	109	7.671797	4.04E-14	4.03E-11
	753	7.512292	1.29E-13	1.29E-10
	883	7.383437	3.26E-13	3.25E-10
	574	7.157633	1.59E-12	1.59E-09
	692	6.998645	4.75E-12	4.74E-09
	751	4.671409	3.40E-06	3.40E-03
	827	4.436248	1.02E-05	1.02E-02
n24s19ratio	827	7.048274	3.39E-12	3.39E-09
	842	5.338887	1.16E-07	1.16E-04
	619	5.23828	1.98E-07	1.98E-04
	753	5.004622	6.62E-07	6.62E-04
	651	4.844443	1.47E-06	1.47E-03
	692	4.457115	9.26E-06	9.25E-03
n24s20ratio	735	8.437353	1.14E-16	1.14E-13
	692	6.942147	6.97E-12	6.96E-09
	827	4.983244	7.38E-07	7.36E-04
	574	4.775123	2.07E-06	2.06E-03
	581	4.421668	1.09E-05	1.09E-02
	271	4.363929	1.41E-05	1.41E-02
	651	4.315933	1.75E-05	1.75E-02
n25_sum	599	5.94109	3.91E-09	3.91E-06
	595	4.980431	7.48E-07	7.47E-04
	355	4.766589	2.15E-06	2.15E-03
	95	4.148188	3.64E-05	3.63E-02

	779	4.129034	3.95E-05	3.95E-02
n26_sum	553	5.783739	9.79E-09	9.78E-06
n26ratio	753	13.88549	3.44E-40	3.43E-37
	733	6.094678	1.57E-09	1.57E-06
	868	4.573627	5.40E-06	5.39E-03
	191	4.483808	8.19E-06	8.18E-03
	893	4.247346	2.37E-05	2.36E-02
	764	4.225972	2.60E-05	2.60E-02
	585	4.195697	2.96E-05	2.96E-02
nds_sum	599	7.700173	3.29E-14	3.28E-11
	595	7.654052	4.62E-14	4.61E-11
	355	6.44541	1.80E-10	1.80E-07
	596	5.874127	5.80E-09	5.79E-06
	404	5.835031	7.28E-09	7.27E-06
	672	4.582881	5.17E-06	5.16E-03
ndssumc18dsratio	404	22.341866	6.46E-90	6.45E-87
	411	10.061077	9.64E-23	9.62E-20
	594	9.385834	4.12E-20	4.11E-17
	407	8.156394	1.03E-15	1.03E-12
	599	6.139521	1.19E-09	1.19E-06
	672	5.326494	1.24E-07	1.24E-04
	355	5.146565	3.20E-07	3.19E-04
	639	4.288581	1.97E-05	1.97E-02
ns18sumc18sratio	640	13.385619	1.14E-37	1.13E-34
	727	7.634024	5.36E-14	5.31E-11
	653	6.86594	1.17E-11	1.16E-08
	29	5.56018	3.47E-08	3.44E-05
	456	-4.954169	8.54E-07	8.47E-04
	484	4.845481	1.47E-06	1.45E-03
	639	4.839013	1.51E-06	1.50E-03
	423	4.43641	1.02E-05	1.01E-02
	547	-4.406201	1.17E-05	1.16E-02
	638	4.38314	1.30E-05	1.28E-02
ns_sum	596	9.110187	4.43E-19	4.42E-16
	595	7.820167	1.34E-14	1.34E-11
	844	6.232165	6.79E-10	6.78E-07
	599	5.422005	7.40E-08	7.39E-05
	594	4.39083	1.25E-05	1.25E-02

	653	4.163639	3.40E-05	3.40E-02
nssum_c18sratio	640	12.220579	4.41E-32	4.37E-29
	653	7.808006	1.48E-14	1.47E-11
	727	7.039539	3.61E-12	3.58E-09
	639	6.603542	6.55E-11	6.50E-08
	655	4.867312	1.32E-06	1.31E-03
	421	4.861663	1.35E-06	1.34E-03
	455	-4.784516	1.98E-06	1.96E-03
	456	-4.777255	2.05E-06	2.03E-03
	29	4.666819	3.48E-06	3.45E-03
	408	4.484458	8.17E-06	8.10E-03
ratio16to24	239	26.395366	9.35E-117	9.34E-114
	267	9.080896	5.68E-19	5.67E-16
	225	6.068046	1.84E-09	1.84E-06
	265	6.051848	2.03E-09	2.03E-06
	238	5.538072	3.92E-08	3.91E-05
	182	4.234334	2.51E-05	2.50E-02
	228	-4.186243	3.09E-05	3.08E-02
ratio20to24	467	5.520404	4.32E-08	0.00004317
	885	5.129255	3.50E-07	0.00034961
ratio22to24	161	5.578405	3.13E-08	3.12E-05
	158	5.100305	4.06E-07	4.05E-04
s18_sum	596	12.243092	6.65E-32	5.82E-29
	595	9.187652	2.90E-19	2.54E-16
	599	6.583464	7.95E-11	6.96E-08
	844	6.116135	1.45E-09	1.27E-06
	984	4.279522	2.08E-05	1.82E-02
s19_sum	751	8.567716	4.00E-17	4.00E-14
	753	8.06741	2.07E-15	2.06E-12
	718	7.128013	1.96E-12	1.96E-09
	717	5.267927	1.69E-07	1.69E-04
	741	5.055546	5.11E-07	5.11E-04
	653	4.686064	3.17E-06	3.17E-03
	844	4.469617	8.74E-06	8.73E-03
	639	4.116907	4.16E-05	4.16E-02
s20_sum	577	5.322199	1.27E-07	0.00012672
	599	4.121325	4.08E-05	0.040785

sumEA	410	4.595995	4.86E-06	0.0048583
	81	4.401182	1.19E-05	0.011923
	541	4.209544	2.79E-05	0.027888
sumc18	510	10.273024	1.35E-23	1.35E-20
	171	4.996389	6.90E-07	6.89E-04
	59	4.901597	1.11E-06	1.11E-03
	369	4.674245	3.36E-06	3.35E-03
	986	4.46661	8.86E-06	8.85E-03
	365	4.33922	1.58E-05	1.57E-02
	364	4.195682	2.96E-05	2.96E-02
sumcer	596	9.340987	6.10E-20	6.08E-17
	595	8.250635	4.98E-16	4.96E-13
	844	6.025281	2.38E-09	2.37E-06
	599	5.926562	4.26E-09	4.25E-06
	594	4.520467	6.91E-06	6.89E-03
sumn24ds	595	7.549455	9.90E-14	9.89E-11
	599	7.513744	1.28E-13	1.28E-10
	355	6.356146	3.15E-10	3.15E-07
	404	5.903611	4.88E-09	4.88E-06
	596	5.873272	5.83E-09	5.82E-06
	672	4.82596	1.61E-06	1.61E-03
	203	4.568054	5.54E-06	5.54E-03
sumn24s	596	8.416516	1.35E-16	1.34E-13
	595	7.423524	2.45E-13	2.45E-10
	844	6.298841	4.50E-10	4.50E-07
	653	5.192108	2.52E-07	2.52E-04
	599	4.343841	1.54E-05	1.54E-02
	753	4.226008	2.60E-05	2.60E-02
sumn22s	596	7.815207	1.40E-14	1.40E-11
	831	7.600732	6.82E-14	6.81E-11
	26	6.221072	7.28E-10	7.27E-07
	595	5.222433	2.15E-07	2.15E-04
	599	5.125262	3.57E-07	3.57E-04
	342	4.127464	3.98E-05	3.97E-02
nssumndssumratio	883	8.275823	4.08E-16	4.07E-13
	692	7.091178	2.52E-12	2.52E-09
	109	5.654825	2.04E-08	2.03E-05

	574	5.478065	5.45E-08	5.43E-05
	827	5.343324	1.13E-07	1.13E-04
sumn26s	553	5.85456	6.50E-09	6.49E-06
totalsphingo	596	9.374469	4.54E-20	4.54E-17
	595	8.263246	4.50E-16	4.49E-13
	844	5.991559	2.90E-09	2.90E-06
	599	5.956585	3.57E-09	3.57E-06
	594	4.573685	5.40E-06	5.39E-03