Genetically-predicted effects of LDL cholesterol lowering therapeutic targets on the risk of diabetes



Katerina Lind

St. Catherine's College, University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2022

Abstract

Genetically-predicted effects of LDL cholesterol lowering therapeutic targets on the risk of diabetes

Katerina Lind, St Catherine College, University of Oxford

A thesis submitted for the degree of Doctor of Philosophy, Trinity 2022

Background: Statins, which lower circulating LDL-C by inhibiting HMG CoA reductase, are a first-line treatment for prevention of cardiovascular disease. Recent clinical and genetic studies have shown that statins modestly increase the risk of incident type 2 diabetes (T2DM). The biological mechanisms of statin-induced T2DM remain unknown. It also remains unclear if patients with risk factors for diabetes are at higher proportional risk of statin-induced T2DM than those without risk factors, as some studies have claimed. Evidence for the effect of another LDL-C lowering drug, ezetimibe, on T2DM is insufficient, but if related to LDL-C, it may also impact on T2DM risk.

Methods: *HMGCR* and *NPC1L1* genetic risk scores (GRS) were used as statin and ezetimibe genetic proxies, respectively. The principles of a drug-target Mendelian randomisation were used to examine the effect of 1 SD (~1 mmol/L) lower genetically predicted LDL-C on T2DM, glycaemic traits and anthropometric measures in approximately 330,000 white British participants in UK Biobank. The heterogeneity of *HMGCR* GRS associations with T2DM and diabetes-related traits was also assessed in subgroup analyses among participants with varying susceptibilities (both genetic and clinical) to T2DM.

Results: 1 SD lower *HMGCR*-predicted LDL-C was associated with a 26% (95% CI 8%, 47%) higher risk of total T2DM, higher HbA1c (0.05 SD 95% CI 0.02, 0.08) and higher levels of adiposity. There was no evidence to support the hypothesis that the proportional effect of *HMGCR* GRS on T2DM in participants was greater in individuals with T2DM risk factors, and instead some subgroup analyses suggested a trend towards lower proportional risk of *HMGCR*-associated T2DM in those at higher risk of T2DM. *NPC1L1* GRS showed a similar but non-significant association with T2DM per 1 SD lower genetically predicted LDL-C to that observed for *HMGCR* and a strong association with HbA1c (0.15 SD, 95% CI 0.09, 0.22) per 1 SD lower *NPC1L1*-predicted LDL-C.

Conclusions: An *HMGCR* GRS associated with lower LDL-C levels was associated with higher T2DM, HbA1c levels and measures of adiposity. Although most cases of T2DM occurred in participants at elevated risk of T2DM, there was no excess proportional risk of T2DM in this subgroup compared to participants at lower risk.

The association of *NPC1L1* with T2DM was underpowered, but directionally consistent with the associations observed for *HMGCR*, and the *NPC1L1* GRS was also associated with higher HbA1c levels but not with higher levels of adiposity.

Acknowledgements

I am extremely grateful to the UK Biobank participants for their vital contribution.

I would like to express my deepest appreciation to my supervisors, Professor Jemma C. Hopewell, Professor Sarah Parish, and Associate Professor David Preiss for their invaluable patience, guidance, and encouragement.

I am also grateful to the members of Hopewell Group especially Dr Alex Stiby, Dr Elsa Valdes-Marquez, Dr Maysson Ibrahim, and Federico Murgia for helping and supporting me through the years.

Finally, I would like to thank the Global Education Programme for providing funding to enable me to undertake this DPhil.

Table of Contents

ABSTRACT..... II

Acknowledgements.....iv

Table of contents.....v

List of abbreviations..... ix

List of figures.....x

List of Tables..... xiv

CHAPTER 1 INTRODUCTION 19

1.1 Epidemiology of cardiovascular disease and statin therapy 19

1.2 Epidemiology of T2DM 26

1.3 Genomics in drug discovery and development.....28

1.4 The need for further research.....35

1.5 Objectives of the thesis.....38

CHAPTER 2 SYSTEMATIC LITERATURE REVIEW.....39

2.1 Objectives of the systematic review 39

2.2 Methods 41

2.3 Results 46
2.3.1 Search 1: *HMGCR* genetic studies 46
2.3.2 Search 2: Meta-analyses of RCTs of statins.....52
2.3.3 Search 3: *NPC1L1* genetic studies 57
2.3.4 Search 4: RCTs of ezetimibe 59

2.4 Discussion.....61

2.5 Summary.....65

CHAPTER 3 OVERALL STUDY DESIGN AND METHODS.....66

3.1 UK Biobank.....66
3.1.1 UK Biobank participant recruitment and baseline assessment.....66
3.1.2 Research ethics and permissions 67

3.2 Data collection and procedures 69
3.2.1 UK Biobank touch screen questionnaire.....69
3.2.2 UK Biobank nurse-led verbal interview 69
3.2.3 UK Biobank anthropometric measurements 70
3.2.4 UK Biobank blood biomarkers.....71

3.2.5 Hospital episode statistics and death registry 72
3.2.6 Genetic data 73

3.3 Type 2 diabetes mellitus phenotype definition 75
3.3.1 Prevalent cases of T2DM 76
3.3.2 Incident cases of T2DM 77
3.3.3 Total T2DM cases 77

3.4 Genetic consortia 78
3.4.1 The Global Lipids Genetics Consortium 78
3.4.2 The DIAbetes Genetics Replication And Meta-analysis 79
3.4.3 The Meta-Analyses of Glucose and Insulin-related traits Consortium 79
3.4.4 The Genetic Investigation of Anthropometric measures 79

3.5 Genetic methods 81
3.5.1 Drug target MR 81
3.5.2 *HMGCR* genetic risk score 82
3.5.3 *NPC1L1* genetic risk score 86

3.6 Statistical methods 88
3.6.1 Main UK Biobank analysis 88
3.6.2 External validation analysis 90

CHAPTER 4 THE ASSOCIATIONS OF *HMGCR* AND *NPC1L1* GRSS WITH LDL-C AND APO B-100 IN UK BIOBANK 93

4.1 Introduction 93
4.1.1 Genetic contributions to LDL-C and relation to Apo-B100 93
4.1.2 Genetic risk score as an instrumental variable 94

4.2 Study design and method 96
4.2.1 Study design 96
4.2.2 Study population 97
4.2.3 Correction for statin use 97
4.2.4 Statistical methods 98

4.3 Results 101
4.3.1 Lipid distributions in UK Biobank population 101
4.3.2 The associations of *HMGCR* index SNPs with LDL-C concentrations 104
4.3.3 The association of *HMGCR* GRS with LDL-C concentration 106
4.3.4 The associations of *NPC1L1* index SNPs with LDL-C concentrations 107
4.3.5 The associations of *NPC1L1* GRS with LDL-C concentration 108
4.3.6 The associations of *HMGCR* and *NPC1L1* GRSs with Apo B-100 concentrations 108

4.4 Discussion 111

CHAPTER 5 THE ASSOCIATIONS OF THE *HMGCR* GRS WITH THE RISK OF TYPE 2 DIABETES MELLITUS, GLYCAEMIC TRAITS, AND ANTHROPOMETRIC MEASURES 114

5.1 Introduction 114

5.2 Study design and method 117

5.2.1 Study population....117
5.2.2 Baseline assessment....117
5.2.3 Total and incident T2DM....118
5.2.4 Statistical methods....118

5.3 Results....119
5.3.1 Baseline characteristics of the study population....119
5.3.2 *HMGCR* GRS and risk of T2DM....123
5.3.3 *HMGCR* GRS and glycaemic traits....127
5.3.4 *HMGCR* GRS and measures of central and general adiposity....130

5.4 Discussion....136
5.4.1 The association of *HMGCR* GRS with T2DM....136
5.4.2 The associations of *HMGCR* GRS with glycaemic traits....139
5.4.3 The associations of *HMGCR* GRS with anthropometric measures....140

5.5 Summary....143

CHAPTER 6 THE ASSOCIATIONS OF THE *HMGCR* GRS WITH THE RISK OF T2DM, GLYCAEMIC TRAITS AND ANTHROPOMETRIC MEASURES BY THE BASELINE RISK OF T2DM....144

6.1 Introduction....144

6.2 Study design and method....148
6.2.1 Study population....148
6.2.2 T2DM risk scores....149
6.2.3 Statistical methods....152

6.3 Results....156
6.3.1 *HMGCR* GRS and total T2DM in participants with different genetic risks of T2DM....156
6.3.2 *HMGCR* GRS and incident T2DM in participants with different absolute risks of T2DM....160
6.3.3 *HMGCR* GRS and the risk of incident T2DM by HbA1c levels....164
6.3.4 *HMGCR* GRS and glycaemic traits in participants with different risks of T2DM....166
6.3.5 *HMGCR* GRS and anthropometric measures in participants with different risks of T2DM....168
6.3.6 *HMGCR* GRS and anthropometric measures....168

6.4 Discussion....173
6.4.1 The effect of *HMGCR* GRS on T2DM....173
6.4.2 The effects of the *HMGCR* GRS on glycaemic traits and anthropometric measures by risk of diabetes....179

6.5 Summary....180

CHAPTER 7 THE ASSOCIATIONS OF THE *NPC1L1* GRS WITH RISK OF T2DM, GLYCAEMIC TRAITS AND ANTHROPOMETRIC MEASURES....181

7.1 Introduction....181

7.2 Study design and method....184
7.2.1 Study population....184
7.2.2 Baseline assessment....184
7.2.3 *NPC1L1* genetic risk score....185

7.2.4 Statistical methods 185

7.3 Results 186
7.3.1 *NPC1L1* GRS and risk of T2DM 186
7.3.2 *NPC1L1* GRS and glycaemic traits 190
7.3.3 *NPC1L1* GRS and measures of central and general adiposity 192

7.4 Discussion 200
7.4.1 The association of *NPC1L1* GRS with T2DM 200
7.4.2 The associations of *NPC1L1* GRS with glycaemic traits 202
7.4.3 The associations of *NPC1L1* GRS with anthropometric measures 203
7.4.4 Limitations of the current study 204

7.5 Summary 205

CHAPTER 8 DISCUSSION 206

8.1 Introduction 206

8.2 Summary of findings 209
8.2.1 The associations of *HMGCR* and *NPC1L1* with T2DM 209
8.2.2 The associations of *HMGCR* and *NPC1L1* with anthropometric traits 222
8.2.3 Strengths and limitations of the study 223
8.2.4 Implications 226
8.2.5 Further research 229

APPENDICES 231

Appendices Chapter 1 231

Appendices Chapter 2 231
2.1 Appendix 2:1 | The associations of genetic variants in the *HMGCR* gene with T2DM, anthropometric measures and glycaemic traits 232
2.2 Appendix 2:2 | The associations of SNPs in the *HMGCR* gene with T2DM, LDL-C and glycaemic/adiposity traits 238
2.3 Appendix 2:3 | The associations of genetic variants in the *NPC1L1* gene with T2DM, anthropometric measures and glycaemic traits 241
2.4 Appendix 2:4 | The associations of SNPs in the *NPC1L1* gene with T2DM, LDL-C and glycaemic/adiposity traits 245
2.5 Appendix 2:5 | Statins RCTs included in the meta-analyses 246
2.6 Appendix 2:6 | RCTs of statins included in each meta-analysis 248

Appendices Chapter 3 250
3.1 Appendix 3:1 | The list of oral hypoglycaemic agents and insulin by class with UK Biobank codes 251
3.2 Appendix 3:2 | UK Biobank non-cancer illnesses codes for diabetes 253
3.3 Appendix 3:3 | List of ICD-10 codes for diabetes 254
3.4 Appendix 3:4 | The algorithm for defining prevalent diabetes cases 256
3.5 Appendix 3:5 | The algorithms for discrimination between type 1 and type 2 diabetes 257
3.6 Appendix 3:6 | The algorithm computing incident cases of T2DM 260
3.7 Appendix 3:7 | SNPs within 100 kb from the *HMGCR* gene which are associated with LDL-C at GWAS level of significance 261

3.8 Appendix 3:8 | The results of the *HMGCR* SNPs clumping procedure....265
3.9 Appendix 3:9 | The distribution of *HMGCR* GRS....266
3.10 Appendix 3:10 | SNPs within 100 kb from the *NPC1L1* gene which are associated with LDL-C at GWAS level of significance....267
3.11 Appendix 3:11| The results of the *NPC1L1* SNPs clumping procedure....268
3.12 Appendix 3:12 | The distribution of the *NPC1L1* GRS....269

Appendices Chapter 4....270
4.1 Appendix 4:1 | The list of LDL-C lowering medications and the UK Biobank medication codes 271
4.2 Appendix 4:2 | LDL-C and Apo B-100 levels in participants on-statin and not on-statin....272

Appendices Chapter 5....273
5.1 Appendix 5:1 | List of UKB-specific hypoglycaemic agents by classes....274
5.2 Appendix 5:2 | Fixed-effect meta-analyses of the associations of *HMGCR* genetic variants with risk of T2DM in UK Biobank and the DIAGRAM consortium....276
5.3 Appendix 5:3 | Fixed-effect meta-analyses of the associations of *HMGCR* genetic variants with HbA1c in UK Biobank and the MAGIC consortium....277
5.4 Appendix 5:4 | The effects of the *HMGCR* SNPs on anthropometric measures in UK Biobank and the GIANT consortium....278
5.5 Appendix 5:5 | Fixed-effect meta-analyses of the associations of *HMGCR* genetic variants with BMI in UK Biobank and the GIANT consortium....280
5.6 Appendix 5:6 | Fixed-effect meta-analyses of the associations of *HMGCR* genetic variants with waist circumference in UK Biobank and the GIANT consortium....281
5.7 Appendix 5:7 | Fixed-effect meta-analyses of the associations of *HMGCR* genetic variants with hip circumference in UK Biobank and the GIANT consortium....282
5.8 Appendix 5:8 | Fixed-effect meta-analyses of the associations of *HMGCR* genetic variants with waist-to-hip ratio in UK Biobank and the GIANT consortium....283
5.9 Appendix 5:9 | Fixed-effect meta-analyses of the associations of *HMGCR* genetic variants with height in UK Biobank and the GIANT consortium....284

Appendices Chapter 6....285
6.1 Appendix 6:1 | 57 SNPs associated with higher risk of T2DM, Morrison meta-analysis and UK Biobank genetic data....286
6.2 Appendix 6:2 | The distribution of the polygenic T2DM GRS....288
6.3 Appendix 6:3 | The distribution of QDiabetes-2018 RS....289
6.4 Appendix 6:4 | Density Plot and Histogram of HbA1c distribution: a comparison of natural units with adjusted for *HMGCR* GRS residuals....290
6.5 Appendix 6:5 | The correlation between HbA1c in mmol/mol and HbA1c residuals....291
6.6 Appendix 6:6 | The associations of anthropometric measures (in SD units) with the *HMGCR* GRS in the subgroups stratified by sex and T2DM GRS....292

Appendices Chapter 7....294
7.1 Appendix 7:1 | Fixed-effect meta-analyses of the associations of *NPC1L1* genetic variants with risk of T2DM in UK Biobank and the DIAGRAM consortium....295
7.2 Appendix 7:2 | Fixed-effect meta-analyses of the associations of *NPC1L1* genetic variants with HbA1c in UK Biobank and the MAGIC consortium....296
7.3 Appendix 7:3 | The effects of the *NPC1L1* SNPs on anthropometric measures in UK Biobank and the GIANT consortium....297
7.4 Appendix 7:4 | Fixed-effect meta-analyses of the associations of *NPC1L1* genetic variants with BMI in UK Biobank and the GIANT consortium....298
7.5 Appendix 7:5 | Fixed-effect meta-analyses of the associations of *NPC1L1* genetic variants with waist circumference in UK Biobank and the GIANT consortium....299

7.6 . Appendix 7:6 | Fixed-effect meta-analyses of the associations of *NPC1L1* genetic variants with hip circumference in UK Biobank and the GIANT consortium.....300
7.7 Appendix 7:7 | Fixed-effect meta-analyses of the associations of *NPC1L1* genetic variants with waist-to-hip ratio in UK Biobank and the GIANT consortium.....301
7.8 Appendix 7:8 | Fixed-effect meta-analyses of the associations of *NPC1L1* genetic variants with height in UK Biobank and the GIANT consortium 302

REFERENCES: 303

List of abbreviations:

The following abbreviations were used in the thesis:

Abbreviation	Description
Apo B100	Apolipoprotein B
BMI	Body mass index
CI	Confidence Intervals
CVD	Cardiovascular disease
FG	Fasting glucose
GRS	Genetic risk score
HbA1c	Glycated haemoglobin
HDL-C	High-density lipoprotein cholesterol
HES	Hospital episode statistics
HMGCR	3-Hydroxy-3-methylglutaryl- coenzyme-A reductase
HR	Hazard Ratio
LD	Linkage disequilibrium
LDL-C	Low-density lipoprotein cholesterol
MAF	Minor allele frequency
MR	Mendelian randomisation
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NPC1L1	Niemann-Pick C1-Like1 protein
OR	Odds Ratio
RCT	Randomised controlled trial
RR	Relative Risk
RS	Risk score
SD	Standard Deviation
SE	Standard Error
SIGN	Scottish Intercollegiate Guidelines Network
SNP	Single nucleotide polymorphism
T2DM	Type 2 diabetes mellitus
WHR	Waist-to-hip ratio

List of figures:

Title	Page no.

FIGURE 1:1 | STATIN (ITEMS) DISPENSED PER 1,000 POPULATION 21

FIGURE 1:2 | EZETIMIBE AND SIMVASTATIN AND EZETIMIBE (ITEMS) DISPENSED PER 1,000 POPULATION 22

FIGURE 1:3 | THE COMPARISON BETWEEN AN RCT AND THE MENDELIAN RANDOMISATION APPROACH 31

FIGURE 2:1 | PRISMA FLOW DIAGRAM FOR THE SEARCH 1 (*HMGCR* GENETIC STUDIES) 46

FIGURE 2:2 | PRISMA FLOW DIAGRAM FOR SEARCH 2 (META-ANALYSES OF RCTS OF STATINS) 53

FIGURE 2:3 | PRISMA FLOW DIAGRAM FOR SEARCH 3 (*NPC1L1* GENETIC STUDIES) 57

FIGURE 2:4 | PRISMA FLOW DIAGRAM FOR SEARCH 4 (RCTS OF EZETIMIBE) 60

FIGURE 3:1 | ALGORITHM FOR SELECTING THE *HMGCR* INDEX SNPS 84

FIGURE 3:2 | ALGORITHM OF THE *NPC1L1* INDEX SNPS SELECTION 87

FIGURE 3:3 | ALGORITHM FOR SELECTING PARTICIPANTS FOR THE ANALYSIS 88

FIGURE 4:1 | COMPARISON OF PLASMA LDL-C DISTRIBUTION IN STATIN USERS AND NON-USERS 101

FIGURE 4:2 | OVERALL PLASMA LDL-C DISTRIBUTION PRIOR AND POST CORRECTION FOR STATIN USE 102

FIGURE 4:3 | COMPARISON OF PLASMA APO B-100 DISTRIBUTION IN STATIN USERS AND NON-USERS 103

FIGURE 4:4 | OVERALL PLASMA APO B-100 DISTRIBUTION PRIOR AND POST CORRECTION FOR STATIN USE 104

FIGURE 4:5 | ASSOCIATIONS OF GENETIC VARIANTS WITH LDL-C IN THE UK BIOBANK AND GLGC 105

FIGURE 4:6 | ASSOCIATIONS OF *HMGCR* GRS WITH BLOOD LIPIDS 106

FIGURE 4:7 | ASSOCIATIONS OF GENETIC VARIANTS IN *NPC1L1* GRS WITH LDL-C 107
FIGURE 4:8 | ASSOCIATIONS OF *NPC1L1* GRS WITH BLOOD LIPIDS..... 108
FIGURE 4:9 | ASSOCIATIONS OF GENETIC VARIANTS IN *HMGCR* WITH LDL-C (SD) AND APO B-100 (SD) CONCENTRATIONS 109
FIGURE 4:10 | ASSOCIATIONS OF GENETIC VARIANTS IN *NPC1L1* WITH LDL-C (SD) AND APO B-100 (SD) CONCENTRATIONS 109
FIGURE 4:11 | ASSOCIATIONS OF HMGCR GRS AND *NPC1L1* GRS WITH APO B-100 110
FIGURE 5:1 | ASSOCIATIONS OF THE *HMGCR* INDEX SNPS WITH THE RISK OF TOTAL T2DM AND LDL-C CONCENTRATIONS 123
FIGURE 5:2 | ASSOCIATIONS OF *HMGCR* GRS WITH THE RISK OF TOTAL T2DM STRATIFIED BY SEX 125
FIGURE 5:3 | META-ANALYSIS OF THE EFFECTS OF *HMGCR* GRS ON RISK OF T2DM IN UK BIOBANK AND THE DIAGRAM CONSORTIUM..... 126
FIGURE 5:4 | EFFECTS OF *HMGCR* GRS ON RISK OF T2DM IN UK BIOBANK AND THE DIAGRAM CONSORTIUM STRATIFIED BY SEX..... 127
FIGURE 5:5 | ASSOCIATIONS OF THE GENETIC VARIANTS IN *HMGCR* WITH GLYCAEMIC TRAITS..... 128
FIGURE 5:6 | THE ASSOCIATION OF *HMGCR* GRS WITH GLYCAEMIC TRAITS PER 1 SD LOWER GENETICALLY PREDICTED LDL-C..... 128
FIGURE 5:7 | META-ANALYSIS OF THE EFFECT OF *HMGCR* GRS ON HBA1C IN UK BIOBANK AND THE MAGIC CONSORTIUM 130
FIGURE 5:8 | ASSOCIATIONS OF GENETIC VARIANTS IN *HMGCR* WITH ANTHROPOMETRIC MEASURES 132
FIGURE 5:9 | ASSOCIATIONS OF THE *HMGCR* GRS WITH ANTHROPOMETRIC MEASURES PER ONE SD LOWER GENETICALLY PREDICTED LDL-C..... 134
FIGURE 5:10 | EFFECTS OF HMGCR GRS ON ANTHROPOMETRIC MEASURES IN THE GIANT CONSORTIUM AND UK BIOBANK STRATIFIED BY SEX..... 135
FIGURE 6:1 | POLYGENIC T2DM RISK SCORE COMPUTATION ALGORITHM 151
FIGURE 6:2 | RISK OF T2DM IN THREE RISK-STRATIFIED SUBGROUPS (BY TERTILES OF POLYGENIC T2DM RISK SCORE)..... 159

FIGURE 6:3 | RISK OF TOTAL T2DM ASSOCIATED WITH 1 SD LOWER *HMGCR* PREDICTED LDL-C (BY TERTILES OF THE POLYGENIC T2DM RISK SCORE) 160
FIGURE 6:4 | RISK OF INCIDENT T2DM IN THREE RISK-STRATIFIED SUBGROUPS (BY QDIABETES-2018 RISK SCORE) 163
FIGURE 6:5 | RISK OF INCIDENT T2DM ASSOCIATED WITH 1 SD LOWER GENETICALLY PREDICTED LOWER LDL-C (BY QDIABETES-2018 RISK GROUPS) 164
FIGURE 6:6 | ASSOCIATION OF *HMGCR* GRS WITH INCIDENT T2DM BY HBA1C (MMOL/MOL) LEVEL 165
FIGURE 6:7 | ASSOCIATION OF *HMGCR* GRS WITH INCIDENT T2DM BY HBA1C (RESIDUALS ADJUSTED FOR *HMGCR* GRS) 166
FIGURE 6:8 | ASSOCIATIONS OF *HMGCR* GRS WITH HBA1C CONCENTRATION IN THE SUBGROUPS BY TERTILES OF T2DM GRS 167
FIGURE 6:9 | ASSOCIATIONS OF *HMGCR* GRS WITH BLOOD GLUCOSE LEVEL CONCENTRATION IN THE SUBGROUPS BY TERTILES OF T2DM GRS 168
FIGURE 6:10 | ASSOCIATIONS OF *HMGCR* GRS WITH ANTHROPOMETRIC MEASURES STRATIFIED BY T2DM GENETIC RISK GROUPS PER 1 SD *HMGCR* PREDICTED LOWER LDL-C. 170
FIGURE 6:11 | ASSOCIATIONS OF *HMGCR* GRS WITH ANTHROPOMETRIC MEASURES STRATIFIED BY T2DM GENETIC RISK GROUPS IN MEN PER 1 SD LOWER GENETICALLY PREDICTED LDL-C 171
FIGURE 6:12 | ASSOCIATIONS OF *HMGCR* GRS WITH ANTHROPOMETRIC MEASURES STRATIFIED BY T2DM GENETIC RISK GROUPS IN WOMEN PER 1 SD LOWER GENETICALLY PREDICTED LDL-C 172
FIGURE 6:13 | RISK OF INCIDENT T2DM IN SUBGROUPS OF PARTICIPANTS WITH AND WITHOUT RISK FACTORS FOR T2DM FROM THE TNT+IDEAL AND JUPITER RCTS 176
FIGURE 6:14 | ASSOCIATIONS OF *HMGCR* GRS WITH INCIDENT T2DM IN PARTICIPANTS WITH NORMAL AND HIGH LEVELS OF FASTING GLUCOSE 177
FIGURE 7:1 | ASSOCIATIONS OF THE *NPC1L1* INDEX SNPS WITH THE RISK OF TOTAL T2DM AND LDL-C CONCENTRATIONS 186

FIGURE 7:2 | ASSOCIATIONS OF THE *NPC1L1* INDEX SNPS WITH TOTAL T2DM STRATIFIED BY SEX 187
FIGURE 7:3 | FIXED-EFFECTS META-ANALYSIS OF THE ASSOCIATION BETWEEN *NPC1L1* GRS WITH RISK OF TOTAL DIABETES IN A FIXED EFFECT MODEL (UK BIOBANK AND DIAGRAM CONSORTIUM) 189
FIGURE 7:4 | SEX-SPECIFIC EFFECTS OF *NPC1L1* GRS ON TOTAL T2DM IN UK BIOBANK AND THE DIAGRAM CONSORTIUM 189
FIGURE 7:5 | ASSOCIATIONS OF THE *NPC1L1* SNPS WITH HBA1C AND LDL-C 190
FIGURE 7:6 | ASSOCIATIONS OF THE *NPC1L1* SNPS WITH NON-FASTING GLUCOSE AND LDL-C. 190
FIGURE 7:7 | ASSOCIATIONS OF *NPC1L1* GRS WITH GLYCAEMIC TRAITS 191
FIGURE 7:8 | FIXED-EFFECT META-ANALYSES OF THE EFFECT OF *NPC1L1* GRS ON HBA1C IN UK BIOBANK AND THE MAGIC CONSORTIUM 192
FIGURE 7:9 | ASSOCIATIONS OF GENETIC VARIANTS IN *NPC1L1* WITH ANTHROPOMETRIC MEASURES 194
FIGURE 7:10 | ASSOCIATIONS OF THE *NPC1L1* GRS WITH ANTHROPOMETRIC MEASURES PER 1 SD LOWER GENETICALLY PREDICTED LDL-C 196
FIGURE 7:11 | FIXED-EFFECTS META-ANALYSIS OF THE ASSOCIATIONS BETWEEN *NPC1L1* GRS WITH ANTHROPOMETRIC MEASURES IN UK BIOBANK AND GIANT 198
FIGURE 7:12 | EFFECTS OF *NPC1L1* GRS ON ANTHROPOMETRIC MEASURES IN THE GIANT CONSORTIUM AND UK BIOBANK STRATIFIED BY SEX 199
FIGURE 8:1 | ASSOCIATIONS OF THE HMGCR SNPS WITH TOTAL T2DM (OR, 95% CI) IN UK BIOBANK, AND LDL-C (SD, 95 CI) IN GLGC 210
FIGURE 8:2 | ASSOCIATIONS OF THE HMGCR SNPS WITH HBA1C (SD, 95% CI) IN UK BIOBANK, AND LDL-C (SD, 95% CI) IN GLGC 210
FIGURE 8:3 | ASSOCIATIONS OF THE HMGCR SNPS WITH T2DM (OR, 95% CI) IN THE DIAGRAM CONSORTIUM, AND LDL-C (SD, 95 CI) IN GLGC 210
FIGURE 8:4 | ASSOCIATIONS OF THE HMGCR SNPS WITH HBA1C (SD, 95% CI) IN THE MAGIC CONSORTIUM, AND LDL-C (SD, 95% CI) IN GLGC 210

FIGURE 8:5 | ASSOCIATIONS OF THE NPC1L1 SNPS WITH TOTAL T2DM (OR, 95% CI) IN UK BIOBANK, AND LDL-C (SD, 95 CI) IN GLGC 211
FIGURE 8:6 | ASSOCIATIONS OF THE NPC1L1 SNPS WITH HBA1C (SD, 95% CI) IN UK BIOBANK, AND LDL-C (SD, 95% CI) IN GLGC 211
FIGURE 8:7 | ASSOCIATIONS OF THE NPC1L1 SNPS WITH T2DM (OR, 95% CI) IN THE DIAGRAM CONSORTIUM, AND LDL-C (SD, 95 CI) IN GLGC 211
FIGURE 8:8 | ASSOCIATIONS OF THE NPC1L1 SNPS WITH HBA1C (SD, 95% CI) IN THE MAGIC CONSORTIUM, AND LDL-C (SD, 95% CI) IN GLGC 211

List of Tables:

Title	Page no.
TABLE 2:1 | SYSTEMATIC REVIEW SEARCH TERMS FOR GENETIC STUDIES	41
TABLE 2:2 | SYSTEMATIC REVIEW SEARCH TERMS FOR META-ANALYSES AND RCTS	43
TABLE 2:3 | FREQUENCIES OF EACH *HMGCR* SNP ANALYSIS ACROSS THE SELECTED MANUSCRIPTS	52
TABLE 2:4 | FREQUENCIES OF EACH *NPC1L1* SNP ANALYSIS ACROSS THE SELECTED MANUSCRIPTS	59
TABLE 3:1 | *HMGCR* INDEX SNPS IN THE UK BIOBANK, GLGC AND ENSEMBL	85
TABLE 3:2 | *NPC1L1* INDEX SNPS IN THE UK BIOBANK, GLGC AND ENSEMBL	87
TABLE 3:3 | POWER CALCULATIONS FOR *HMGCR* AND *NPC1L1* GRSS	90
TABLE 4:1 | MEAN LEVELS OF LDL-C AND APO B-100 IN UK BIOBANK PARTICIPANT ON-STATIN, NOT ON-STATIN AND OVERALL CORRECTED BY STATIN USE	102
TABLE 4:2 | ASSOCIATIONS OF THE *HMGCR* GENETIC VARIANTS WITH LDL-C IN UK BIOBANK AND GLGC	105
TABLE 4:3 | ASSOCIATIONS OF THE *NPC1L1* GENETIC VARIANTS WITH LDL-C IN UK BIOBANK AND GLGC	107
TABLE 4:4 | ASSOCIATIONS OF THE *HMGCR* AND *NPC1L1* GENETIC VARIANTS WITH APO B-100 IN UK BIOBANK	109
TABLE 5:1 | BASELINE CHARACTERISTICS OF 334,516 EUROPEAN UK BIOBANK PARTICIPANTS BY T2DM STATUS AT BASELINE ASSESSMENT AND SEX	120
TABLE 5:2 | ASSOCIATIONS OF THE *HMGCR* INDEX SNPS WITH TOTAL T2DM AND LDL-C CONCENTRATIONS	124
TABLE 5:3 | ASSOCIATIONS OF *HMGCR* GENETIC VARIANTS WITH RISK OF T2DM IN THE DIAGRAM CONSORTIUM AND UK BIOBANK	125

TABLE 5:4 | EFFECTS OF THE *HMGCR* SNPS ON HBA1C IN UK BIOBANK AND THE MAGIC CONSORTIUM 129
TABLE 6:1 | BASELINE CHARACTERISTICS OF 334,515 PARTICIPANTS OF THE UK BIOBANK IN THREE SUBGROUPS BY POLYGENIC T2DM RISK SCORE 157
TABLE 6:2 | BASELINE CHARACTERISTICS OF 320,622 PARTICIPANTS OF THE UK BIOBANK WHO WERE FREE FROM T2DM AT BASELINE IN THREE SUBGROUPS BY QDIABETES-2018 RISK SCORE 161
TABLE 7:1 | ASSOCIATIONS OF THE *NPC1L1* INDEX SNPS WITH TOTAL T2DM AND LDL-C CONCENTRATIONS 187
TABLE 7:2 | ASSOCIATIONS OF THE *NPC1L1* INDEX SNPS WITH TOTAL T2DM IN UK BIOBANK AND THE DIAGRAM CONSORTIUM 188
TABLE 7:3 | EFFECTS OF *NPC1L1* GRS ON HBA1C IN UK BIOBANK AND THE MAGIC CONSORTIUM 191

Chapter 1 Introduction

1.1 Epidemiology of cardiovascular disease and statin therapy

Atherosclerotic cardiovascular disease (CVD) remains the leading cause of mortality and morbidity worldwide[1]. Currently, it accounts for approximately one third of all deaths globally[2]. Over the last decade, the absolute number of deaths from CVD has grown by 12.5%[3] and reached 18.6 million in 2019[2]. However, the recent increase in number of deaths from CVD reflects the aging and growing population rather than an increase in mortality rates. Age-standardised analysis showed, that by year 2000, the age-standardized mortality rate from CVD had decreased by 30% (from 398 to 278 deaths per 100,000 standard population) compared to the 1960s[4]. The main factors contributing to the decline in CVD mortality were identification and treatment of risk factors[5] such as blood pressure, blood cholesterol, smoking and diabetes along with advances in surgical interventions[4], such as coronary artery bypass grafting and coronary stenting. The recognition of the central role of low-density lipoprotein cholesterol (LDL-C) in the formation of atherosclerotic plaques led to the development of a very effective LDL-C lowering class of drugs – the statins.

Statins work by inhibiting HMG-CoA (hydroxymethylglutaril-coenzyme A) reductase, the rate-limiting enzyme in the cholesterol biosynthesis pathway[6]. The effect of statins on LDL-C concentration depends on the intensity (dose) of the statin. According to The National Institute for Health and Care Excellence (NICE) guidance[7], statins are grouped into three intensity categories by the achieved percentage reduction in LDL-

C: low-intensity statins (e.g. 10 mg of simvastatin daily) reduce LDL-C concentrations by approximately 20%-30%, moderate-intensity statins (e.g. 20 mg of simvastatin) by 31% to 40%[8] and high-intensity statins (e.g. 40-80 mg of atorvastatin) by greater than 40%.

The Cholesterol Treatment Trialists' Collaboration meta-analyses of 170,000 participants from large-scale (more than 1000 participants), long-term (2 years or more) randomised controlled trials (RCT) of statins examined the effect of the reduction in LDL-C concentration on the risk of CVD morbidity and mortality[9]. Across 21 trials of statin vs. control, LDL-C difference at 1 year was - 1.07 mmol/L; five trials of more versus less intensive statin regimes showed a -0.51 mmol/L 1-year LDL-C difference. The proportional reductions in major vascular events per 1 mmol/L reduction in LDL-C was 21% (RR 0.79, 95% CI 0.77, 0.81) for statin vs. control in 21 trials[9]. The corresponding proportional vascular events risk reduction in the 5 more intensive vs. less statin trials was 28% (95% CI 22-34%)[9]. The authors concluded that a reduction in LDL-C of 2 mmol/L (which could be achieved with a high intensity statin regimen) could reduce the risk of vascular events by 40-50%[9].

Prescription guidelines and trends

Statins are among the most commonly prescribed medications in the UK[10]. The British Heart Foundation estimated that 7-8 million adults in the UK are currently taking statins for primary or secondary prevention of CVD[11]. **Figure 1:1** shows a trend in statin dispensing from community pharmacies in England between 1998 and 2018. The data were obtained from OpenPrescribing[12] and show the number of items dispensed per

1,000 population (each year was adjusted for England's mid-year population size). One item represents a month supply of statins.

From 1998 to 2008, there has been a rapid year-on-year increase in number of items. As a result, the number of dispensed items increased from 103 to 918 items per 1,000 population. The next decade has witnessed a slow but progressive increase in the number of dispensed items; in 2018, the number of dispensed items increased by 37% compared to 2008.

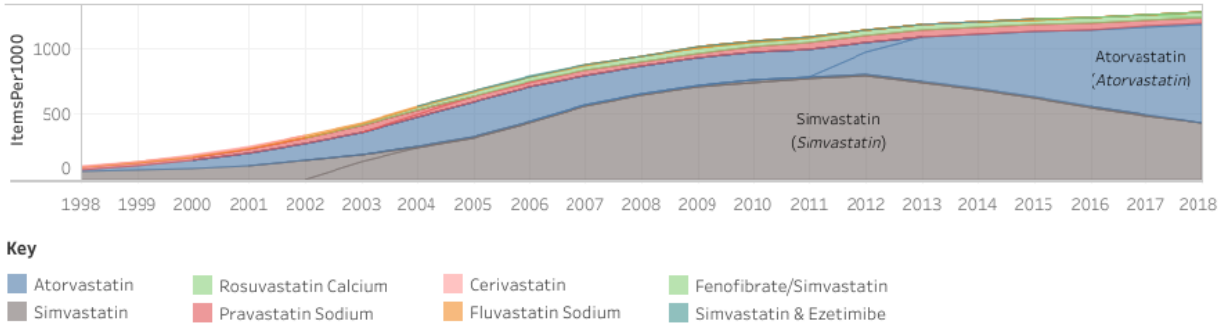


Figure 1:1 | Statin (Items) dispensed per 1,000 population
The figure was generated using an online method provided by OpenPrescibing with parameters: Chapter/section 2:12, statins only. OpenPrescribing.net, The DataLab, University of Oxford, 2022. The number of items is shown per 1,000 population adjusted for England's mid-year population size each year.

In 2007, NICE approved ezetimibe - a new type of lipid-lowering medication, often prescribed in conjunction with statin therapy for patients who cannot achieve a recommended level of LDL-C with a statin alone or who report that they cannot tolerate a statin. Ezetimibe's prescription steadily increased between 2007 and 2010 (**Figure 1:2**). After reaching the peak of 45 items per 1,000 population in 2010, there were five

years of decline until 2015, when the prescription rate declined to 30 items per 1,000 population and it has stabilised at this level since then.

The prescription of simvastatin and ezetimibe, as a combination therapy, increased to 3 items per 1,000 population within the first four years and remained on this level until 2011. However, the following three years witnessed a sharp decline in prescriptions to 20,477 items (0 per 1,000 population) per year in 2017/2018.

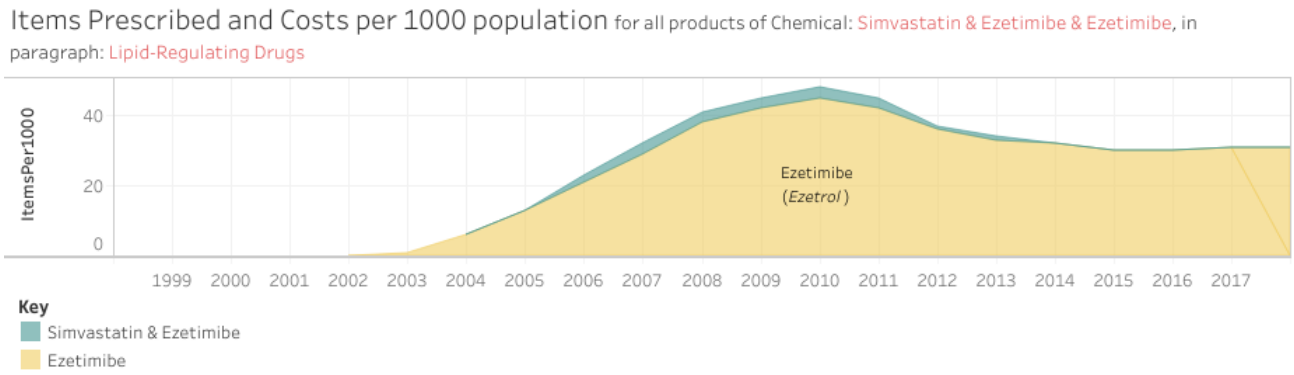


Figure 1:2 | Ezetimibe and simvastatin and ezetimibe (Items) dispensed per 1,000 population
The figure was generated using an online method provided by OpenPrescibing with parameters: Chapter/section 2:12, statins only. OpenPrescribing.net, The DataLab, University of Oxford, 2022. The number of items is shown per 1,000 population adjusted for England's mid-year population size each year.

According to NICE guidelines[13] and Scottish Intercollegiate Guidelines Network (SIGN) guidelines[14], all patients with established atherosclerotic CVD should be offered high intensity statin therapy (atorvastatin 80 mg). For primary prevention, NICE[13] currently recommends the initiation of statin therapy in patients with ≥10% risk of CVD in the next 10 years. Formal CVD risk assessment in England and Wales is done with the QRISK2[15] tool in patients from 30 to 84 years old. SIGN[14] guidance, which applies in Scotland, recommends the ASSIGN[16,17] risk prediction model for CVD

risk estimation, with statin therapy recommended for patients with 10-year absolute risk of the cardiovascular event ≥20%[14].

Safety and adverse effects of statins

Multiple large-scale and long-term RCTs of statins have demonstrated that statin therapy lowers LDL-C and reduces mortality and morbidity from atherosclerotic CVD[6].

Although RCTs have demonstrated that statins are safe and well-tolerated by most adults[18], there are two well documented adverse effects of statin therapy – myopathy and new-onset type 2 diabetes mellitus (T2DM)[8]. This study focuses on the effect of statins on new-onset T2DM.

The first detailed investigation of the effect of statins on T2DM came from the West of Scotland Coronary Prevention Study (WOSCOPS) [9], which randomised 5,974 middle-aged men with moderately elevated LDL-C to pravastatin 40 mg or placebo group. In a post hoc analysis, WOSCOPS reported a 30% decrease in the risk of incident T2DM on pravastatin versus placebo.

By contrast, JUPITER (Justification for the Use of Statins in Primary Prevention: An Intervention Trial Evaluating Rosuvastatin)[11], one of the largest placebo-controlled trials with nearly 18,000 participants, and the first RCT which pre-specified incident T2DM as an outcome, reported a relative increase of 24% in incident T2DM in the rosuvastatin group compared to the placebo group. The discrepancy between WOSCOPS and JUPITER results led to various attempts to pool the data across major statin RCTs.

Evidence of a causal relationship between statin therapy and an increased risk of T2DM emerged from a meta-analysis of 13 placebo-controlled and standard-care controlled RCTs of statins[19] published in 2010. The meta-analysis showed a 9% increase in risk of incident T2DM in the intervention group compared to the control group. It was followed by a meta-analysis of five RCTs of more intensive versus less intensive statin regimens by Preiss *et al.*[20] in 2011, which demonstrated a dose-dependent association between the statin therapy and the risk of incident T2DM. However, T2DM was a pre-specified outcome in a very small number of the RCTs included in this meta-analysis and, therefore, the definition of T2DM varied substantially across the trials.

It has been suggested that the relative risk of statin-associated incident T2DM is higher in patients with higher BMI, elevated glycated haemoglobin, and impaired fasting glucose[6,18,21]. In a recent American Heart Association scientific statement on statin safety and associated adverse events[18], it was also suggested that the diabetogenic effect of statins was higher in patients with several pre-existing risk factors for diabetes.

To date, it is unclear if the diabetogenic effect of statins is due to HMG-CoA reductase inhibition alone or if it may be relevant to other LDL-C lowering therapies. Additional evidence may come from exploring evidence of any such effect with ezetimibe. Ezetimibe's lipid-modulating mechanism of action differs from statins. It lowers LDL-C by inhibiting the intestinal absorption of dietary and biliary cholesterol. Ezetimibe selectively binds with Niemann-Pick C1 like 1 (NPC1L1) protein, which acts as a cholesterol carrier in the membrane of the enterocyte brush border, and therefore reduces the overall delivery of cholesterol to the liver[22].

Evidence for ezetimibe

The Improved Reduction of Outcomes: Vytorin Efficacy International Trial (IMPROVE-IT) is the only large-scale, double-blind RCT of ezetimibe versus placebo, both added to background simvastatin. The trial included 18,144 patients with acute coronary syndrome[23]. Participants were randomised to combination therapy with ezetimibe (10 mg) added to simvastatin (40 mg) or simvastatin (40 mg) monotherapy. LDL-C levels in the combination arm were 24% lower than in the monotherapy arm[24]. At seven years of follow-up, ezetimibe/simvastatin was superior to simvastatin monotherapy in reducing the risk of major cardiovascular events (HR 0.94, 0.89-0.99, P = 0.016).

The effect of ezetimibe on the risk of T2DM was not defined as an outcome in IMPROVE-IT RCT and was not reported in their publications. However, the American College of Cardiology published an expert opinion about the effect of ezetimibe added to the statin therapy on new-onset T2DM[25]. The analysis of 12,254 participants without T2DM at randomisation found that participants randomised to statin plus ezetimibe did not experience higher risk of T2DM (HR 1.04, 95% CI 0.94, 1.15)[26], than those randomised to statin plus placebo. The trial was not designed to assess the effect of combination therapy on the risk of T2DM; the analysis was most likely underpowered to detect this association. Moreover, the result from one RCT is not sufficient to conclude that ezetimibe is not associated with an increased risk of T2DM. Following a successful phase 3 trial, no further big placebo-controlled trials of ezetimibe can be anticipated. Consequently, genetic studies may be particularly relevant for exploring potential on-target effects.

1.2 Epidemiology of T2DM

T2DM is a complex metabolic disease, characterised by insulin resistance, impaired glucose tolerance, and chronic hyperglycaemia[27-29]. The diagnosis of T2DM is preceded by a period of elevated levels of blood glucose, which is known as pre-diabetes. The cut off point for diagnosing T2DM is HbA1c ≥48 mmol/mol or fasting plasma glucose ≥7.0 mmol/L[30].

Chronic hyperglycaemia caused by T2DM is associated with long-term microvascular complications and damage of various organ systems[31]. Several studies[32,33] demonstrated a substantial increase in cost associated with treating and ongoing management of T2DM and its complications, which require regular biomedical tests, long-term medications, consumables and monitoring devices, primary care consultations and hospital in-patient treatments.

Several modifiable and non-modifiable factors alter the risk of T2DM. Modifiable risk factors include increased BMI, sedentary lifestyle, a low-fibre, high glycaemic index diet, and alcohol consumption[34]. BMI and central fat distribution measured by waist-to-hip ratio (WHR) are known to be strong predictors of T2DM[35]. Non-modifiable risk factors are old age, ethnicity, genetic predisposition to T2DM, and low birth weight[35]. T2DM is heritable, and genetic risk scores (GRSs) have been successfully developed that predict disease.

Currently, T2DM is the ninth leading cause of death globally and is responsible for more than 1 million deaths annually[36]. T2DM increases the risk of cardiovascular events and also leads to the development of microvascular disease (retinopathy, neuropathy, nephropathy) in many people. From 2000 to 2016 the rates of deaths due

to vascular complications of T2DM had increased by 30.8% globally[37]. In addition to the higher risk of death from vascular disease, T2DM was also associated with premature death from other non-communicable[38] and communicable diseases[39]. The prevalence of T2DM is increasing at an alarming rate in all age groups. According to the first Diabetes Atlas[40], published in 2000 by the International Diabetes Federation (IDF), 151 million adults aged 20-79 years old (corresponding to 4.6% of this age bracket) had T2DM. In 2019 the number of people affected by T2DM increased to 463 million (corresponding to 9.3% of this age bracket). It has been estimated that the global prevalence of T2DM will increase to 643 million over the next decade and will reach 784 million by year 2045[40].

In the UK 4.9 million people are living with T2DM and 13.6 million are at high risk of developing the disease[41]. The incidence rate of T2DM is continuing to rise, leading to an increased burden on the National Health Service (NHS). In 2019 the total cost of treating T2DM and its complications reached £14 billion pounds[42], and the annual cost of prescribing treatments through the NHS was £1.07 billion in 2018-2019[43]. If current trends continue, 5.5 million people will have T2DM by 2030, considerably increasing diabetes related clinical and public health expenditure.

Considering the substantial direct cost of T2DM for public health services and indirect social and productivity cost of the disease, more research is needed to understand the effect of LDL-C lowering medications such as statin and ezetimibe on the risk of T2DM in general population and in the subgroup of people with pre-diabetes.

1.3 Genomics in drug discovery and development

GWAS and large international genetic consortia

In the past decade, genome wide association studies (GWAS) have been increasingly used to explore the genetic architecture and aetiology of complex diseases[44], including CVD[45,46] and T2DM[47]. Global collaborative efforts have resulted in large-scale genetic consortia reporting the associations between single nucleotide polymorphisms (SNPs) and modifiable exposures (lipids, biomarkers of glycaemia, and measures of adiposity) and diseases (CVD, T2DM) in hundreds of thousands of participants.

The Global Lipids Genetics Consortium (GLGC)[45] has undertaken large-scale meta-analyses to examine the genetic architecture of lipid levels– including HDL-C, LDL-C, triglycerides, and total cholesterol. The 2013 GLGC meta-analysis (the most recent available at the time of commencing this thesis) included 46 studies and >100,000 participants of mainly European descent (37 studies with primary European population and 8 studies with Asian, African, and Caribbean populations). Each individual GWAS included in the meta-analysis reported SNPs with minor allele frequency (MAF)>1%. The GLGC meta-analysis identified 57 novel loci associated with LDL-C, which together explained 14.6% of the variance of LDL-C[45]. GLGC reported a fixed-effect meta-analysis of the SNP effects on lipid traits[48], and consistent with other genetic consortia, made summary level data available to the research community.

The DIAGRAM (DIAbetes Genetics Replication And Meta-analysis)[47] consortium is another large-scale international GWAS meta-analysis that was designed to improve understanding of the genetic architecture of T2DM. It provides summary level

data for the associations between genetic variants and T2DM, based on 149,821 participants including 34,840 cases of T2DM in mostly European descent population. DIAGRAM identified 63 T2DM susceptibility loci which, when combined in a GRS, could be used as a genetic predictor of lifetime susceptibility to T2DM.

In addition to the insights into genetic architecture and pathophysiology of complex diseases, large-scale genetic consortia provide data on the associations between genetic variants and phenotypes, which could be effectively translated to the drug discovery pipeline[49]. The recent advances in using genetic data for drug discovery applications include genetic drug target validation[50,51], repurposing existing drugs[41,52] and elucidating adverse drug effects[51,53,54].

Mendelian randomisation for predicting adverse drug events

The majority of drug candidates undergoing preclinical testing or drugs at Phase I and Phase II of RCT fail due to safety or efficacy[53]. Although considerable efforts and funds are invested into elucidating the safety profile of a new drug, it remains a major hurdle for drug development. Some rare side effects and adverse events may not emerge during relatively short Phase I and II trials and manifest at the later stages of development, such as Phase III RCTs or during the post-approval period when the number of users of the medication increases significantly. By the time a new drug is approved for clinical use, a limited number of participants will have been treated within RCTs, therefore rare adverse events may not have been apparent. Often, due to a limited number of participants and small number of adverse events, a meta-analysis of several long-term and large-scale RCTs may be needed to identify a safety concern.

Genetic methods and data from large genetic consortia are increasingly being used to inform efficacy and potential adverse events of the medications[50,51,55]. The majority of drugs exert their effects by perturbing the functions or production of proteins[56]. Genetic variants that alter the expression or function of those proteins, which are causally associated with a disease risk factor, can be used as a genetic proxy of the corresponding drugs. Currently, only a limited number of GWAS on the proteome has been published; therefore, the adaptation of this approach is commonly used. Instead of using the association with a protein, most single-protein MR studies use a downstream biomarker approach. Genetic variants associated with a biomarker of interest are selected from a gene (or proximity from this gene) known to encode a protein drug target. Then, the association of these genetic variants with an outcome is estimated through their associations with biomarkers. This method has been referred to as Drug Target Mendelian randomisation (Drug Target MR)[57].

Drug Target MR is often described as being analogous to a RCT[58], which is regarded as the “gold standard” of biomedical investigation[59]. In Drug Target MR, a randomly inherited genotype is used for randomisation that provides a biological proxy for the randomisation in RCTs. Genetic variants, which are associated with the outcome through (i) their effect on the expression of a protein drug-target or (ii) their effects on a downstream biomarker act as an instrumental variable. Individuals carrying trait-decreasing alleles, which act as a proxy of a drug, can be considered to be randomised to a treatment group, whereas carriers of the other (trait-increasing) alleles can be considered to be randomised to a control group. Participants with a genetic variant experience a lifelong effect, which mimics exposure to the drug in question and therefore can help to predict whether the proposed pharmacological interventions will produce expected health benefits and if they are associated with other health

consequences. However, unlike the pharmacological intervention which tends to be relatively short in duration, genetic variants represent a cumulative lifelong exposure to protein perturbation. Therefore, the magnitude of the effects of the genetic variants estimated with MR studies for a given reduction in the biomarker could be different to what is observed in clinical research[60]. **Figure 1:3** presents a comparison between an RCT and a Drug Target MR.

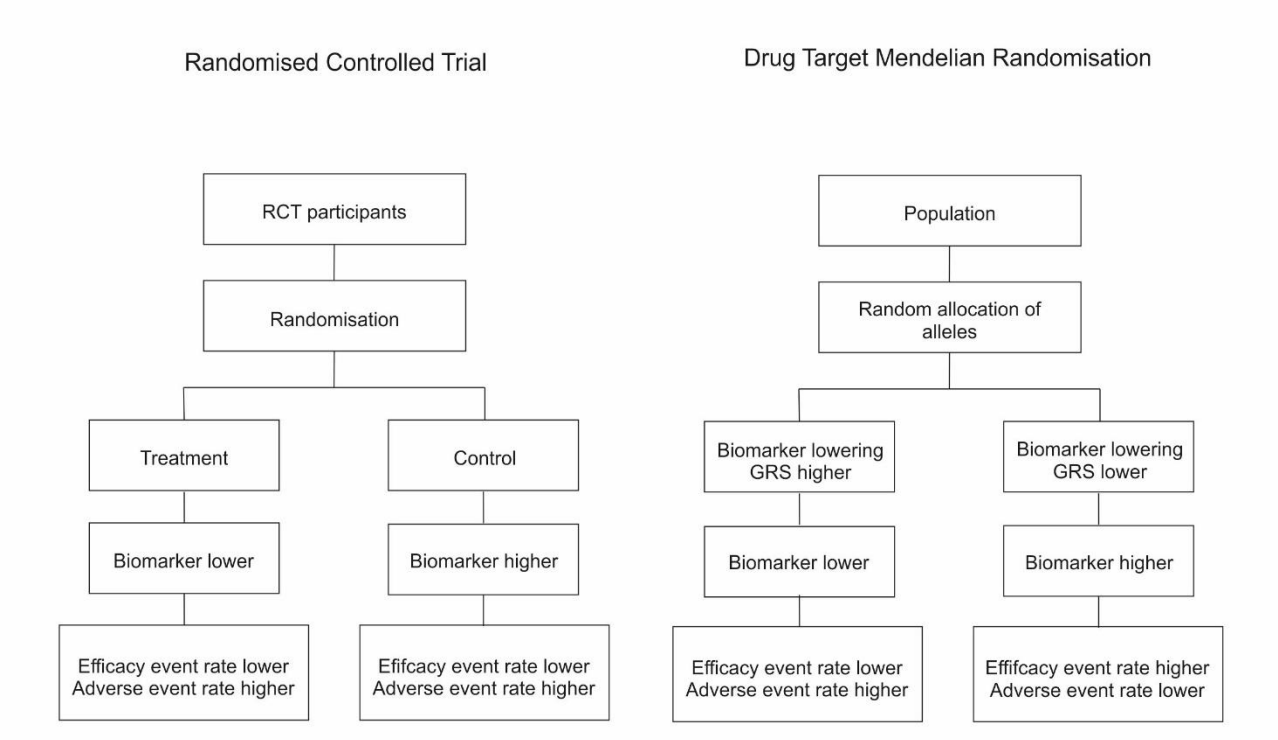


Figure 1:3 | The comparison between an RCT and the Mendelian randomisation approach

Drug Target MR studies are increasingly used in cardiovascular research[61]. For example, after large-scale RCTs[62-65] established the efficacy of statins for prevention of CVD, several Drug Target MR studies showed similar results[30,66,67] Genetic studies were also able to show the effect of statins on new-onset diabetes by examining the effect of naturally occurring variation in *HMGCR* on the risk of T2DM[67].

Ezetimibe has been less intensively studied than statins, with only one large-scale, double-blind, RCT conducted to date, namely IMPROVE-IT[68]. The *NPC1L1* gene encodes Niemann–Pick C1-Like 1, which is the molecular target of ezetimibe. Drug Target MR studies[66] using genetic variants in *NPC1L1* showed very similar associations to *HMGCR* in the risk of CVD per one SD unit lower *NPC1L1*-predicted LDL-C. Similarly, a Drug Target MR study[30] provided genetic evidence supporting an effect of NPC1L1 inhibition on the increased risk of T2DM, however, the magnitude of the genetic effect of NPC1L1 inhibition was much larger than was reported for *HMGCR*. The potential diabetogenic effect of ezetimibe became the subject to further controversy due to the report[25,69] from IMPROVE-IT RCT stating that ezetimibe added to statin therapy does not increase risk of T2DM.

The rationale and limitations of Drug Target Mendelian randomisation

The underlying rationale of Drug Target MR is that if the biomarker is causal risk factor for the disease, then the genetic variants, located in the coding protein for this biomarker should be also associated with the disease[61].

An instrumental variable constructed from these genetic variants should satisfy several assumptions[70], including (i) it must be associated with the biomarker; (ii) it must not be associated with the outcome directly, an instrumental variable should exert the effect only through its association with the biomarker; (iii) the instrumental variable must not be associated with any confounder of the association between exposure and outcome.

The use of genetic variants or GRS as an instrumental variable in a Drug Target MR study gives this method significant methodological advantages over traditional

observational studies. Genetic variants are randomly assigned at conception, rendering them independent from other biological or environmental factors. Therefore, MR studies reduce the chance of biases due to confounding[70]. Additionally, MR studies exclude the risk of reverse causation[71].

The choice of the second LDL-C lowering targets Two main factors were considered when deciding between the two genetic drug targets, *NPC1L1* and *PSCK9*. Firstly, we assessed the differences in biological mechanisms of action between proposed targets and statins. Secondly, we compared prescription trends for two medications. While *PCSK9* inhibitors lower LDL-C through a similar to statins' biological mechanism of action, *NPC1L1* acts via a contrasting biological mechanism.

Moreover, ezetimibe is a commonly prescribed lipid-lowering medication, whereas NICE had not approved the use of *PSCK9* inhibitors when the research project was planned. Therefore, *NPC1L1,* a genetic proxy for ezetimibe, was chosen as a second drug target.

Limitations of the Drug Target MR methodology include:

- limited data on the genetic associations with proteins levels; therefore, a need to use proxy associations with biomarkers;
- the effect of the population structure (e.g. if there is a variability in allele frequencies, disease and exposure rates in different ethnic subgroups of the population);

- pleiotropy (the situation where a polymorphism used as an instrumental variable is affecting more than one biological pathway and influencing several intermediate phenotypes);
- linkage disequilibrium (statistical correlation between polymorphisms, due to non-random association of alleles at different loci of different populations);
- inadequate statistical power and weak instrument (as individual polymorphisms explain a small proportion of the variance in the exposure, a large sample size is needed to detect the association between SNP and outcome).

1.4 The need for further research

Value of further research

The growing and aging population increasingly needs effective management of LDL-C levels to lower mortality and morbidity from CVD. Statins remain the most prescribed and affordable lipid-modulating agent with the rates of initiation exceeding any other medication. Although the cardiovascular benefits of statin therapy hugely outweigh the risks of developing T2DM, statins' effect on glucose homeostasis is a growing concern for public health due to potentially increasing economic and social costs associated with T2DM and its complications. A better understanding of the relationship between statins and risk of T2DM and whether there are groups of patients at particular risk of statin induced T2DM may help design new therapeutic approaches and cost-effective preventive measures to reduce the number of incident T2DM cases within those who require therapeutic lipid modification.

A Drug Target MR study can provide additional genetic evidence of the effect of statins' genetic proxy on not only the risk of T2DM and glycated haemoglobin levels, but also diabetes-related phenotypes, such as anthropometric measures of central and general adiposity (key drivers of diabetes risk).

Individual RCTs of statins lacked sufficient power to robustly examine the modest diabetogenic effect of statins in subgroups of patients stratified by their baseline risk of T2DM. Nonetheless, it has been suggested that statins' effect on T2DM is limited to pre-diabetic patients only (with those who have normal glycaemic control experiencing little or no extra risk of incident T2DM when commencing statin therapy)[8,18]. In this thesis the effect of a *HMGCR* GRS on risk of T2DM, glycaemic

traits and anthropometric measures will be examined in participants with different susceptibilities to T2DM. It will also assess the effect on T2DM and related traits in men and women, as fat distribution significantly differs between sexes. If statin's effect on T2DM is mediated by increased adiposity, the effect may be different in men, who have predominantly visceral fat deposition compared to female subcutaneous fat deposition.

Moreover, another lipid-modifying drug ezetimibe reduces the concentration of LDL-C via a different biological mechanism of action to statin therapy. Its effect on T2DM may therefore differ from the effect of statins. The current study will also use MR genetic tools to elucidate the effect of the genetic proxy of ezetimibe on T2DM, glycaemic traits, and adiposity measures.

Finally, there are very limited data on the effect of ezetimibe on T2DM. A previous MR study[30] showed an increased risk of T2DM, whereas the only large-scale RCT of ezetimibe found no association with T2DM[69]. The current study will also use MR genetic tools to elucidate the effect of the genetic proxy of ezetimibe on T2DM, glycaemic traits, and adiposity measures.

The comparison of the effects of *HMGCR* and *NPC1L1*, two genes encoding drug targets with different biological mechanisms of action, on the risk of T2DM will contribute to the ongoing investigation on whether an increased risk of T2DM is associated with a particular biological pathway.

Very little work has been undertaken to elucidate the genetically predicted effects of LDL-C lowering therapeutic targets on the risk of T2DM, glycaemia and anthropometric traits. Several studies[30,66,67] mainly used overlapping data (prior to UK Biobank releasing individual-level genetic data for its entire cohort). UK Biobank data

allows an analysis of the population not included in previous genetic studies. In addition, UK Biobank data provide an opportunity to run a better-guided subgroup analysis in participants with different clinical and genetic predispositions to T2DM, which has not been done before. Furthermore, while the effects of statins on the risk of T2DM were reported by the meta-analysis of large RCTs of statins[67], there was no current evidence in the literature of the effect of ezetimibe. This research project aims to address the current literature's shortcomings and contribute to knowledge of the relationship between genetically instrumented lower LDL-C and the risk of T2DM.

Opportunities in the UK Biobank

UK Biobank is an ongoing, population-based, cohort study, which holds genetic and phenotypic data and detailed medical history for 0.5M participants[72] The scope, quality, and diversity of data, such as genome-wide genotype information, diverse phenotypic and medical data, including anthropometric measures and blood biomarkers of glycaemia for each participant, provided a unique opportunity for a MR study. The T2DM outcome, which caused inconsistency in the previous MR studies and RCTs, will be well defined in the UK Biobank due to the links with hospital episode statistics, data on medication use, and self-reported medical history. Moreover, diabetes-related measurements such as measures of general and central adiposity and blood biomarkers of glycaemia were collected following the same protocol for each participant, which will ensure consistency not possible to achieve in previous studies.

1.5 Objectives of the thesis

The objectives of this thesis are to use the UK Biobank and genetic data from publicly available consortia:

(1) to investigate the association between genetic variation in the *HMGCR* gene and the risk of T2DM in men and women; to examine the effect of *HMGCR* GRS on glycaemic traits and measures of central and general adiposity in men and women.

(2) to investigate, whether the effect of *HMGCR* on the risk of T2DM differs between subgroups of participants at different risks of T2DM; to examine whether the effects of *HMGCR* on blood biomarkers of glycaemic control and anthropometric measures differ between subgroups of participants at different risks of T2DM.

(3) to investigate the effect of *NPC1L1* GRS on total T2DM; to examine the effect of *NPC1L1* GRS on glycaemic traits and adiposity in men and women.

Chapter 2 Systematic literature review

2.1 Objectives of the systematic review

Statins are commonly prescribed drugs for the prevention of cardiovascular events in high-risk patients. Years of use in clinical practice and numerous clinical trials have proven their therapeutic efficacy and safety[8]. While substantially decreasing the risk of developing CVD, recent meta-analyses[19] linked statin therapy with an increased risk of developing T2DM. As T2DM was not a pre-specified outcome in most of the RCTs of statins, large-scale meta-analyses were needed to demonstrate a causal association of statin therapy with incident T2DM. Although previous studies demonstrated that the risk of developing T2DM due to statin therapy is relatively low and is outweighed by the health benefits[8,66,67], it is important to understand if this increased risk of T2DM is an off-target effect of statins or part of the biological pathway of 3-Hydroxy-3-methylglutaryl-coenzyme-A reductase inhibition. It is also important to understand if the effects vary in different patient subgroups and discover the likely mechanisms mediating the increased risk of T2DM.

Since statins' effect on T2DM has been identified, other lipid-modulating drugs' effects on glucose homeostasis have also come under scrutiny. Ezetimibe is another lipid-modulating drug, which reduces the level of LDL-C by limiting exogenous and biliary cholesterol absorption in the small intestine through inhibiting the Niemann-Pick C1-Like1 protein[22]. Little is known about ezetimibe's effect on glucose homeostasis and on adiposity traits.

This review summarised the available published evidence of the association between genetic proxies of both therapies (*HMGCR* for statins and *NPC1L1* for ezetimibe) and T2DM, adiposity measures and glycaemic traits. It also reviewed the evidence from large meta-analyses of RCTs of statins and results of the major RCT of ezetimibe to elucidate the effects of these therapies on the risks of T2DM, glycaemic traits, and adiposity measures.

This chapter presents four systematic searches to identify RCTs and genetic studies that have addressed the following research questions:

Search 1: Is there an association between genetic variants in *HMGCR* and the risk of T2DM, glycaemia, and adiposity?

Search 2: Does statin therapy affect the risk of T2DM, glycaemia, and adiposity?

Search 3: Is there an association between genetic variants in *NPC1L1* and the risk of T2DM, glycaemia, and adiposity?

Search 4: Does ezetimibe therapy affect the risk of T2DM, glycaemia, and adiposity?

2.2 Methods

Ovid Medline, Ovid Embase and The Cochrane Library were used for systematic searches. Additionally, data on genetic associations between polymorphisms in *HMGCR* and *NPC1L1* and LDL-C, T2DM, HbA1c and adiposity measures were extracted from GLGC[45], DIAGRAM[47], MAGIC[73], and GIANT[74] international consortia.

For the genetic parts of the review, the following search strategy was used (simplified):

(HMGCR AND GLYCAEMIC TRAITS) OR (HMGCR AND ADIPOSITY TRAITS) OR (HMGCR AND DIABETES) OR (HMGCR AND MENDELIAN RANDOMISATION)

(NPC1L1 AND GLYCAEMIC TRAITS) OR (NPC1L1 AND ADIPOSITY TRAITS) OR (NPC1L1 AND DIABETES) OR (NPC1L1 AND MENDELIAN RANDOMISATION)

The full description of each searching term is shown in **Table 2:1**.

Table 2:1 | Systematic review search terms for genetic studies

Short term	Full term	Comments
HMGCR	(*HMGCR* and (gene* or score or risk or association or effect or SNP or single nucleotide polymorphism or rsid or loc*)).mp.	Any paper which contains *HMGCR* and one of the terms in brackets in the default search fields. Truncated terms gene* and loc* include all terms with these roots and various possible endings, therefore this will retrieve “genetic”, “genetics”, “gene”, “locus”, “loci” etc.
NPC1L1	(*NPC1L1* and (gene* or score or risk or association or effect or SNP or single nucleotide polymorphism or rsid or loc*)).mp.	Any paper which contains *NPC1L1* and one of the terms in brackets in the default search fields. Truncated terms gene* and loc* include all terms with these roots and various possible endings, therefore this will retrieve “genetic”, “genetics”, “gene”, “locus”, “loci” etc.

GLYCAEMIC TRAITS	(glycaemic traits or glucose or HbA1c or glycated h#emoglobin or h#emoglobin A1c or insulin).mp.	Any paper which contains one of these terms in the default fields. Special command for “h#emoglobin” and “h#emoglobin A1c” retrieves both British and American spelling of the words.
ADIPOSITY TRAITS	(BMI or body mass index or obesity or adiposity or waist circumference or waist hip ratio or bodyweight or weight).mp.	Any paper which contains one of these terms in the default fields.
DIABETES	diabetes.mp. or Diabetes Mellitus, Type 2/	Any paper which contains “diabetes” as a keyword in the default fields or “Diabetes Mellitus, Type 2” as a subject heading.
MENDELIAN RANDOMI-SATION	mendelian randomi#ation.mp.	Any paper which contains one of these terms in the default fields, including British and American spelling of the term “randomisation”.

The searches for RCTs of statins and ezetimibe were different due to the substantial difference in the number of RCTs conducted for both therapies. For statins, the following search strategy was used:

STATINS AND (GLYCAEMIC TRAITS OR DIABETES) AND META-ANALYSIS

For ezetimibe, at first the search strategy was the same as for statins, but there were no meta-analyses available, therefore the search strategy was adjusted and included RCTs rather than meta-analyses of RCTs:

EZETIMIBE AND (GLYCAEMIC TRAITS OR DIABETES) AND RCT

The full description of each searching term is shown in Table 2:2.

Table 2:2 | Systematic review search terms for meta-analyses and RCTs

Short term	Full term	Comments
STATINS	exp Hydroxymethylglutaryl-CoA Reductase Inhibitors/ or statin*.tw. or atorvastatin.tw. or cerivastatin.tw. or fluvastatin.tw. or lovastatin.tw. or pravastatin.tw. or simvastatin.tw.	Any paper which contains exploded term *HMGCR* as a subject heading. “Explod” or exp command is used to include the selected term and all more specific terms, related to the search term, or statins. The search is also looking for the various statins trade names in the text word (tw.) fields – title and abstract.
EZETIMIBE	exp ezetimibe.	Any paper which contains exploded term ezetimibe as a subject heading.
GLYCAEMIC TRAITS	(glycaemic traits or glucose or HbA1c or glycated h#emoglobin or h#emoglobin A1c or insulin).mp.	Any paper which contains one of these terms in the default fields. Special command for “h#emoglobin” and “h#emoglobin A1c” retrieves both British and American spelling of the words.
DIABETES	diabetes.mp. or Diabetes Mellitus, Type 2/	Any paper which contains “diabetes” as a keyword in the default fields or “Diabetes Mellitus, Type 2” as a subject heading.
RCT	randomi#ed controlled trial.mp.	Any paper which contains one of these terms in the default fields, including British and American spelling of the term “randomised”.
META ANALYSIS	meta-analysis.mp.	Any paper which contains “meta-analysis” as a keyword in any of the default fields.

Criteria for considering studies for this review:

Genetic studies: peer reviewed English language studies reporting the association between genetic variants in *HMGCR* and/or *NPC1L1* and T2DM or glycaemic/adiposity traits. Data on genetic associations from the DIAGRAM, MAGIC, and GIANT consortiums.

RCT meta-analyses: peer reviewed English language meta-analyses of RCTs of clinical trials of statin therapy reporting study-specific estimates of the effects of statin therapy on glycaemic/adiposity traits and/or the risk of incident T2DM. Studies in the meta-analysis should use either a drug versus placebo (or standard care) comparison of different statins, or different doses/intensities of statins.

RCT of ezetimibe: trials reporting the effect of ezetimibe on T2DM, or glycaemic/adiposity traits. The selection of trials was restricted to those with at least 1000 participants.

Exclusion criteria: Studies which were published before 2007 were not included. All papers with non-human subjects, conference papers and papers looking at the non-lipid modifying variants in *HMGCR* or *NPC1L1*, as well as meta-analyses which included RCT with less than 1000 participants, were also excluded from the review.

Electronic searches

The searches were conducted in October 2017. For each search, all papers were checked for duplicates and then a systematic screening of titles and abstracts for eligibility was performed. After exclusion of irrelevant papers, the final eligibility decision was made by assessing the full texts.

From the genetic papers (Search 1 and Search 3), the following information was extracted: authors, year of publication, number of participants and their ethnicity, point estimate and confidence intervals for the association between a genetic variant and an adiposity/glycaemic trait, odds ratio for T2DM. The results were presented per effect allele (the allele which is associated with lower level of LDL-C) for individual SNPs or per 1 SD change in LDL-C for genetic risk scores.

From meta-analyses of RCTs of statins (Search 2) and RCTs of ezetimibe (Search 4), the following information was extracted: authors, year of publication, number and names of RCTs, sample size of the meta-analysis, the number of events, the effect on T2DM (Odds Ratio (OR), Hazard Ratio (HR), or Relative Risk (RR)).

2.3 Results

2.3.1 Search 1: *HMGCR* genetic studies

The initial search retrieved 100 genetic papers. After the title screening was conducted, 41 papers were selected for the full text review. 7 papers were selected for the systematic review. Detailed results of the search are presented in a PRISMA flow diagram bellow (**Figure 2:1**).

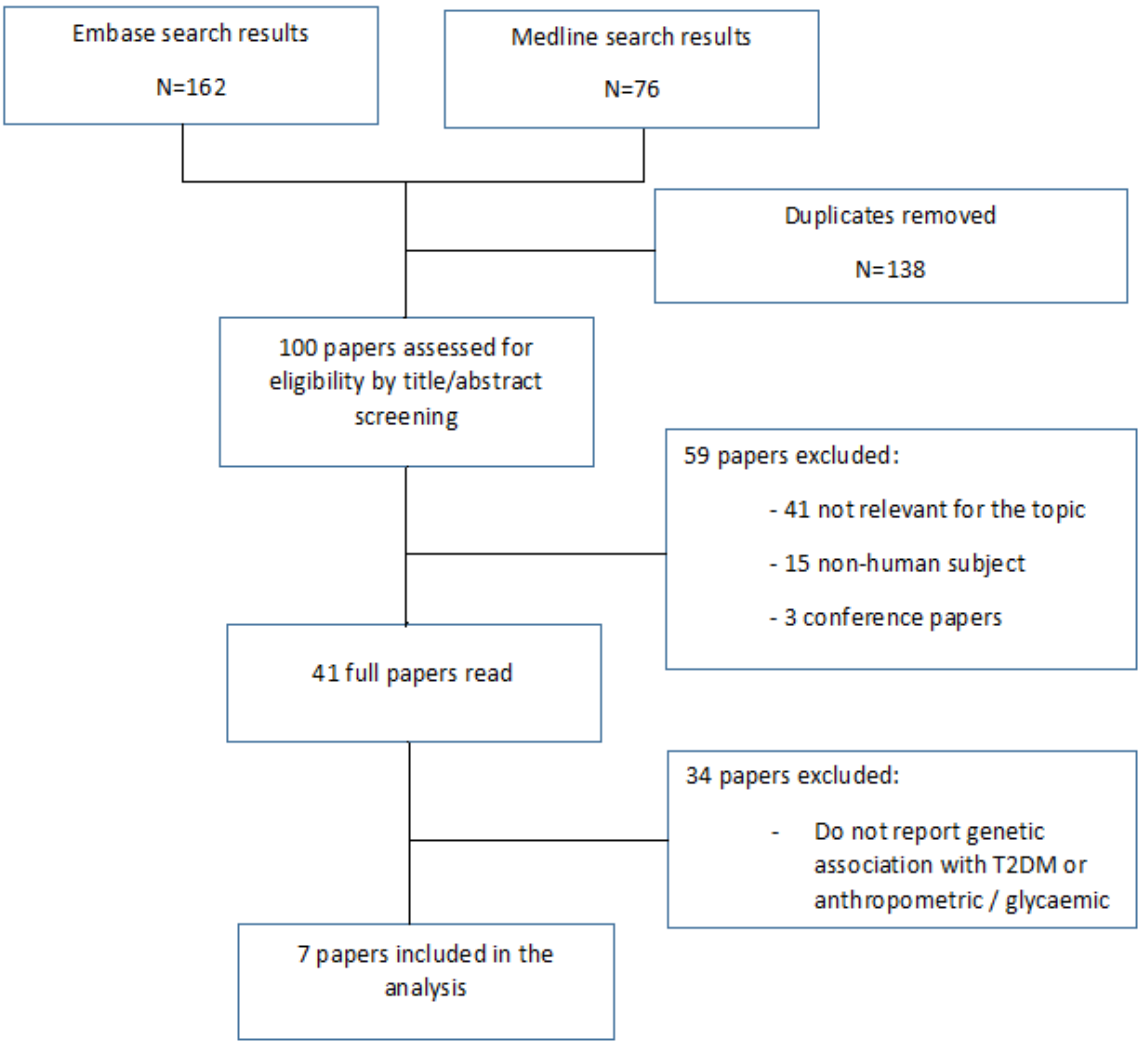


Figure 2:1 | PRISMA flow diagram for the Search 1 (*HMGCR* genetic studies) Associations with T2DM

Individual SNPs

Several genetic studies examined the effects of LDL-C lowering polymorphisms in and in close proximity to the *HMGCR* gene on the risk of T2DM. Swerdlow *et al.*[67] meta-analysed 43 genetic studies with 223,463 participants of European ancestry. This meta-analysis used two genetic variants in the *HMGCR* gene as a proxy for statin therapy and examined their effects on the risk of T2DM and adiposity traits.

For the primary analysis of the effects of rs17238484, data were available from 43 studies with 195,444 participants. For the secondary analysis of the effects of rs12916, data were available from 21 studies with 94,652 participants. LDL-C lowering alleles in both SNPs were associated with higher risks of T2DM; rs17238484 effect on T2DM was 1.02 (95% CI 1.00, 1.05) and rs12916 effect on T2DM was 1.06 (95% CI 1.03, 1.09) per LDL-C lowering allele (Appendix 2.1).

The associations of rs17238484 and rs12916 with the risk of T2DM were also reported in a subsequent study by Ference *et al.*[51], which used genetic data from 181,111 participants (including from the DIAGRAM consortium) and 10,635 cases of T2DM. Ference et *al.*[51] reported positive associations of LDL-C lowering alleles in these SNPs with the risk of T2DM, and no associations with T2DM for three other *HMGCR* SNPs (rs10066707, rs2303152, rs5909) included in the analysis (Appendix 2.1).

The most recent paper by Lotta *et al.*[30] looked at the associations between LDL-C lowering alleles in several genes (Appendix 2.2) as proxies of established cardiovascular medicines and T2DM. The paper reported associations of three *HMGCR* polymorphisms with the risk of T2DM. There were 320,946 participants including 55,241 cases of T2DM available for rs12916, and 270,269 participants including 50,773 cases of T2DM for rs16872526 and rs5744707. Only rs12916 had a

statistically significant association with T2DM (OR 1.03, 95% CI 1.01, 1.05); rs16872526 was directionally concordant but its association with T2DM did not reach the 0.05 level of statistical significance. The third SNP rs5744707 showed a directionally discordant non-significant effect on T2DM (Appendix 2.1).

Another genetic association analysis[75] of 6,200 European participants examined the effects of four lipid-lowering Drug Target genes on common disease, including T2DM (327 cases). In this weakly powered study, one SNP in *HMGCR* (rs3846663), selected based on its association with LDL-C, showed no association with T2DM (OR 1.0, 95% CI 0.8, 1.2) per LDL-C lowering allele.

HMGCR GRS

Two studies[30,51] combined *HMGCR* genetic variants in a *HMGCR* GRS. Ference *et al.*[51] combined 5 SNPs (rs17238484, rs12916, rs10066707, rs2303152, rs5909) within 100 kb of the *HMGCR* gene to construct a genetic instrument explaining more variation in LDL-C. The effects of each SNP on LDL-C were estimated in the GLGC consortium, and the effects on T2DM were estimated in up to 181,111 participants, including 10,635 cases of T2DM. The study reported 1 SD lower *HMGCR* genetically predicted LDL-C was associated with OR 1.59 (95% CI 1.25, 2.02) higher risk of T2DM.

Lotta *et al.*[30] combined 3 SNPs in the *HMGCR* gene (rs12916, rs5744707, rs16872526) to assess the effect on T2DM. The individual SNPs' associations with LDL-C were assessed by the GLGC consortium, and the associations with T2DM were estimated in a meta-analysis of five genetic studies (InterAct, DIAGRAM, UK Biobank, T2D-GENES, GoT2D) with 50,775 cases of T2DM and 270,269 controls. The study

reported that a 1 SD lower *HMGCR* genetically predicted LDL-C was associated with 39% (95% CI 1.12, 1.73) higher risk of T2DM.

Associations with glycaemic traits

Five studies reported associations of genetic variants in *HMGCR* with glycaemic traits. Lotta *et al.*[30] assessed the effects of five SNPs and a *HMGCR* GRS with fasting plasma glucose and fasting plasma insulin. For this analysis, Lotta *et al.* [30] used the associations between the SNPs and glycaemic outcomes from the MAGIC consortium, which reported the association with fasting insulin for 133,010 participants, and with fasting glucose for 108,552 participants. Only one SNP rs12916 was associated with higher fasting insulin (0.01 SD 95% CI 0.00, 0.02) per LDL-C lowering allele. The effects of other SNPs on both traits were quantitatively consistent and directionally concordant but did not reach the 0.05 level of statistical significance. When combined in an *HMGCR* GRS, a 1 SD lower genetically predicted LDL-C was associated with 0.16 SD (95% CI 0.05, 0.27) higher level of fasting insulin and had no significant effect on fasting glucose (0.02 SD 95% CI -0.07, 0.11).

Swerdlow *et al*.[67] estimated the effect of two SNPs on plasma glucose (23 studies with 73,490 participants) and plasma insulin (12 studies with 37,353 participants). Each LDL-C lowering allele in rs17238484 was associated with 1.62% (95% CI 0.53, 2.72) higher plasma insulin and 0.23% (95% CI 0.02, 0.44) higher plasma glucose. Associations of rs12916 with plasma glucose and plasma insulin were directionally concordant with rs17238484, but not statistically significant (Appendix 2.1).

Wurtz *et al.*[76] has also reported the positive association between rs12916 and glucose in a study comparing the effects of statin therapy with the effects of the *HMGCR*

genotype. Genetic data were available for 27,914 participants from 8 population-based cohorts. Each LDL-C lowering allele in rs12916 was associated with 0.01 SD (95% CI 0.00, 0.02, P =0.0006) higher glucose.

Tragante *et al.*[77] examined 5 SNPs within *HMGCR* and reported a positive association with a study specific "glycaemic burden composite", derived from a meta-analysis of associations of four glycaemic traits (T2DM, fasting insulin, fasting glucose, fasting proinsulin and HbA1c) with p <0.05; however individual associations of the genetic variants with each trait were not provided.

Associations with anthropometric measures

Two studies[30,67] reported genetic associations between *HMGCR* and anthropometric measures. The first meta-analysis, which systematically assessed the effect of genetic variants in *HMGCR* on adiposity, was conducted by Swerdlow *et al.*[67]. Both *HMGCR* SNPs in that study were associated with increased body weight, BMI, waist circumference and hip circumference. In the population of almost 143,113 participants from 30 studies, each LDL-C lowering allele in rs17238484 was associated with 0.30 kg (95% CI 0.18, 0.43) higher bodyweight, and 0.11 kg/m^2 (95% CI 0.07, 0.14) higher BMI. In the population of 69,163 participants from 19 studies, it was associated with 0.32 cm (95% CI 0.16, 0.47) greater waist circumference, and 0.21 cm (95% CI 0.10, 0.32) greater hip circumference. In a population of 95,469 participants, it was associated with higher waist-to-hip ratio (0.001, 95% CI 0.000, 0.002). The second *HMGCR* SNP (rs12916), which was used for a subsidiary analysis, had directionally concordant and quantitatively consistent associations with these anthropometric measures (Appendix 2.1).

The second study by Lotta *et al.*[30] used data from 222,495 participants from the DIAGRAM 2016 consortium to assess the associations between three *HMGCR* SNPs and an *HMGCR* GRS with BMI, and data from 222,047 participants to assess the effect on waist-to-hip ratio.

Each LDL-C lowering allele in rs12916 was associated with higher BMI (0.02 SD, 95% CI 0.01, 0.02) and with higher waist-to-hip ratio (0.01 SD, 95% CI 0.00, 0.01). Rs16872526 was associated with higher BMI (0.02 SD, 95% CI 0.01, 0.03), but not with higher waist-to-hip ratio. The third SNP, rs5744707, was not associated with these traits. When combined in a *HMGCR* GRS, the associations of genetic variants with BMI and waist-to-hip ratio were 0.23 SD (95% CI 0.15, 0.30) and 0.09 SD (95% CI 0.01, 0.18) per 1 SD lower genetically predicted LDL-C respectively.

Overall, these studies reported eleven SNPs within *HMGCR* associated with T2DM, glycaemic traits or adiposity measures. The effects of the individual SNPs within *HMGCR* and *HMGCR* GRS on T2DM, glycaemic traits and adiposity measured are summarised in Appendix 2.1.

The most studied SNP, rs12916, was analysed in 6 papers and its associations with T2DM / glycaemic and adiposity traits were reported in the three major consortia. Four studies reported negative associations of rs5909 with the risk of T2DM, two studies reported the association of rs3846663 with T2DM and rs17238484 with T2DM or glycaemic / adiposity traits. Other SNPs were only represented in a single analysis. **Table 2:3** shows frequencies of each SNP analysis across the selected manuscripts.

Table 2:3 | Frequencies of each *HMGCR* SNP analysis across the selected manuscripts

SNP	Lotta *et al.* (2016)	Swerdlow *et al.* (2015)	Swerdlow *et al.* (2016)	Ference *et al.* (2016)	Tragante *et al.* (2016)	Knouff *et al.* (2008)	Wurtz *et al.* (2016)	N of papers reported
rs12916	x	x	x	x	x		X	6
rs5909				x				1
rs17238484		x	x	x				2
rs3846663					x	x		2
rs10066707				x				1
rs3846662					x			1
rs2303152				x				1
rs7703051					x			1
rs12654264					x			1
rs16872526	x							1
rs5744707	x							1

Three major genetic consortiums (DIAGRAM, GIANT and MAGIC) provided publicly available GWAS summary statistics, including the associations of rs12916 and rs5909 with T2DM, adiposity and glycaemic traits. The full results of the associations between genetic variants in *HMGCR* with T2DM, adiposity measures and glycaemic traits are shown in the Appendix 2.2.

2.3.2 Search 2: Meta-analyses of RCTs of statins

Statins and incident T2DM

The initial literature search retrieved 10,473 meta-analysis papers. After title assessment and removal of duplicate papers, 1,001 papers were retained for abstract screening, and 108 papers were selected for full text assessment. Of these, 8 papers

were eventually included. Detailed results of the search are presented in a PRISMA flow diagram (**Figure 2:2)**.

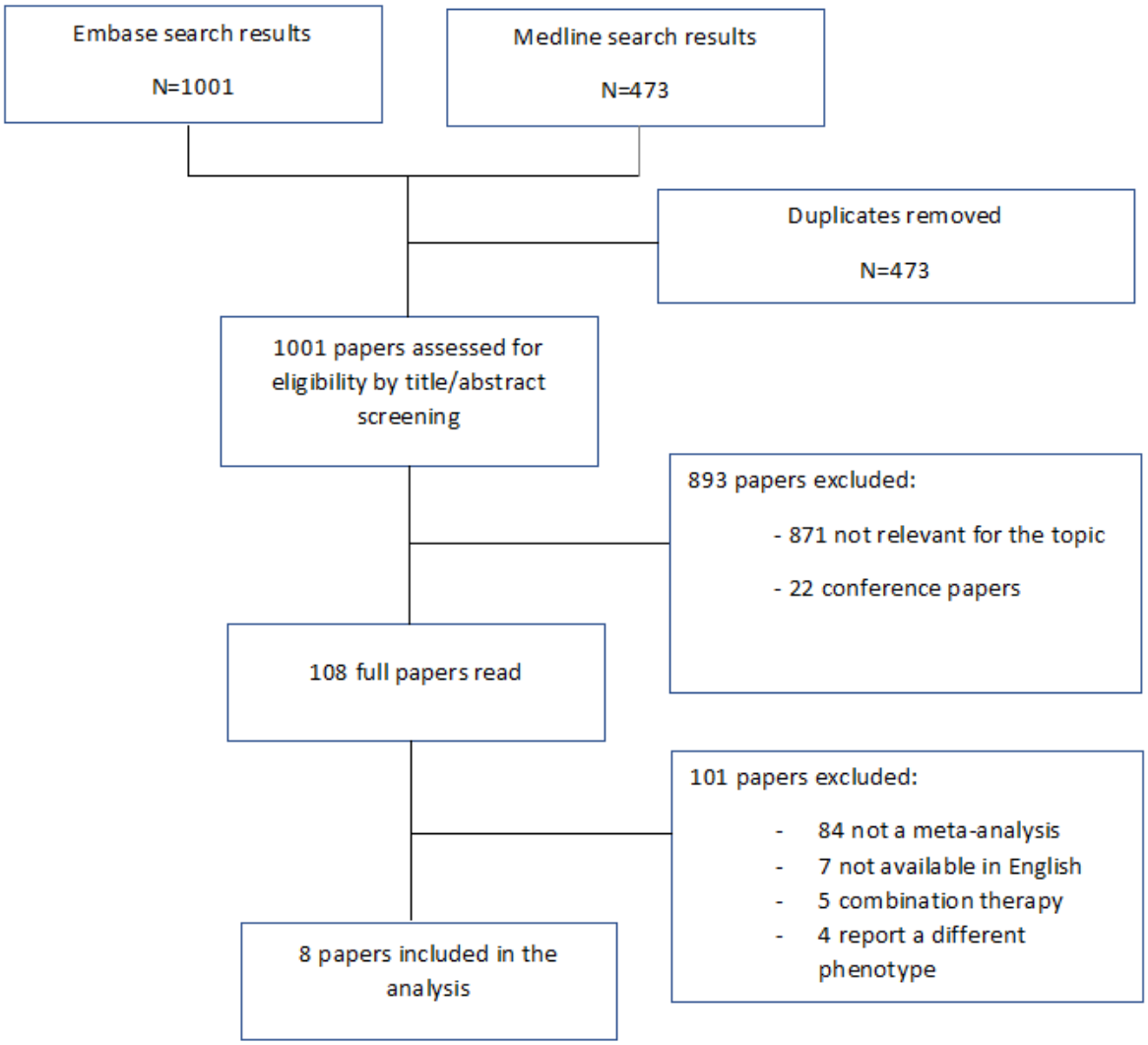


Figure 2:2 | PRISMA flow diagram for search 2 (meta-analyses of RCTs of statins)
The RCTs of statins included in these meta-analyses are described in Appendix 2.5.

.

One of the first meta-analyses of the association between statin therapy and T2DM was published in 2009 by Coleman et al[78]. Based on the results of one of the early cardiovascular outcome trials, the West of Scotland Coronary Prevention Study (WOSCOPS)[9] which reported a substantial 30% decrease in risk of incident T2DM in the pravastatin treatment arm compared to placebo, the objective of the meta-analysis

was to examine the effect of statin therapy on T2DM. Five randomised controlled trials of statins, including WOSCOPS, with a total population of 39,791 patients, were included in this meta-analysis. 1,407 participants who were free from T2DM at baseline developed the disease. The authors concluded that statins, as a class, were not associated with the risk of developing T2DM (RR 1.03 95% CI 0.89, 1.19). When WOSCOPS results were excluded, the meta-analysis showed a significant 14% (95% CI 1.02, 1.28) increase in the risk of incident T2DM. However, the authors could not explain the difference in the effect of statins on T2DM between WOSCOPS and other RCTs.

Published the same year, another meta-analysis[79] also explored the effect of statin therapy on incident T2DM. 6 RCTs with 57,593 participants were included in the analysis. In addition to the RCTs from Coleman's *et al*.,[78] this meta-analysis included JUPITER[63], the first RCT, that reported a statistically significant effect of statins on the risk of T2DM. With the WOSCOPS trial included, there was no association with incident T2DM (RR 1.06, 95% CI 0.93, 1.23). In a sensitivity analysis excluding WOSCOPS, there was a small increase in risk of incident T2DM in the treatment arm (RR 1.13, 95% CI 1.03, 1.23).

A year later, a larger meta-analysis by Sattar *et al*.[19] demonstrated that statin therapy was associated with an increased risk of incident T2DM. The study analysed published and unpublished evidence from 13 placebo-controlled and standard-care controlled cardiovascular outcome statin trials with 91,140 participants (45,619 on placebo, 45,521 on statin). The results showed a 9% (95% CI 1.02, 1.17) increase in risk of incident T2DM associated with statin therapy compared to the control group.

Preiss *et al.*[20] then investigated if the risk of T2DM varied by intensity of statin therapy. This meta-analysis included 32,752 (16,480 on high intensity statin, 16,344 on moderate intensity statin) non-diabetic participants from five cardiovascular outcome RCTs. All trials used atorvastatin (80 mg) or simvastatin (80 mg) for an intensive regimen and smaller doses of the same drug (20-40 mg) or pravastatin 40 mg for a moderate-dose regimen. The authors reported a dose-dependent association of statin therapy with risk of diabetes. Patients allocated to the high intensity statin therapy experienced a 12% (OR 1.12, 95% CI 1.04, 1.22) higher risk of incident diabetes than those allocated to moderate-dose therapy.

The same year Cai *et al.* [80] published the results of a meta-analysis of RCTs of statins. The meta-analysis included 95,102 non-diabetic participants from 14 trials. Unlike the previous study, which stratified participants by the statin dose, this meta-analysis combined RCTs of statins aiming to reduce the LDL-C to a specific target level. The high target group for LDL-C was set at ≥ 2.59 mmol/L, the moderate target group was set at 1.8–2.59 mmol/L, and the low target group was set at <1.8 mmol/L. Like the previous meta-analyses, Cai *et al.*[80] reported a target-dependent increase in risk of T2DM. Patients in the low target group had a 33% (95% CI 1.14, 1.56) increased risk of incident T2DM, those in the moderate target group had a 16% (95% CI 1.06, 1.28) increase and participants in the high target group had no extra risk of incident T2DM (OR 1.01, 95% CI 0.92, 1.10).

In 2015, Swerdlow *et al.*[67] combined data on 96,418 participants free from diabetes at baseline from 15 statin placebo-controlled and standard care-controlled RCTs. The mean LDL-C reduction at one-year was 1.07 mmol/L (95% CI 0.44, 1.70). The risk of new-onset T2DM was 11% higher (95% CI 1.03, 1.20) among those allocated to statin therapy than among those allocated to placebo or standard care.

Statins and glycaemic traits

To better understand possible underlying biological mechanisms of the effect of statin therapy on T2DM, several meta-analyses have examined the associations between statins and glycaemic traits. Baker *et al*[81] reported inconsistent effects of different statins on insulin sensitivity in 1,146 non-diabetic patients from 16 studies. The authors concluded that statins, as a class, do not exert a worsening effect on insulin sensitivity.

Erqou *et al.*[82] examined the effect of statins on glycaemia in patients with diabetes. 9,696 participants from 9 placebo-controlled or standard treatment were included in the analysis. By the end of the average 3.6-year follow-up period, HbA1c in the patients who were randomised to statin therapy was 1.3 mmol/mol (95% CI 0.4, 2.2) higher than in patients from the control group.

Statins and adiposity traits

Swerdlow *et al.*[67] examined the effect of statin therapy on bodyweight. This meta-analysis included 91,393 non-diabetic participants from 15 trials. An increase in bodyweight (0.33 kg, 95% CI 0.25, 0.42) was observed in the statin vs. placebo/standard care controlled trials but not in the intensive statin vs. moderate dose trials (-0·15 kg, 95% CI -0.39, 0.08). A potential weakness of Swerdlow's study is that it used data from multiple studies and, therefore, could not explore heterogeneity between RCTs in the individual characteristics of the participants. Moreover, it's unclear if the adiposity measures were done at the recruitment or during the follow-up period when adiposity could potentially impact mortality. Thus, the evidence of the effect of statins on body weight might not be sufficiently robust.

2.3.3 Search 3: *NPC1L1* genetic studies

The initial literature search identified 6 papers. Two papers passed eligibility criteria and were included in the review. Detailed results of the search are presented in a PRISMA flow diagram (**Figure 2:3**).

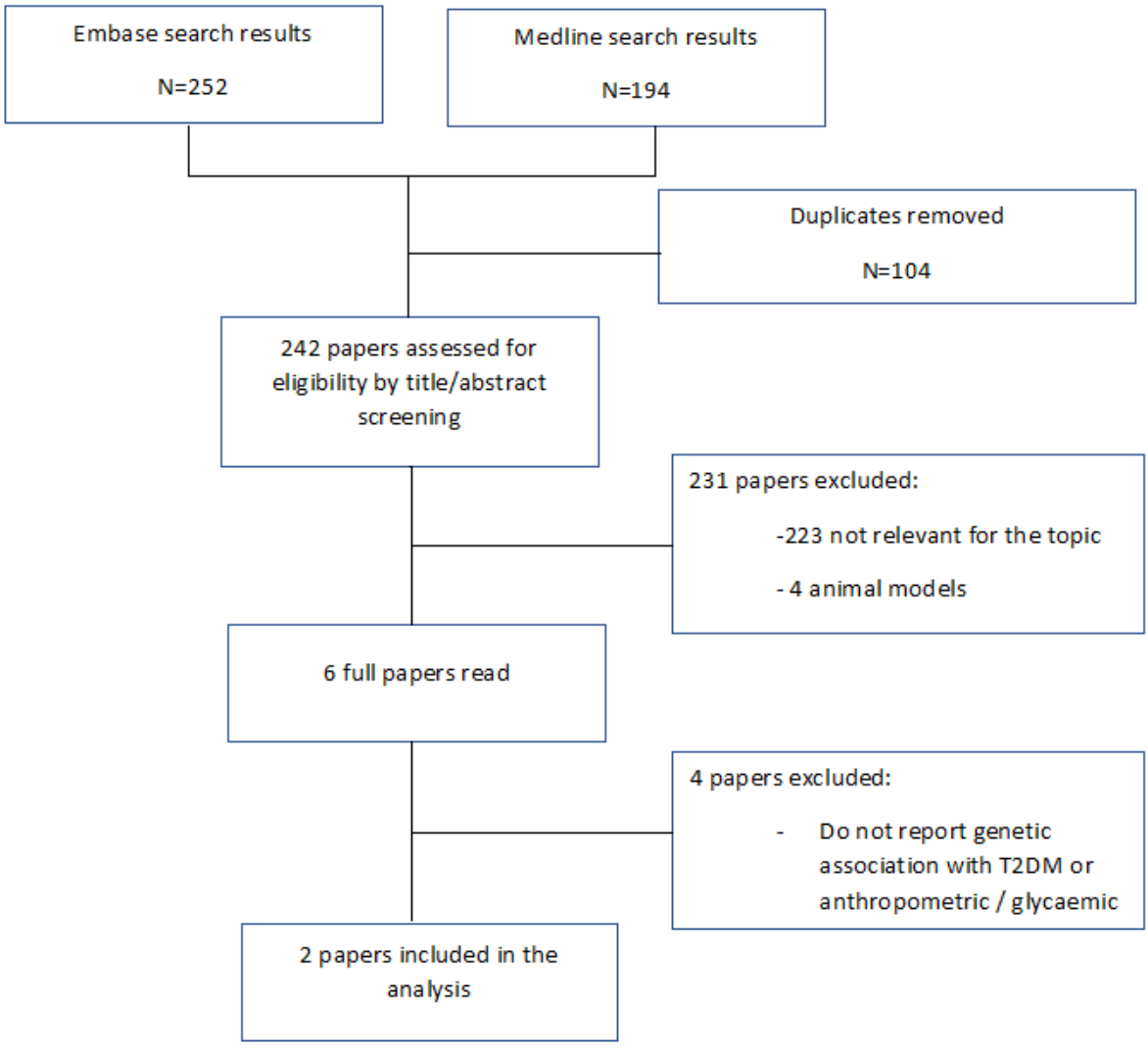


Figure 2:3 | PRISMA flow diagram for Search 3 (*NPC1L1* genetic studies)

The effect of genetic variants in *NPC1L1* on the risk of T2DM was reported by Lotta *et al.*[30]. A meta-analysis of 3 genetic studies (EPIC-InterAct, UK Biobank, DIAGRAM), with 50,775 cases of T2DM and 270,269 controls, was used to examine genetic associations of *NPC1L1* in participants of European descent. Two *NPC1L1* SNPs

(rs2073547 and rs217386) were associated with lower LDL-C and higher risk of T2DM. There were no significant associations with BMI, waist-to-hip ratio, fasting glucose, or fasting insulin (Appendix 2.3). Combined in a GRS, *NPC1L1* was associated with T2DM (OR 2.42, 95% CI 1.70, 3.43 per 1 SD lower genetically predicted LDL-C). There was no material difference between the effect of *NPC1L1* GRS on the risk of T2DM in men (OR 2.17, 95%CL 1.00, 4.71) and women (OR 2.27 95% CI 0.97, 5.33). *NPC1L1* GRS was not associated with anthropometric measures (BMI 0.08 kg/m² 95% CI -0.04, 0.20, waist-to-hip ratio -0.01 95% CI -0.14, 0.12) or with glycaemic traits (fasting glucose 0.14 SD 95% CI -0.05, 0.33, fasting insulin -0.15 95%CO -0.37, 0.07) per 1 mmol/L lower LDL-C. The authors concluded that *NPC1L1* genotype is associated with lower LDL-C levels and higher risk of T2DM.

Another study by Ference *et al.*[51] compared the effect of lower genetically predicted LDL-C (by *NPC1L1, HMGCR* and *PCSK9*) on the risk of T2DM. The study included 112,772 participants of European descent from 14 prospective cohort or case-control studies with 10,635 cases of T2DM. The *NPC1L1* GRS, comprised of 5 SNPs (rs217386, rs2073547, rs7791240, rs10234070 and rs2300414), was not associated with the risk of T2DM (OR 1.07 95% CI 0.98, 1.18 per 10 mg/dl, which is equivalent to OR 1.31 95% CI 0.93, 1.89 per 1 mmol/L lower genetically determined LDL-C).

NPC1L1 SNPs included in two studies are shown in **Table 2:4** below.

Table 2:4 | Frequencies of each *NPC1L1* SNP analysis across the selected manuscripts

SNP	Lotta *et al.* (2016)	Ference *et al.* (2015)	N of papers reported
rs2073547	x	x	2
rs2073547	x	x	2
rs7791240		x	1
rs10234070		x	1
rs2300414		x	1

2.3.4 Search 4: RCTs of ezetimibe

The initial search retrieved 242 papers after duplicates were removed. One paper passed eligibility criteria and was included in the review. Detailed results of the search are presented in a PRISMA flow diagram **(Figure 2:4).**

There was an exception for one conference paper[26] reporting the effect of ezetimibe on T2DM in the IMProved Reduction of Outcomes: Vytorin Efficacy International Trial[83] (IMPROVE-IT). The exception was made because it was the only large trial of ezetimibe in which the effect on T2DM was reported at a conference by the key investigators. No other RCT of ezetimibe assessed its effect on T2DM. Therefore, the exception was made for the conference paper reporting evidence of the possible adverse effect of ezetimibe on the risk of T2DM.The largest RCT of ezetimibe conducted to date, IMPROVE-IT[83], recruited 19,144 post-acute coronary syndrome participants. 9,067 participants were randomised to ezetimibe 10 mg /simvastatin 40 mg and 9,077 participants to simvastatin 40 mg with placebo. A post-hoc analysis of participants without T2DM at baseline found no difference in the risk of new-onset T2DM between treatment arms (HR 1.04, 95% CI 0.94, 1.15)[25].

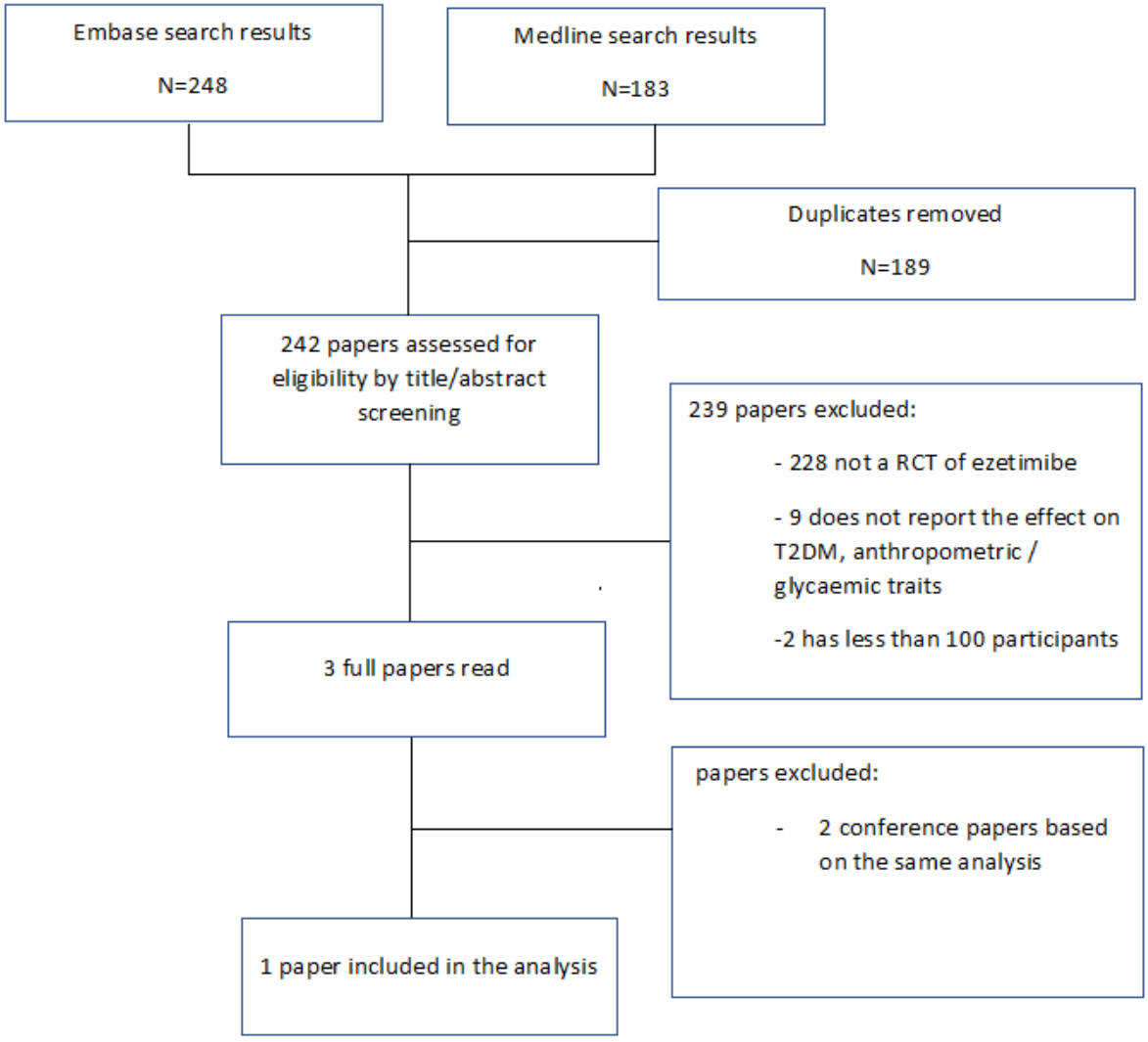


Figure 2:4 | PRISMA flow diagram for Search 4 (RCTs of ezetimibe)

2.4 Discussion

Four searches were performed to systematically examine the available evidence from genetic studies and from RCTs regarding the effects of statins and ezetimibe on the risk of T2DM, glycaemic traits and anthropometric measures. The review included seven studies which used *HMGCR* genetic variants as proxies of statin therapy, and two studies which used *NPC1L1* genetic variants as proxies of ezetimibe, to assess the associations between genetically determined lower LDL-C and glycaemic traits and anthropometric measures. These studies used largely overlapping genetic data available from major genetic consortia. The analysis of the association between *HMGCR* GRS and T2DM showed that genetically lower LDL-C was associated with higher risk of T2DM, and the results were concordant with meta-analyses of RCTs of statins. However, there was an apparent discrepancy in the effect of *NPC1L1* GRSs on risk of T2DM in two genetic studies. While one study (which used 5 SNPs for the *NPC1L1* GRS) found no association with T2DM and was concordant with the results of IMPROVE-IT, the other study (with 2 SNPs in the GRS) showed a 2-fold higher risk of T2DM. The disagreement in the results of the effect of *NPC1L1* GRS on the risk of T2DM and results from the only large RCT of ezetimibe render it difficult to draw any robust conclusion about the effect of NPC1L1 inhibition on T2DM. There is therefore a need for more genetic and clinical evidence before any firm conclusions can be made.

Although nine meta-analyses of RCTs of statins met the inclusion criteria, there was considerable overlap in the included trials. The two largest and most recent papers[20,84] have best summarised the evidence of the effect of statins on the risk of T2DM. These

studies also established a dose-dependent association between statins as a class and an increased risk of incident T2DM. Statin therapy was also associated with a small increase in body weight[67], which supports the evidence of a diabetogenic effect of statins, as weight gain plays an important role in the aetiology of diabetes[85].

Two papers[86,87] which did not satisfy the inclusion criteria for this systematic literature review (as they were not meta-analyses) offered an interesting interpretation of the post-hoc results of several RCTs of statins. Both papers suggested that the diabetogenic effect of statins was limited to the participants with pre-diabetes. At the same time, those with normal glycaemic control appeared to have no increased risk of incident T2DM.

Waters *et al*.[86] compared the effect of statin therapy on the risk of incident T2DM among 15,056 participants of the TNT and IDEAL RCTs. In both trials, participants with prevalent cardiovascular disease were randomised to a low-dose statin (10 mg atorvastatin in TNT, 20-40 mg simvastatin in IDEAL) or a high-dose statin (80 mg atorvastatin in both trials). A further stratification by the number of pre-existing risk factors for T2DM (such as fasting blood glucose > 100 mg/dl, fasting triglycerides > 150 mg/dl, BMI > 30 kg/m^2, and hypertension) allocated people to the low-risk group (0-1 risk factor) or high-risk group (2-4 risk factors). The study showed no material difference in the incidence of T2DM among participants with 0-1 risk factors for T2DM randomised to low- vs high-dose statins (HR 0.97, 95% CI 0.77, 1.22). The results among participants with 2-4 risk factors for T2DM were different – those randomised to the high-dose statin had a 24% (HR 1.24, 95% CI 1.08, 1.42) higher risk of incident T2DM than those in the low-dose statin group. The authors concluded that participants at higher risk of T2DM at baseline experienced an excess risk of statin-associated

incident T2DM, whereas those at low risk of T2DM at baseline had no increased risk associated with a high-dose statin regimen.

The second paper[87] assessed the effect of rosuvastatin on the risk of incident T2DM in 17,603 participants without prior cardiovascular disease or T2DM from the JUPITER RCT. In this post-hoc analysis, 6,095 participants with no risk factors for T2DM (such as metabolic syndrome, impaired fasting glucose, body mass index ≥ 30 kg/m^2, or HbA1c > 6%) were included in the low-risk T2DM group. The remaining 11,508 participants with at least one risk factor for T2DM were included in the high-risk T2DM group. In the low-risk group, allocation to rosuvastatin 20 mg was not associated with an increased risk of incident T2DM (HR 0.99, 95% CI 0.45, 2.21). Randomization to rosuvastatin 20 mg in the high-risk group was associated with a 28% (HR 1.28, 95% CI 1.07, 1.54) increase in incident T2DM. The authors concluded that the statin-induced risk of T2DM may be limited to participants with pre-existing risk factors for T2DM.

While not formally included in this review, these papers set the context for further investigation of the risk factors associated with the development of incident T2DM in patients prescribed statins.

There have been few major RCTs conducted with ezetimibe and only one RCT has assessed the effect of ezetimibe on the risk of incident T2DM in patients. The trial reported no association between ezetimibe and risk of T2DM (HR 1.04, 95% CI 0.94, 1.15). There was no randomised evidence from large RCTs about an effect of ezetimibe on glycaemia or adiposity. Lack of evidence for the association of ezetimibe with T2DM and diabetes-related traits is to be expected; in the case of statin therapy, which has a stronger effect on lowering LDL-C, 13 large RCTs were needed to

generate enough data to establish an association with increased risk of T2DM. One trial of a lipid-lowering agent with a considerably weaker effect on LDL-C is highly unlikely to generate solid evidence for any effect on T2DM.

Both evidence from RCTs and from genetic studies showed associations between *HMGCR* GRS and statin therapy with slightly higher body weight. Neither IMPROVE-IT nor genetic papers on *NPC1L1* assessed the effect ezetimibe or its genetic proxy on anthropometric measures.

The systematic review showed evidence supporting an increased risk of T2DM associated with pharmacological inhibition of HMGCR and *HMGCR*-predicted genetically lower LDL-C. However, there was insufficient evidence to attribute the increased risk of T2DM to the participants with pre-diabetes or to conclude that the diabetogenic effects of statins did not affect those with normal glycaemic control. Moreover, this systematic literature review yielded little evidence on the associations with glycaemic traits and anthropometric measures. Reported associations were often obtained from largely overlapping data from large genetic consortia. Far less research has been done on the effect of ezetimibe and *NPC1L1* genotype on T2DM. Only one paper reported the associations of genetic variants in *NPC1L1* with glycaemic traits and anthropometric measures.

2.5 Summary

Evidence from meta-analyses of RCTs of statins confirmed the association between statin therapy and increased risk of incident T2DM. Statin therapy was also associated with slightly higher body weight compared to placebo or standard care. Genetic studies of *HMGCR* demonstrated concordant results.

The only large RCT of ezetimibe showed no association between ezetimibe and incident T2DM. The review of genetic studies also failed to produce compelling evidence of the effect of the *NPC1L1* genotype on risk of T2DM.

The systematic review showed evidence supporting an increased risk of T2DM associated with pharmacological inhibition of HMGCR and *HMGCR*-determined genetically lower LDL-C. However, there was insufficient evidence to attribute the increased risk of T2DM to the participants with pre-diabetes or to conclude that the diabetogenic effects of statins did not affect those with normal glycaemic control. Moreover, this systematic literature review yielded little evidence of the associations with glycaemic traits and anthropometric measures. Reported associations were often obtained from largely overlapping data from large genetic consortia.

Far less research has been done on the effect of ezetimibe and *NPC1L1* genotype on T2DM. Only one paper reported the associations of genetic variants in *NPC1L1* with glycaemic traits and anthropometric measures.

Chapter 3 Overall study design and methods

This chapter describes the data and methods used throughout the thesis, including the UK Biobank study and large international genetic consortia, procedures followed to define the disease outcomes and related phenotypes, construction of the genetic instruments, and statistical methods used to answer the research questions. Additional objective-specific details on the methods and statistical analyses are provided in subsequent chapters, where relevant.

3.1 UK Biobank

The UK Biobank is a large-scale population-based prospective cohort study that enrolled 502,536 middle-aged adults from the UK general population, selected from the National Health Service (NHS) population-based register, limited to those living within reasonable traveling distance from the assessment centres[88]. The primary goal of the UK Biobank is to learn about the causes of common life-threatening and disabling diseases and improve their prevention, diagnosis, and treatment[89]

3.1.1 UK Biobank participant recruitment and baseline assessment

In 2006 UK Biobank sent more than 9 million invitations to join the study to patients registered with the NHS in England, Wales, and Scotland[90]. Eligible candidates aged 40-69 years and living within 25 miles of one of 22 assessment centres were identified

from the NHS records[90]. The Department of Health received demographic information and contact details of the eligible candidates from the electronic NHS records and sent personalised recruitment letters to 9,238,433 individuals[91].

With about a 6% response rate[92], UK Biobank had a total of 576,926 candidates who agreed to participate in the study, and 503,317 participants attended an assessment centre between 2006 and 2010[90]. During the baseline assessment, participants completed a touch screen questionnaire with questions covering such topics as socio-demographic characteristics, lifestyle exposures, psychological state, cognitive function, family health history, and personal medical history[93]. The touch screen questionnaire was followed by a nurse-led interview, which collected more detailed information on the demographic and medical history of the study participants. A nurse collected detailed information about cancer and non-cancer illnesses, operations, and use of medication (if a participant has mentioned a medical condition or indicated use of any medication in the touchscreen questionnaire)[94]. All participants had body composition and blood pressure measured after the completion of the interview. Physical measures of waist and hip circumference were followed by weight and body fat measurement by bioelectrical impedance[95]. Hospital episode statistics (HES)[96] were available for the periods prior to and post enrolment to the study. Prior HES records dating back to April 1997, and post enrolment data up to March 2017, were used in the analyses presented in this thesis.

3.1.2 Research ethics and permissions

Any study involving human participants, including UK Biobank, is subject to the framework of national legislation and governance. However, to address the

unprecedented scale of the study, the number of biological materials to be collected and stored, and the need to win and maintain considerable public trust, UK Biobank developed an additional ethical framework – UK Biobank Ethics and Governance Framework[97]. Along with a number of different regulators, such as the Human Tissue Authority, the Charity Commission and Companies House and the Information Commissioner's Office, UK Biobank is overseen by the Ethical Advisory Committee, which provides the advice on ethical issues related to the development, maintenance and use of the UK Biobank.

An amendment to the original Research Ethics Committees approval to the UK Biobank was granted by North West Multi-centre Research Ethics Committee on 10 May 2016 (REC reference 16/NW/0274). This allowed access to the participants' electronic health records in England and Wales, UK Biobank had approval from the Patient Information Advisory Group, and, in Scotland, from the Community Health Index Advisory Group.

UK Biobank data were accessed under approved project application 14568.

.

3.2 Data collection and procedures

The following sections describe the data available, focusing on aspects of the UK Biobank data that provide information regarding diabetes and associated traits that are material to this thesis.

3.2.1 UK Biobank touch screen questionnaire

A touch screen self-completion questionnaire was used to collect data about lifestyle, family history and early life exposures, medical history, and medication use[93]. Several diabetes-specific questions included "Has a doctor ever told you that you have diabetes?" and "What was your age when the diabetes was first diagnosed?". Two other questions asked about the use of hyperglycaemic medication, namely "Did you start insulin within one-year of the diagnosis of diabetes?" and "Do you regularly take any of the following medications? (You can select more than one answer)" with "insulin" as an answer option. The use of other diabetes medications was considered during the verbal interview (Section 3.2.2).

3.2.2 UK Biobank nurse-led verbal interview

The verbal interview aimed to collect detailed information about the sociodemographic characteristics, occupation, and health of the participants[94]. Questions asked during the interview were linked to the participants' answers in the touch screen questionnaire. A computer system notified an interviewer if the participant had answered in the touch screen questionnaire that a doctor had told them that they had

cancer or other non-cancer illnesses. Participants who reported being told by a doctor that they had diabetes were asked about the type of diabetes and year of diagnosis be the nurse.

Similarly, each participant who indicated the use of prescription or over-the-counter medicine was asked the name of the medication. Information about medications was recorded either by a generic or trade name. UK Biobank recorded current medications, including regular prescription medications and regular treatment taken weekly, monthly, or quarterly. Prescribed but not taken or recently stopped medications were not included.

Data collected at the verbal interview were used to identify people who were taking oral hypoglycaemic agents or insulin injections. Oral hypoglycaemic agents and insulin codes used for the analysis are listed in Appendix 3.1 and were verified by a clinician. UK Biobank codes (from the UK Biobank coding system) for diabetes are shown in Appendix 3.2.

3.2.3 UK Biobank anthropometric measurements

The following measurements were selected for the project based on biological relevance to T2DM and included: bodyweight, height, waist circumference, hip circumference, waist-to-hip ratio, body mass index (BMI), body fat mass, and lean body mass.

Bodyweight, body fat mass and lean body mass were measured in kg by the Tanita BC-418 MA body composition analyser. Body fat mass and lean body mass were estimated by bioelectrical impedance.

Height was measured with a Seca 202 height measure (in cm). BMI (kg/m^2) was calculated by dividing weight (in kg) by the square of height (in m). Waist and hip circumferences were measured in cm using a Wessex non-stretchable sprung tape measure. The waist circumference was measured at the level of the umbilicus; the hip circumference was measured around the widest portion of the buttocks[95]. The waist-to-hip ratio was calculated by dividing the waist measurement by the hip measurement.

3.2.4 UK Biobank blood biomarkers

UK Biobank measured various blood serum biomarkers which were selected due to their established associations with common chronic diseases. Blood samples were analysed centrally using automated analysers. In addition, several quality control procedures and method performance protocols were implemented to ensure accuracy and precision of laboratory testing (the quality control pipelines are described in the UK Biobank serum biomarker data accompanying document, and other puplications[98,99]).

For the current study, two circulating biomarkers of lipid metabolism (LDL-C and Apo B 100) and two biomarkers of glucose homeostasis (HbA1c and non-fasting glucose) were of particular relevance.

Direct LDL-C was measured on Beckman Coulter AU5400 analysers using an enzymatic selective protection analysing method (coefficient of variation <2% and reportable range 0.26 to 10.3 mmol/L). Apo B was also measured on Beckman Coulter AU5800 analysers using an immuno-turbidimetric method (coefficient of variation <3% and reportable range 0.4 to 2 g/L). Non-fasting glucose was measured by Beckman

Coulter AU5800 using an enzymatic analysing method (reportable range 0.6 to 45 mmol/L, coefficient of variation <2%). Finally, HbA1c was analysed using the VARIANT II TURBO Hemo-globin Testing System with a reportable range of 15-184 mmol/mol, coefficient of variation <2%). Participants who reported the use of glucose-lowering medication or had HbA1c values outside of the reportable range were excluded from the analysis of the effect of GRSs on HbA1c and non-fasting glucose.

3.2.5 Hospital episode statistics and death registry

Hospital Episode Statistics (HES) is a database containing information about all hospital admissions at NHS hospitals. The regularly updated HES dataset is linked to the UK Biobank participants' records and contains a broad scope of information about each hospital admission, including primary and secondary diagnoses, and surgical procedures. Information about diagnosis is recorded in the World Health Organization's ICD-10 (International Classification of Diseases and Related Health Problems) format.

Data on hospital admissions were available for all participants. However, at the time of data analysis, there was no access to primary care data[96]. That is especially relevant to identifying incident T2DM cases, as the disease is mainly diagnosed and treated in primary care and therefore not all diabetes cases may be reliably identified using HES data[100].

All participants were followed up for cause-specific mortality by linking with NHS death registry data. Every six months, NHS Digital provided information on mortality for participants from England and Wales[101]. For Scottish participants, the NHS Central Register, a part of the National Records of Scotland, provided mortality data

quarterly[101]. Mortality data contain information on the date of death, age at the time of death, and ICD-10 codes for primary and contributory causes of death. For the current project, death records were analysed, and participants with ICD-10 codes for T2DM in the death certificates were classified as having diabetes. Death records up to March 2017 were included in the analyses.

3.2.6 Genetic data

Blood samples collected at the baseline assessment of all UK Biobank participants were genotyped. Two different genotyping arrays were used; a small proportion of samples (11%, 36,688 participants from the study population) were genotyped with a purpose-designed Applied Biosystems™ UK BiLEVE Axiom™ Array by Affymetrix, which included 807,411 variants. The remaining 297,828 (89%) samples were genotyped using another purpose-designed Applied Biosystems™ UK Biobank Axiom™ Array (825,927 markers)[102] Reference panels based on data from the UK10K, Haplotype Reference Consortium and 1000 Genomes Phase 3 were used for imputation, which increased the number of SNPs available for analysis up to 92 million[103].

To measure the ethnic population structure of the participants, UK Biobank performed principal components analyses and computed 40 principal components (PC) which explain the most variation in ethnicity[103]. For the current project only white British participants with genetically confirmed ancestral origins were selected.

Several quality controls (QC) were applied to the selected participants (the entire quality control pipeline is described elsewhere) [72,103]. The QC criteria included a check for:

- Duplicate samples
- Genetic and self-reported sex discordancy
- Low genetic call rate (the proportion of samples in which the confident genotype call could be made) and high missing rate
- Relatedness up to 3rd degree
- Withdrawn consent

In total, 334,516 unrelated White British UK Biobank participants passed genetic QC and were included in the current research project.

3.3 Type 2 diabetes mellitus phenotype definition

Cases of T2DM were algorithmically defined following a slightly modified procedure suggested by Eastwood et al[104]. A wide range of data was used to classify participants into diabetic and non-diabetic groups and further discriminate between different types of diabetes. The algorithms utilised self-reported information about diagnosis and use of hypoglycaemic drugs, as well as HES records, pre and post UK Biobank baseline assessment date, and the Death Registry. Blood biomarkers of glucose homeostasis were not initially available when analyses to identify T2DM were performed for this project and are therefore not included in the algorithms to identify cases of diabetes.

The major challenge in the phenotype definition was to discriminate between different types of diabetes due to inconsistency in HES records and lack of specification in the self-reported data. For example, it was not uncommon for a participant to have records of type 1 and type 2 diabetes in different hospitalisation records (10.4% of participants with T2DM have a record of T1DM) or to have an unspecified diabetes record (11.7% of participants have a record of unspecified diabetes). To address this, two decisions were made. First, the type of diabetes was defined following the first record of diabetes in HES. Second, in the case of discrepancy between self-reported data and HES records, priority was given to the self-reported data as a more accurate description of the disease.

3.3.1 Prevalent cases of T2DM

Prevalent cases of T2DM were defined using data from the UK Biobank clinic (self-reported information and glucose-lowering drugs) and HES records of T2DM events prior to the baseline UKB assessment date. The algorithm for prevalent T2DM is based on several logical rules and cross-tabulation of different data, shown in Appendix 3.4. Discrimination between cases of T1DM and T2DM took into account the variation in the age of onset of the disease in different ethnic groups. The threshold between the age of onset of T1DM and T2DM for participants with white British ancestry was 36 years[104], whereas, for the participants with Asian, African and Caribbean ancestry, it was 30 years[105]. Participants diagnosed with diabetes prior to the thresholds were considered to have T1DM.

The list of hypoglycaemic drugs which were included in the algorithm is shown in Appendix 3.1. The use of metformin alone (i.e. without other hypoglycaemic drugs or history of diabetes) was not considered sufficiently strong evidence of diabetes due to the potential of this drug to target other comorbidities such as polycystic ovarian syndrome and non-diabetic hyperglycaemia.

The algorithms begin with identifying all cases of diabetes and then progress to discriminate between types of diabetes and estimate of the likelihood of the selected type of diabetes. Consequently, T2DM cases were classified into ‘possible’ and ‘probable’ categories based on the strength and quantity of evidence supporting the allocation to T2DM. Participants identified as having probable T2DM were defined as T2DM cases in all analyses that follow. All cases of unspecified diabetes, T1DM and possible T2DM were excluded from the analyses. Participants who did not report diabetes during the baseline assessment but had prior ICD-10 codes for T2DM in the

HES prior to the date of recruitment records were also classified as prevalent T2DM cases.

3.3.2 Incident cases of T2DM

Definitions of incident T2DM differed from the algorithm suggested by Eastwood[105]. In the original paper, derivation of incident cases of T2DM was limited to a subset of Welsh participants whose records were linked to primary care data. However, at the time of the current analysis, primary care data for the whole UK Biobank population were not available. Therefore, the following algorithm was developed: all cases of T2DM found in the HES records or death certificates with a post-baseline assessment date and which were not recorded at the UK Biobank baseline assessment were classified as incident T2DM Appendix 3.6.

3.3.3 Total T2DM cases

The combination of prevalent and incident cases of T2DM constituted total T2DM cases.

3.4 Genetic consortia

Global-scale genetic consortia data were used for two purposes (i) to estimate the associations of genetic variants with the exposure (LDL-C levels) and (ii) to assess the effect of genetic variants and GRSs on HbA1c and anthropometric measures in an independent dataset for the validation analysis.

3.4.1 The Global Lipids Genetics Consortium

The Global Lipids Genetics Consortium (GLGC) is an international research collaboration seeking to identify genetic variants associated with lipid traits[45]. The first GLGC genome-wide meta-analysis included 46 studies with 188,577 European participants and 7,898 non-European participants[45]; 37 studies consisted of the European population, and eight studies consisted of participants with non-European ancestry[45]. Fasting total cholesterol, high-density lipoprotein-cholesterol (HDL-C), and triglycerides were measured in the majority of studies. LDL-C levels were directly measured in ten studies (24% of participants) and were estimated using the Friedewald formula with missing values assigned to the participants with triglycerides >400mg/dL in the rest of the studies. In most studies, participants who were prescribed lipid-lowering medications were excluded from the meta-analyses; three studies, however, included those who were known to be on lipid-lowering therapies, but the percentage of such participants in each study was fewer than 2%, and therefore no adjustments were made.

The current project used genetic associations (beta-estimates and standard errors) between individual SNPs and LDL-C reported by GLGC 2013[45] to compute two genetic instrumental variables for the MR based analysis.

3.4.2 The DIAbetes Genetics Replication And Meta-analysis

The DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) consortium is an international research collaboration dedicated to discovering the genetic basis of T2DM[47]. DIAGRAM includes 34,840 cases of T2DM and 114,981 controls of European descent from 12 GWAS. In addition, Sex-specific and sex-combined genetic associations with T2DM from the DIAGRAM study were used for external validation analyses of the effects of *HMGCR* and *NPC1L1* on the risk of T2DM.

3.4.3 The Meta-Analyses of Glucose and Insulin-related traits Consortium

The Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC) combines data from 52 GWAS to identify genetic variants that regulate glycaemic and metabolic traits in European ancestry dominated population (68.0%)[106]. In addition, sex-combined data for HbA1c from 281,416 individuals without T2DM[106] were used for an external validation analysis of the associations between *HMGCR* and *NPC1L1* GRSs, and HbA1c.

3.4.4 The Genetic Investigation of Anthropometric measures

The Genetic Investigation of Anthropometric measures (GIANT) consortium is an international collaboration aiming to discover the genetic architecture of human body composition, including anthropometric measures of obesity[74] and height[107]. The sex-

specific systematic meta-analysis of 114 studies (>320,000 individuals of European ancestry) was used to conduct an external validation analysis of the effect of *HMGCR* and *NPC1L1* GRSs on BMI, waist and hip circumference, waist-to-hip ratio, and height[108]. These data release reported sex-specific and sex-combined genetic associations (beta-estimates and standard errors) with anthropometric measures, excluding the UK Biobank population.

3.5 Genetic methods

This research project used the principles of MR to assess the causal relevance of the associations between the genetic proxies of statin therapy and the risks of T2DM and effects on diabetes-related glycaemic and anthropometric measures. Then it repeated the analysis for another LDL-C lowering drug target – a genetic proxy of ezetimibe. This section describes the Mendelian randomisation approaches and the computation of *HMGCR* and *NPC1L1* GRSs.

3.5.1 Drug target MR

Unlike most MR studies, which use multiple polymorphisms from different loci as a GRS to examine the effect of a risk factor on the disease outcome, the current project focuses on the MR analysis of a single protein. It limits the instrumental variable to genetic variants located in the encoding gene and several adjacent base pairs regulating the expression of the gene. As data on the genetic associations with proteins levels are not available, the associations with a biomarker (LDL-C) were used as a proxy. Variants in low linkage disequilibrium were weighted by their effect on the exposure biomarker and then combined into a GRS.

Two weighted GRSs, one for *HMGCR* and another for *NPC1L1*, were used as instrumental variables in the current project. Unlike the example in the introduction (**Figure 3:1**), where GRS was classified into two groups to mimic the allocation to the treatment and control groups, we used continuous scale GRSs.

Several steps were taken to address the potential limitations of the MR method. First, only White British participants were included in the study population to avoid a population stratification bias. Ethnic background was self-reported by participants by selecting from a list of pre-specified categories. Although 94% of participants reported 'white' ethnic background, about 22,000 of them originated outside Europe, therefore, to account for potential genetic variability between ethnically diverse participants properly, all statistical models were adjusted for 40 principal components.

Second, LD clumping procedure was performed at the stage of selecting the variants to ensure independence between the chosen SNPs (the computation of the *HMGCR* GRS is described in the following section in more details).

The current project used a two-sample MR analysis, which applies estimates of the associations between genetic variants with the exposure (LDL-C) from GLGC. Then, the association of the genetic variants with the risk of T2DM was estimated in an independent sample of UK Biobank participants. Finally, those estimates were combined to assess the effect of life-long exposure to *HMGCR*-predicted (or *NPC1L1*-predicted) lower LDL-C on the risk of T2DM, and effects on glycaemic and anthropometric measures.

3.5.2 *HMGCR* genetic risk score

Using a GRS provides greater power to explore the association between a genotype and an outcome than single genetic variants. Therefore, independently inherited polymorphisms (SNPs) within the *HMGCR* gene were combined in the *HMGCR* GRS. Publicly available GWAS summary data on the association between individual genetic variants and LDL-C from the GLGC consortium was used to select the SNPs for the

GRS. First, 2,818 SNPs within 100kb of the *HMGCR* gene were selected (base pair positions ranged between chr5:74,757,929 and chr5:74,532,154 human assembly GRCh37/hg19). Then, the strength of selected SNPs' associations with LDL-C, reported by GLGC, was assessed and 113 SNPs associated with LDL-C at GWAS level of significance ($p \leq 5x10\text{-}8$) were selected. Next, PLINK software and 1000 Human Genome data (Phase 3, the European subpopulation) were used to run a clumping procedure. Clumping selected index SNPs in low linkage disequilibrium with each other ($R2<0.2$) and therefore with independent effects on LD-C. The approach for selecting the five relevant *HMGCR* SNPs selection (rs12173076, rs10515198, rs3857388, rs7711235, rs12916) is shown in the **Figure 3:1**. The results of the clumping procedure are shown in Appendix 3.8.

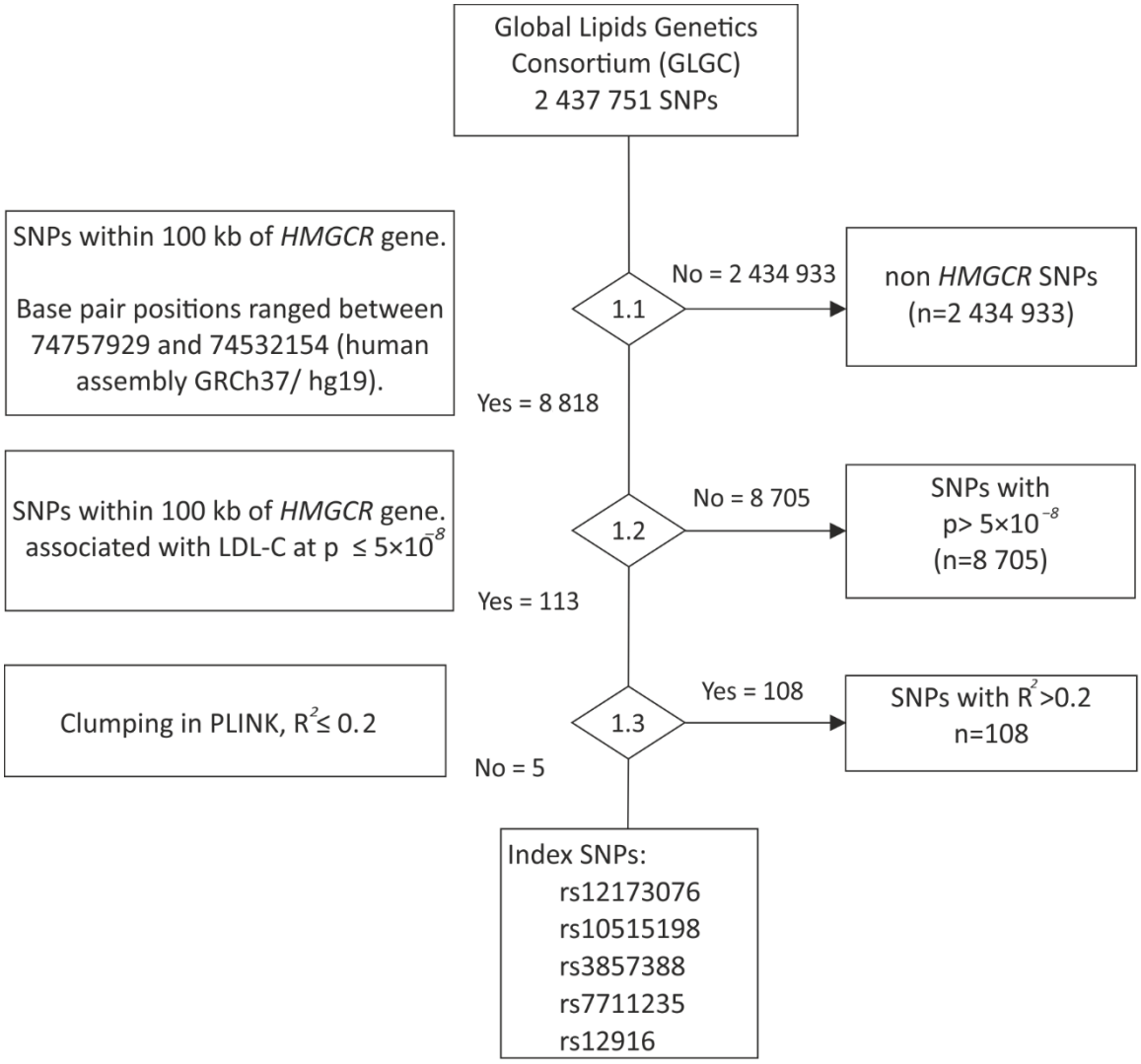


Figure 3:1 | Algorithm for selecting the *HMGCR* index SNPs

The exposure alleles were associated with lower levels of LDL-C based on the estimates reported by GLGC. Several checks were employed to ensure directional concordance between UKB genotype data and GLGC data (**Table 3:1**):

- effect alleles in GLGC and UKB data were compared; GLGC reported trait-increasing alleles as effect alleles, whereas for the current project, alleles associated with lower LDL-C were used as effect alleles to proxy the effect of lipid-lowering medication. The effect estimates from GLGC were re-

parameterised to the LDL-C lowering allele and compared to LDL-C lowering alleles in the UK Biobank; there was no discrepancy between the two datasets.

- effect allele frequencies for European populations in UKB, GLGC and Ensembl were compared. The difference between frequencies is within 3%, which indicates that the frequencies in all three populations are similar.

- checks were made for ambiguous (A/T or G/C pairs) SNPs and strand flipping. There were no ambiguous SNPs.

Table 3:1 | *HMGCR* index SNPs in the UK Biobank, GLGC and Ensembl

RSID	Minor allele UKB	Minor allele GLGC	Minor allele frequency UKB	Minor allele frequency GLGC	Minor allele frequency Ensembl	Freq. difference between UKB and GLGC	Freq. difference between UKB and Ensembl population
rs12916	C	C	0.40	0.43	0.41	-0.03	-0.01
rs12173076	G	G	0.11	0.12	0.11	-0.01	0.00
rs10515198	A	A	0.09	0.10	0.10	-0.01	-0.01
rs3857388	C	C	0.12	0.13	0.12	-0.01	0.00
rs7711235	G	G	0.23	0.27	0.26	-0.03	-0.03

To calculate the *HMGCR* GRS for each participant, the number of exposure alleles in each SNP was multiplied by its respective effect on LDL-C in standard deviation (SD) units as reported by GLGC. The weighted effects of five index SNPs were then summed to create a weighted GRS. The distribution of the *HMGCR* GRS is shown in Appendix 3.9.

3.5.3 *NPC1L1* genetic risk score

To compute the weighted *NPC1L1* GRS, 278 SNPs within 100 kb of the *NPC1L1* gene were selected (base pair positions ranged between chr7:44,552,134 and chr7:44,580,914 human assembly GRCh37/hg19). Then, the strengths of selected SNPs' associations with LDL-C, reported by GLGC, were assessed, and 26 SNPs associated with LDL-C at GWAS level of significance ($p \leq 5x10\text{-}8$) were selected. Next, PLINK software and 1000 Human Genome data (Phase 3, European subpopulation) were used to run a clumping procedure. The clumping procedure selected index SNPs in low linkage disequilibrium with each other ($R2<0.2$) and had an independent effect on LDL-C. The algorithm of the two *NPC1L1* SNPs selection (rs2073547 and rs217386) is shown in **Figure 3:2**. Appendix 3.10 shows the list of SNPs within 100 kb of the *NPC1L1* gene. Appendix 3.11 shows the results of the clumping procedure, and Appendix 3.12 shows the algorithm for selecting the two *NPC1L1* index SNPs (rs2073547, rs217386).

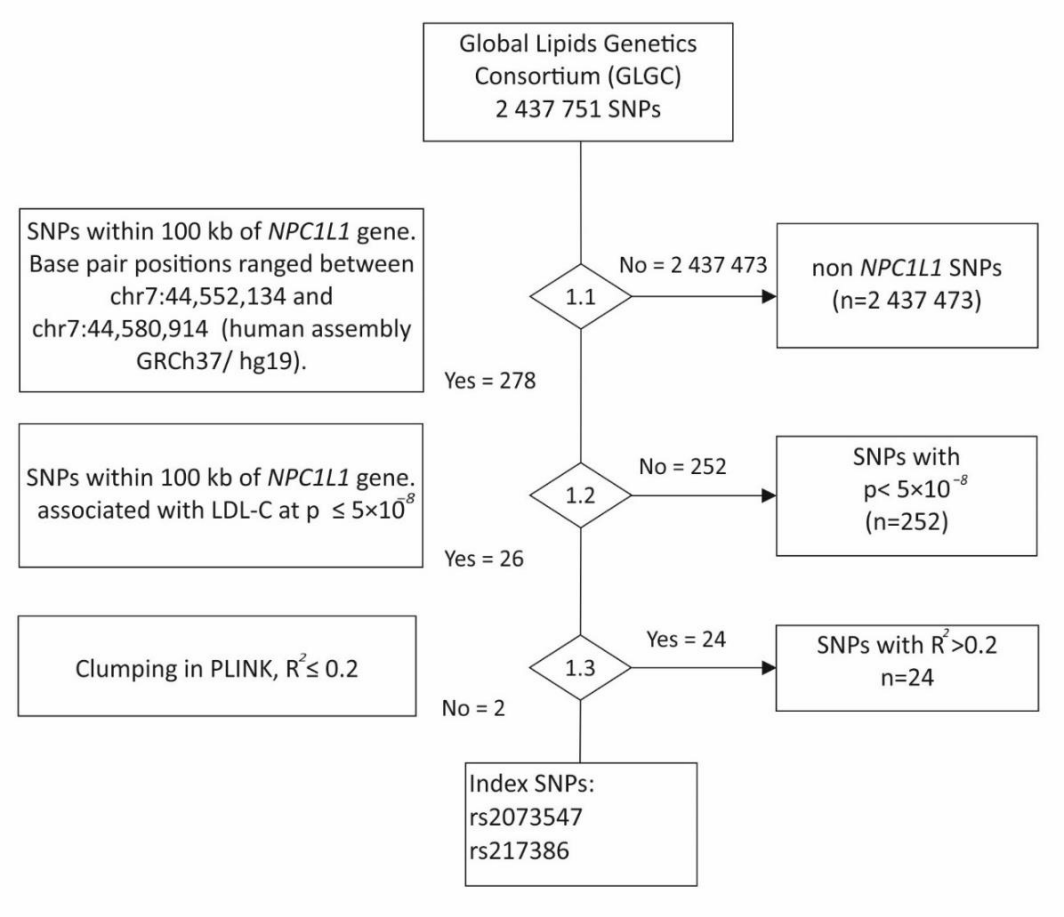


Figure 3:2 | Algorithm of the *NPC1L1* index SNPs selection

Similar to the approach for *HMGCR,* checks were conducted to ensure concordance of genetic data between UK Biobank and GLGC. Minor allele frequencies in UK Biobank, GLGC, and Ensembl were similar, implying ethnically similar populations in both datasets (**Table 3:2**).

Table 3:2 | *NPC1L1* index SNPs in the UK Biobank, GLGC and Ensembl

RSID	Minor allele UKB	Minor allele GLGC	Minor allele frequency UKB	Minor allele frequency GLGC	Minor allele frequency Ensembl	Freq. difference between UKB and GLGC	Freq. difference between UKB and Ensembl population
rs2073547	G	G	0.19	0.19	0.21	0.00	-0.02
rs217386	A	A	0.42	0.41	0.43	0.01	-0.01

3.6 Statistical methods

3.6.1 Main UK Biobank analysis

334,516 unrelated White British UK Biobank participants were included in analyses of the effects of the *HMGCR* GRS *and NPC1L1* GRS on total T2DM and anthropometric measures. In addition, a subset of 320,622 patients without T2DM at baseline was used to assess the effects of these GRSs on incident T2DM.

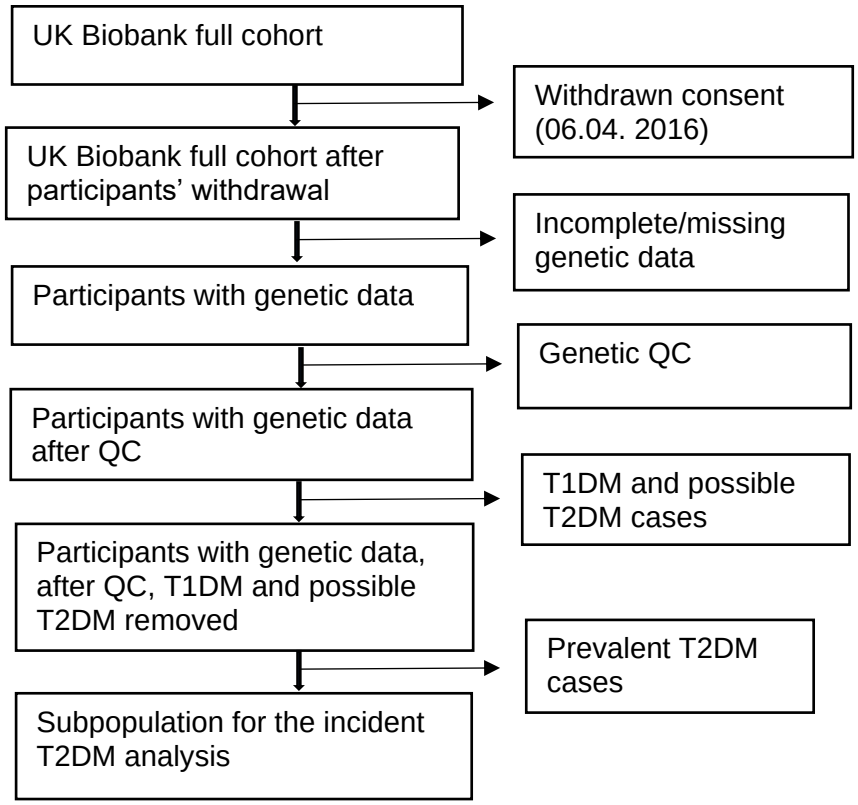


Figure 3:3 | Algorithm for selecting participants for the analysis

Figure 3:3 shows the algorithm for selecting participants for the analysis. 309,458 participants who were not on hypoglycaemic drugs were included in the analyses of the effects of the GRSs on glycaemic traits. Five independent SNPs in *HMGCR*, and

2 SNPs in *NPC1L1*, which were associated with LDL-C concentrations, were combined into respective weighted GRSs for *HMGCR* and *NPC1L1* as described above in section 3.3.2-3.5.3. The association of each index SNP with T2DM was assessed with a logistic regression model. Odds ratios (OR) and 95% confidence intervals (95% CI) were estimated per copy of the LDL-C lowering allele (based on the effect observed in GLGC) [45]. The models were adjusted for sex, array and 40 principal components (to adjust for population stratification). An additive genetic model (also referred to as the per allele model) was used for the analyses, assuming the effect on a trait for one copy of the effect allele was β and 2*β for two copies of the allele.

The association of each GRS with T2DM was assessed with logistic regression models. OR and 95% CI were estimated per 1 SD lower genetically predicted LDL-C (1 SD is approximately equivalent to 1 mmol/L in the GLGC, see Chapter 4, section 4.2.1 for more details)[45].

The associations of individual SNPs and GRSs with continuous glycaemic traits and anthropometric measures were evaluated using linear regression models, similarly, adjusted for sex, array and 40 principal components. Estimates and confidence intervals for each SNP was reported per copy of the LDL-C lowering allele and, for the GRS, per 1 SD lower genetically predicted LDL-C.

To assess the interactions between sex and GRSs, a multiple linear regression model with an interaction term was fitted:

$$Y = \alpha + \beta 1 * GRS + \beta 2 \times sex + \beta 3 \times GRS * sex + adjustments$$

Where Y is a dependent variable (an anthropometric trait), β1 is the effect of GRS in women and β1 + β3 is the effect of the GRS in men. P-value was used to assess if

sex was an effect modifier in the association between the genetic instrument and an anthropometric trait.

ower calculation for the MR study was performed with an online power calculator developed by Brion et al[109]. The power of the MR analysis to detect the effect of 1 SD lower genetically predicted LDL-C on risk of total T2DM was performed for a sample size of 334,516 participants with a proportion of T2DM cases 6.1%. A type 1 error rate of 5% was selected. The proportion of variance in the exposure (LDL-C) explained by *HMGCR* GRS and *NPC1L1* GRS was computed as 0.36% and 0.14% respectively. The effect sizes for both GRSs were taken from Ference's study[51] (OR 1.59 for *HMGCR* and OR 1.31 for *NPC1L1*). The results of power calculations are shown in **Table 3:3** below and indicate sufficient statistical power for the analyses for *HMGCR* and low statistical power for *NPC1L1*.

Table 3:3 | Power calculations for *HMGCR* and *NPC1L1* GRSs

GRS	Number of variants	Variance explained (%)	Calculated power (%)
HMGCR	5	0.36	100.0
NPC1L1	2	0.14	35.0

3.6.2 External validation analysis

External validations of the effects of *HMGCR* and *NPC1L1* on T2DM were performed in an independent dataset (the DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) 1000G GWAS meta-analysis Stage 1 Summary statistics[73], which is the most up-to-date release excluding UK Biobank data at the time of analysis). The same genetic variants as used in the UK Biobank analyses within the *HMGCR* and *NPC1L1* genes were used in the DIAGRAM consortium. For the summary level data reported

by the DIAGRAM consortium, the percentage of the variance of a trait explained by the index SNPs included in the GRS was estimated by:

$$R^2 = \beta^2[2*MAF(1-MAF)]$$

where β is a point estimate of the effect of an index SNP on trait, and MAF is its minor allele frequency. The variance explained by the GRS using summary level data was estimated by summing up R^2 computed for the index SNPs included in the GRS:

$$\text{GRS Variance explained} = \sum_{i=1}^{n} R_i^2$$

The percentage of variance explained by the GRS was estimated by:

$$(\text{Variance explained})*100$$

To assess the effects of the *HMGCR* and *NPC1L1* GRSs on the risk of T2DM in summary-level datasets, 2-sample MR was used. This approach integrated summary statistics from the GWAS reporting the effect on exposure (LDL-C)[110] and summary statistics from the GWAS reporting the effect on outcome (T2DM)[47]. The combined effect of the GRSs on the risk of T2DM, which was estimated in consortium summary-level data and UK Biobank individual-level data, was estimated by a fixed-effects meta-analysis.

External validation of the association between the GRS and HbA1c used summary-level data from the European population published by MAGIC[106] (described above in section 3.4.3). The data were not sex-specific and did not allow for a sex-stratified comparison.

External validation analysis for the associations of the GRSs with anthropometric measures (BMI, waist and hip circumferences, waist-to-hip ratio and height) were

performed on the independent dataset from the Genetic Investigation of Anthropometric measures (GIANT) consortium[107] (see section 3.4.4 for more details). Summary-level data stratified by sex were available for the anthropometric measures, which describe general and central adiposity and are the risk factors for T2DM.

Chapter 4 The associations of *HMGCR* and *NPC1L1* GRSs with LDL-C and Apo B-100 in UK Biobank

4.1 Introduction

This chapter assessed the associations between genetic instruments and blood lipids in UK Biobank. It examined the associations of individual genetic variants and GRSs with (i) LDL-C and (ii) Apo B-100. The associations with LDL-C were then compared with the results reported by GLGC.

4.1.1 Genetic contributions to LDL-C and relation to Apo-B100

LDL-C is an established modifiable risk factor for CVD. Although it has been estimated that genetics explains up to 90% of the variability in plasma LDL-C levels[111], known genetic variants explain only a small proportion of genetic contribution to an individual's LDL-C levels[112]. Previous GWASs[45,110,113] have shown that more than 50 loci are associated with levels of LDL-C. It was estimated that, combined, known genetic variants explain 19.5% of the total variance in LDL-C[114]. Genetic variants in the genes that encode the targets of pharmacological modulation of LDL-C levels, such as *HMGCR* and *NPC1L1*, contribute to this variation in LDL-C.

Apo B-containing lipoproteins are a related determinant of atherosclerosis. Apo B is produced in two main isoforms, namely Apo B-48 and Apo B-100. Apo B-48 is unique to the production of chylomicrons and chylomicron remnants, whereas Apo B-100 is present in all very low-density lipoprotein (VLDL), intermediate-density lipoprotein

(IDL), LDL, and Lp(a) particles.[115] Several studies suggest that Apo B-100 may be a better predictor of cardiovascular risk than LDL-C, as it represents the combined effects of all atherogenic lipoproteins on arteries' walls[116-118].

LDL-C levels are highly correlated with Apo B-100. Each LDL particle contains one molecule of Apo B-100; therefore, changes in LDL-C concentration are generally proportional to changes in Apo B-100 concentration[116]. However, these changes are not equal in magnitude, as LDL is one of several lipids containing one molecule of Apo B-100. For example, an analysis of 11 RCTs of lipid-modifying therapy including 17,035 participants, demonstrated that the moderate-dose statin therapy decreased average LDL-C level by 42.1% and Apo B-100 by 33.1% from the baseline levels (i.e. the decrease in Apo B-100 constituted 79% of the decrease in LDL-C)[115].

4.1.2 Genetic risk score as an instrumental variable

The associations between genetic variants included in the *HMGCR* and *NPC1L1* GRSs, and LDL-C, are important for the following analyses examining the effects of these GRSs on the risk of T2DM, and effects on glycaemic traits and anthropometric measures. As both GRSs are used as instrumental variables for the 2-sample single protein MR analysis (see Chapter 3.5. Genetic methods), the GRS-exposure associations must be sufficiently well powered to avoid bias from a weak instrument. A weak instrument is defined as an instrumental variable which explains a small proportion of the variance in the exposure and is assessed with the F statistics in the regression of the exposure on the instrumental variable[70]. The associations of the *HMGCR* and *NPC1L1* GRSs with LDL-C were estimated using data from GLGC[45]. However, confirmation of the associations between genetic variants and levels of LDL-

C in the UK Biobank dataset would provide an additional validation for the GRSs. Moreover, GLGC has never analysed data on genetic associations with Apo B-100, whereas UK Biobank provides both genetic information and Apo B-100 concentrations for the full cohort.

The aims of this chapter were (1) to assess the associations of the genetic variants in *HMGCR* and *NPC1L1* loci on LDL-C concentrations in UK Biobank and to compare UK Biobank estimates with the estimates provided by GLGC; (2) to assess the associations of the selected genetic variants in *HMGCR* and *NPC1L1* loci on Apo B-100 concentrations.

4.2 Study design and method

4.2.1 Study design

The current project used principles of 2-sample single protein MR analysis to assess potential adverse effects of lipid modifying drug targets on the risk of T2DM and effects on related glycaemic and anthropometric measures. This method allowed a GRS based on a particular target gene to act as a proxy of the pharmacological intervention (e.g. statins or ezetimibe) and assessed the effects of 1 SD lower genetically predicted LDL-C on the risk of T2DM, and on HbA1c, random glucose and measures of central and general adiposity. The main assumption of this approach was that the GRS is a genetic proxy of lipid lowering medication and that it therefore exerted its action on lipid fractions through the same biological pathway as the relevant drug. As 2-sample single protein MR principles were used, the SNP-exposure estimates were derived from an independent dataset to avoid bias of the SNP-exposure effect.

GLGC reports per-allele effects on LDL-C in SD units but does not state its exact value in mmol/L[45]. To estimate the equivalent LDL-C value in mmol/L, a weighted average of the SD of the mean LDL-C was calculated and weighted by the number of participants in each of the studies included in the GWAS. The mean LDL-C (mmol/L) and the number of participants in each study is shown in the Appendix 4.2. Similar to the other projects[119,120], the value of 1 SD of LDL-C in GLGC was estimated as equal to 1 mmol/L.

4.2.2 Study population

UK Biobank participants of white British ancestry were included in this analysis. Participants with no available measures of LDL-C or Apo B-100, or with levels outside the reportable range of 0.26-10.3 mmol/L for LDL-C and 0.4-2 g/L for Apo B-100, were excluded from the analysis[98].

4.2.3 Correction for statin use

UK Biobank participants treated with statin therapy have lower levels of LDL-C and Apo B-100 than their untreated levels would be. To account for this, the levels of these measures were corrected in those participants who reported the use of statins. The full list of statins and UK Biobank medication codes is shown in Appendix 4.1. The correction factor for LDL-C was based on the average effect of statins as reported by Cholesterol Treatment Trialists' Collaborators in the meta-analysis of 14 RCTs of statins, including 90,056 participants[121]. Analogous to clinical practice, the CTT study included compliant and non-compliant participants. The weighted average reduction in LDL-C levels at one year of therapy was 29%, therefore the baseline LDL-C levels of UK Biobank participants who reported use of statins were divided by 0.7 to provide an estimate of their untreated LDL-C levels. This method of correction for the statin treatment effect was previously used by Peloso *et al*.[122]

As Apo B-100 is an integral part of each LDL particle, the levels of Apo B-100 in statin-users were adjusted using a similar approach. The correction factor for Apo B-100 was based on data from the run-in phase of the Heart Protection Study, which examined the changes in LDL-C and Apo B-100 in response to 40 mg simvastatin treatment in compliant participants[123]. The study included 18,705 high-risk participants aged 40-80

who had blood total cholesterol of ≥3.5 mmol/L. The proportional responses to simvastatin 40 mg were reported as the changes in $\log_e$ levels of LDL-C and Apo B-100 before commencing statin and following 4-6 weeks of therapy. There were proportional reductions in LDL-C and Apo B-100 (the study average reduction in LDL-C and Apo B-100 were -1.39 mmol/L (approximately -42.4%) and -0.37 g/L (approximately -32.8%) respectively)[62]. To account for a proportional difference in the statin effect on Apo B-100 and LDL-C, the measured concentration of Apo B-100 in UK Biobank was divided by a correction factor of 0.8167 if a participant reported the use of statins.

4.2.4 Statistical methods

Strengths of the associations between index SNPs in *HMGCR* and *NPC1L1* and concentrations of LDL-C and Apo B-100 were examined using linear regression adjusted for sex, array and 40 PCs to account for population stratification. The validity of linear regression tests between genetic variants and a trait depends on the trait distribution; any departure from normality in phenotype distribution inflates the risk of Type I error and reduces statistical power. To avoid false positive results from violation of the normality assumption, a two-stage method is widely used in genotype-phenotype association studies[124]. At the first stage a trait is regressed on covariates (intercept, age, age^2, and sex) to obtain residuals. Then, residuals are ranked and converted into z-scores using an inverse-normal transformation. The transformation produces an asymptotically normal distribution of any continuous trait. At the second stage, rank-normalised residuals are regressed on genetic variants with secondary adjustments (sex, genetic PCs, genotyping array)[125].

The linear regression model equation included the following data: for each of n participants there was a continuous LDL-C trait Y_i, genotype g_i, and a $p \times 1$ vector $x_i=(x_{i1},....,x_{ip})$ of covariates (intercept, age, age^2, and sex). Let $y= (Y_1,....,Y_n)$ denote the $n \times 1$ sample phenotype vector, g the $p \times 1$ genotype vector, and X the $n \times p$ covariate design matrix. The untransformed linear regression model:

$$y = X\beta_X + g\beta_G + \boldsymbol{\epsilon}, \qquad (1)$$

where $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\boldsymbol{I})$ is an $n \times 1$ residual vector, β_G is the genetic effect, and $\boldsymbol{\beta}_X$ is the covariate effect.

The direct inverse transformation model (2) is directly comparable to the untransformed model (1), with $\boldsymbol{y}$ being replaced by $\boldsymbol{z}$, which is the inverse normal transformed phenotype:

$$z = X\beta_X + g\beta_G + \boldsymbol{\epsilon}, \qquad (2)$$

The effects of GRSs on LDL-C and Apo B-100 were estimated by regressing lipid levels corrected for statins on the weighted GRS.

The validation analysis was carried out for the effects of genetic variants on LDL-C concentration only. There was no data available for the association between genetic variants and Apo B-100. In the validation analysis, index SNPs' effects on LDL-C were

extracted from GLGC[45]. The strengths of the association between GRS and LDL-C in the summary level data was estimated by F-statistics as:

$$Mean(F) = \frac{1}{n}\sum_{i=1}^{n}\frac{(b_n)^2}{(se_{\mathrm{n}})^2} \qquad (3)$$

where b_x and $se_{x,}$ are the effect size and standard error on the exposure (LDL-C).

4.3 Results

4.3.1 Lipid distributions in UK Biobank population

There were 316,731 participants with directly measured LDL-C, including 55,534 participants on statin therapy, and 261,197 not on statins. The mean level of LDL-C in on-statin participants was 24.3% lower than in those who were not on statins (2.83 mmol/L vs 3.74 mmol/L respectively) before the correction for statin use. The distributions of LDL-C levels in participants on-statin and not on-statin are shown in **Figure 4:1.**

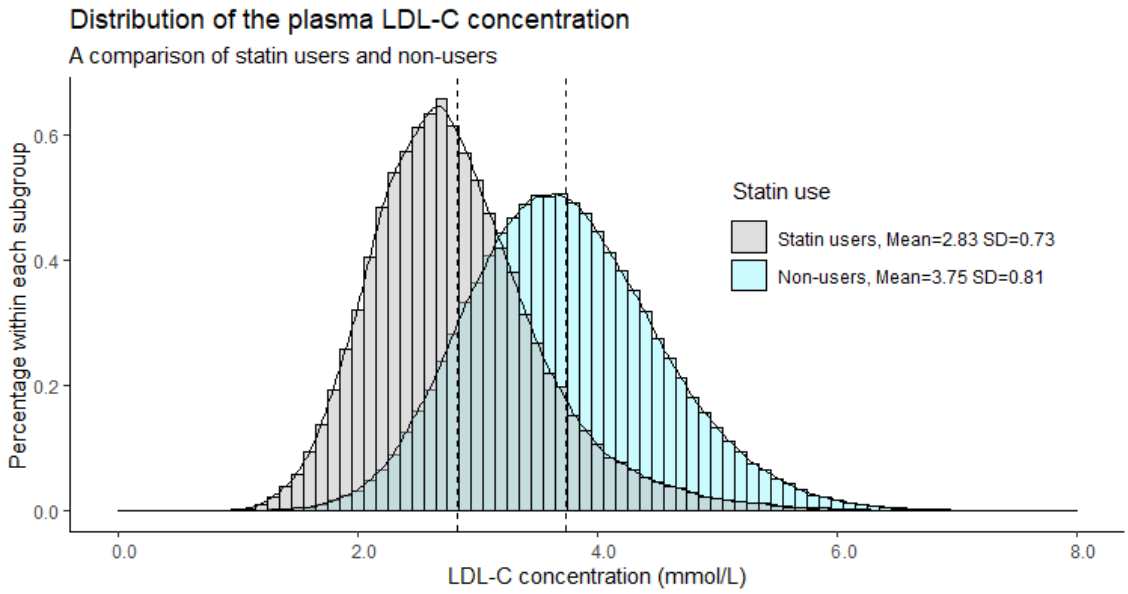


Figure 4:1 | Comparison of plasma LDL-C distribution in statin users and non-users
The number of participants with directly measured LDL-C is 316,731, including 55,534 participants on-statin, and 261,197 not on-statin. Bin width is scaled to 0.1 mmol/L, dotted lines represent the means

After correction for statin use, the overall mean LDL-C level increased to 3.79 mmol/L from the uncorrected level of 3.58 mmol/L) (**Table 4:1** and **Figure 4:2**). Appendix 4.2 provides further details of LDL-C levels before and after correction.

Table 4:1 | Mean levels of LDL-C and Apo B-100 in UK Biobank participant on-statin, not on-statin and overall corrected by statin use

	Off-statin Mean ± SD (N)	On-statin Mean ± SD (N)	Overall (corrected for statin use) Mean ± SD (N)
LDL-C (mmol/L)	3.74 ± 0.81 (261,197)	2.83 ± 0.73 (55,534)	3.79 ± 0.86 (316,731)
Apo B-100 (g/L)	1.07 ± 0.23 (261,727)	0.88 ± 0.21 (55,632)	1.07 ± 0.23 (317,359)

Data are presented as mean ± standard deviation (number of participants in each group). The correction factor for LDL-C was 0.7, for Apo B-100 it was 0.8167

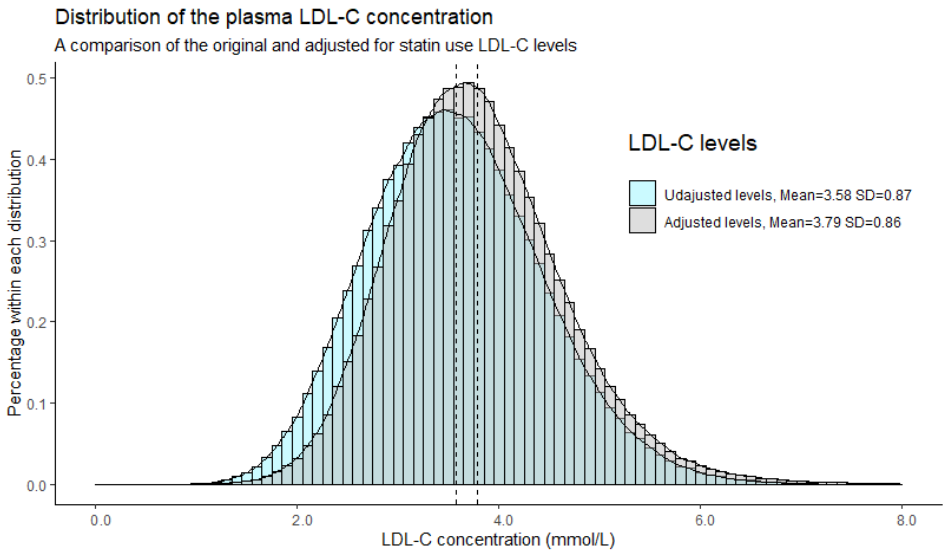


Figure 4:2 | Overall plasma LDL-C distribution prior and post correction for statin use

The overall number of participants with directly measured LDL-C is 316,731, including 55,534 participants on statin, whose raw values of LDL-C were corrected by 0.7. Bin width was scaled to 0.1 mmol/L, dotted lines represent the means

There were 317,359 participants with measured Apo B-100, of whom 55,632 were on-statin, and 261,727 were not on-statin (**Figure 4:1**, **Figure 4:3**). The mean level of Apo B-100 in treated participants was 18.1% lower than in those who were not on-statins (0.88 g/L vs 1.07 g/L respectively) (**Table 4:1**). After correction for statin use, the mean Apo B-100 for all participants increased to 1.07 g/L from the uncorrected level of 1.04 g/L (**Table 4:1** and **Figure 4:4**). Appendix 4.2 provides further details of Apo B-100 levels before and after correction.

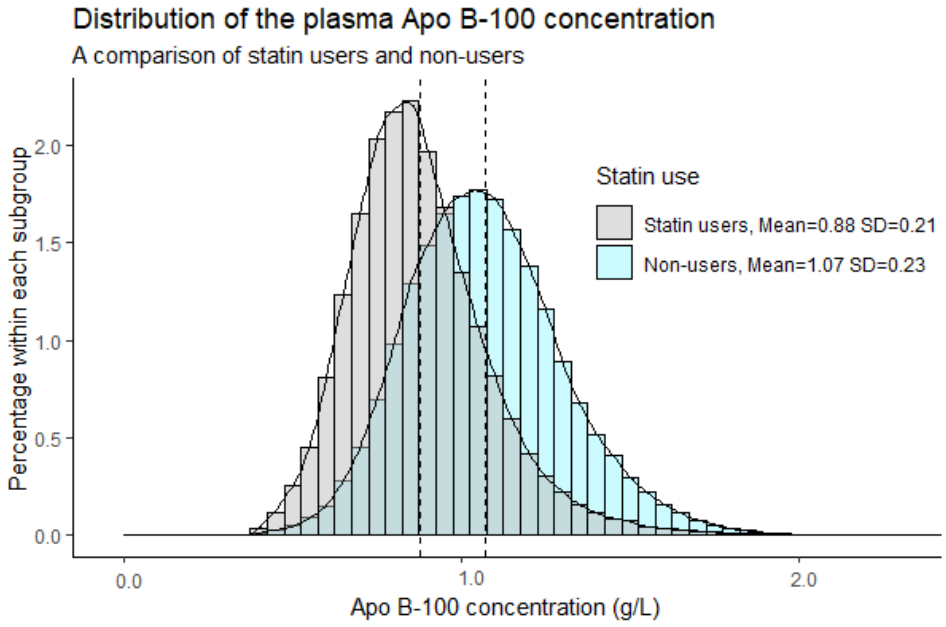


Figure 4:3 | Comparison of plasma Apo B-100 distribution in statin users and non-users

The overall number of participants with directly measured Apo B-100 is 317,359, including 55,632 participants on-statin, and 261,727 not on-statin. Bin width is scaled to 0.05 g/L, dotted lines represent the means.

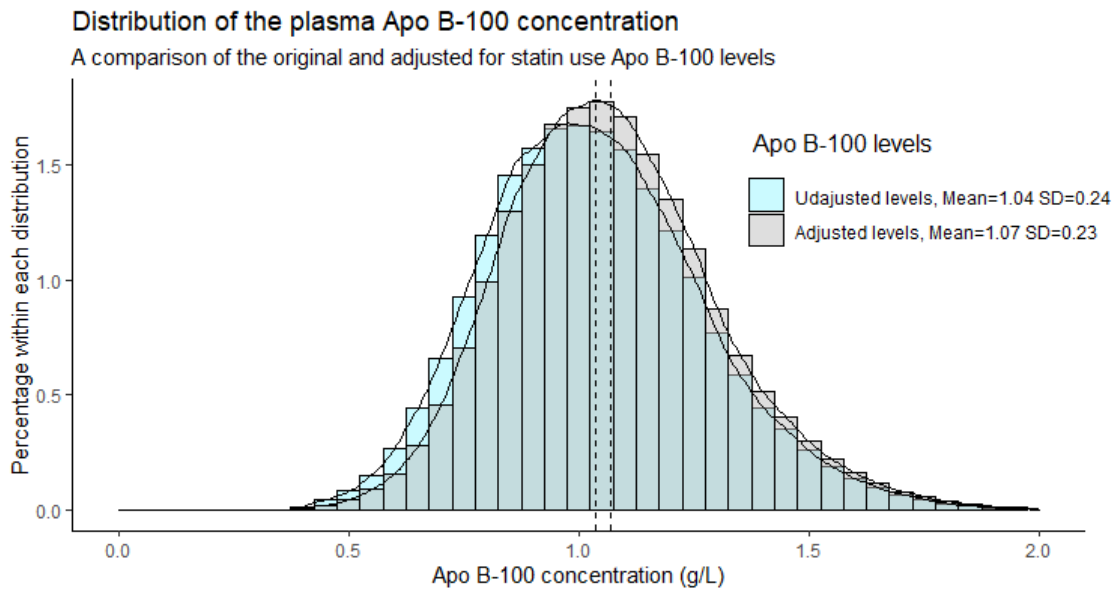


Figure 4:4 | Overall plasma Apo B-100 distribution prior and post correction for statin use

The overall number of participants with directly measured Apo B-100 is 317,359, including 55,632 participants on-statin, whose raw values of Apo B-100 were corrected by 0.8167. Bin width was scaled to 0.05 g/L, dotted lines represent the means.

4.3.2 The associations of *HMGCR* index SNPs with LDL-C concentrations

Associations of the selected genetic variants with LDL-C in UK Biobank are reported with respect to the LDL-C lowering allele in GLGC. The associations of the five *HMGCR* index SNPs with LDL-C levels were negative and directionally concordant (**Table 4:2)**.

Those associations were compared with the associations reported by GLGC. The results were directionally concordant and quantitatively consistent in general. The effect sizes of three SNPs (rs12916, rs12173076, and rs10515198) were slightly larger in UK Biobank, than in GLGC **(Figure 4:7, Figure 4:5).**

Table 4:2 | Associations of the *HMGCR* genetic variants with LDL-C in UK Biobank and GLGC

RSID	Effect allele / other allele	UK Biobank			GLGC		
		Effect (95% CI), SD	Sample size	Effect allele frequency	Effect (95% CI), SD	Sample size	Effect allele frequency
rs12916	T /C	-0.084 (-0.089, -0.079)	316,731	0.599	-0.073 (-0.081, -0.066)	173,026	0.569
rs12173076	T /G	-0.076 (-0.083, -0.068)	316,731	0.887	-0.065 (-0.076, -0.054)	173,012	0.877
rs10515198	G /A	-0.067 (-0.075, -0.058)	316,731	0.906	-0.060 (-0.072, -0.048)	172,939	0.897
rs3857388	T /C	-0.032 (-0.040, -0.024)	316,731	0.887	-0.042 (-0.054, -0.031)	89,724	0.872
rs7711235	A /G	-0.029 (-0.035, -0.023)	316,731	0.771	-0.038 (-0.050, -0.025)	168,357	0.732

The effect allele is an LDL-C lowering allele. Effect represents the change in LDL-Cl levels in SD units per copy of the effect allele. 1 SD unit in GLGC is equivalent to 1.00 mmol/L, in UK Biobank after correction for statins 1 SD was 0.85 mmol/L.

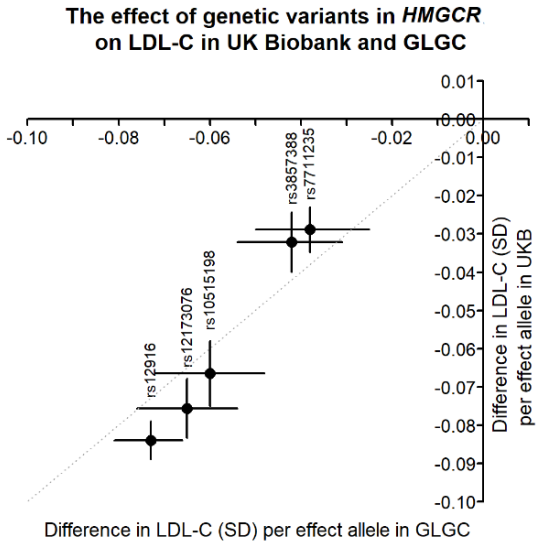


Figure 4:5 | Associations of genetic variants with LDL-C in the UK Biobank and GLGC

The sample size of the UK Biobank dataset was 316,731, the sample size of the GLGC dataset varied between the SNPs, with the smallest number for rs12916 and highest for rs12173076 (89,724 and 173,026 respectively). 1 SD unit in GLGC is equivalent to 1.00 mmol/L, in UK Biobank after correction for statins 1 SD was 0.85 mmol/L.

4.3.3 The association of *HMGCR* GRS with LDL-C concentration

When combined in a *HMGCR* GRS, 1 SD lower genetically predicted LDL-C was associated with a -0.607 SD (95% CI -0.64, -0.57) LDL-C concentration (**Figure 4:6**). Several methodological factors may explain the difference between genetically estimated in GLGC and measured in the UK Biobank levels of LDL-C. Firstly, the proportion of participants on lipid-lowering medication was based on self-reported data, which may need to be more accurate. Secondly, the Drug Target MR assumes a linear genetic effect, but it might not be. Therefore, a small degree of heterogeneity between the two studies was expected.

HMGCR GRS explained 0.36% of the variance in LDL-C in the UK Biobank and 0.60% of the variance in GLGC. The F-values were 1028 and 494 in the UK Biobank and GLGC respectively, indicating that in both datasets the *HMGCR* GRS is reliably associated with LDL-C levels.

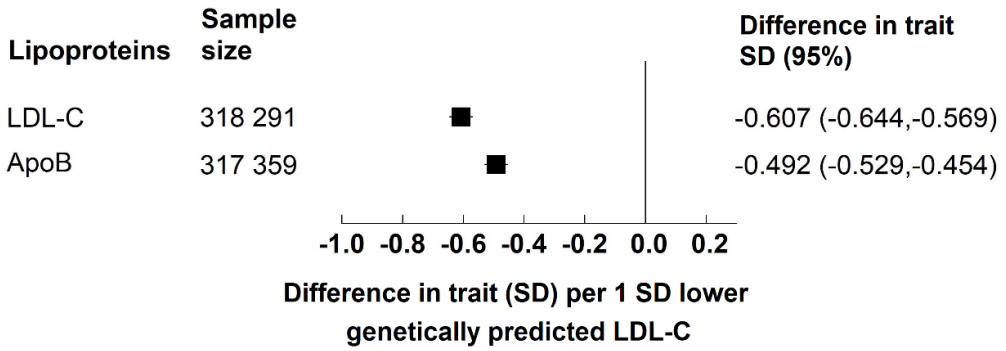


Figure 4:6 | Associations of *HMGCR* GRS with blood lipids
The associations of GRS are shown in SD units per 1 SD lower genetically predicted LDL-C, directly measured LDL-C values were available for 318,291 participants, Apo B-100 values for 317,359 participants of the UK Biobank.

4.3.4 The associations of *NPC1L1* index SNPs with LDL-C concentrations

The associations of the *NPC1L1* index SNPs with LDL-C were negative and directionally concordant (**Table 4:3**). Compared to the genetic associations reported by GLGC 2013[45], the associations observed in the UK Biobank were directionally concordant and quantitatively consistent (**Figure 4:7**).

Table 4:3 | Associations of the *NPC1L1* genetic variants with LDL-C in UK Biobank and GLGC

RSID	Effect allele / other allele	UK Biobank			GLGC		
		Effect (95% CI), SD	Sample size	Effect allele frequency	Effect (95% CI), SD	Sample size	Effect allele frequency
rs2073547	A/G	-0.052 (-0.058, -0.045)	316,731	0.817	-0.049 (-0.058, -0.039)	169,889	0.806
rs217386	A/G	-0.033 (-0.038, -0.028)	316,731	0.424	-0.036 (-0.044, -0.029)	173,021	0.408

The effect allele is an LDL-C lowering allele. Effect represents the change in LDL-Cl levels in SD units per copy of the effect allele. 1 SD unit in GLGC is equivalent to 1.00 mmol/L, in UK Biobank after correction for statins 1 SD was 0.85 mmol/L.

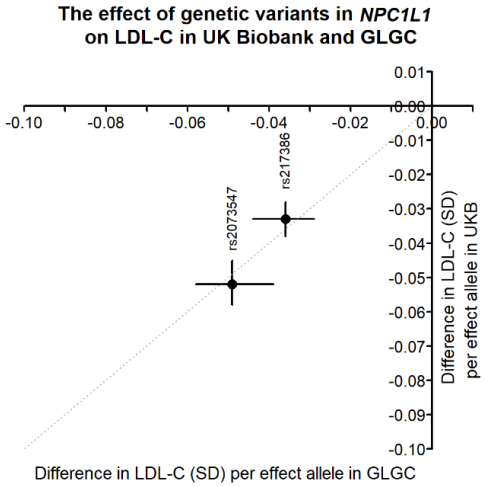


Figure 4:7 | Associations of genetic variants in *NPC1L1* GRS with LDL-C
The sample size of the UK Biobank dataset was 317,359, the sample size of the GLGC dataset was 169,889 for rs2073547, and 173,021 for rs217386. 1 SD unit in GLGC is equivalent to 1.00 mmol/L, in UK Biobank after correction for statins 1 SD is 0.85 mmol/L.

4.3.5 The associations of *NPC1L1* GRS with LDL-C concentration

When combined in the *NPC1L1* GRS, 1 SD lower genetically predicted LDL-C was associated with a -0.72 SD (95% CI -0.80, -0.64) LDL-C concentration (**Figure 4:8**). *NPC1L1* GRS explained 0.14% of the variance in LDL-C in the UK Biobank and GLGC.

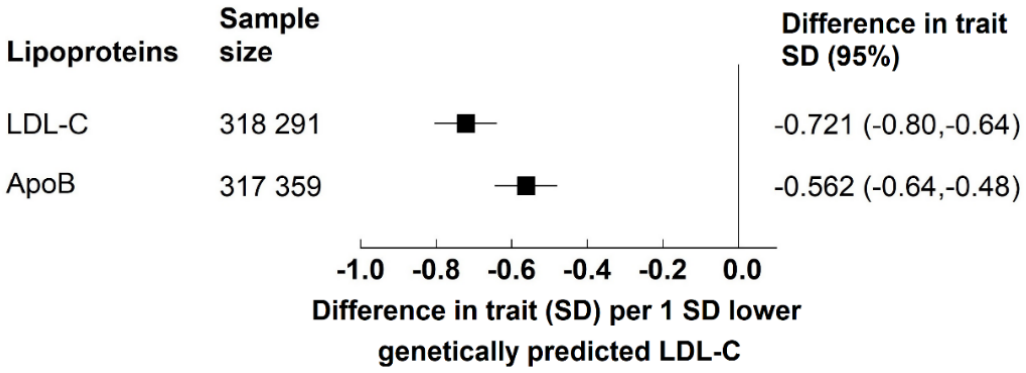


Figure 4:8 | Associations of *NPC1L1* GRS with blood lipids
The associations of GRS are shown in SD units per 1 SD lower genetically predicted LDL-C, directly measured LDL-C values were available for 318,291 participants, Apo B-100 values for 317,359 participants of the UK Biobank.

4.3.6 The associations of *HMGCR* and *NPC1L1* GRSs with Apo B-100 concentrations

The associations of *HMGCR* and *NPC1L1* LDL-C lowering alleles with Apo B-100 concentration were directionally concordant and quantitatively consistent (**Table 4:4**). *HMGCR* GRS was associated with -0.492 SD (95% CI -0.53, -0.45) Apo B-100 per 1 SD genetically predicted LDL-C. *NPC1L1* GRS was associated with -0.562 SD (95% CI -0.64, -0.48) Apo B-100 per 1 SD lower genetically predicted LDL-C). **Figure 4:9** and **Figure 4:10** shows the associations of the *HMGCR* and *NPC1L1* genetic variants with LDL-C and Apo B-100 in SD units.

Table 4:4 | Associations of the *HMGCR* and *NPC1L1* genetic variants with Apo B-100 in UK Biobank

	UK Biobank			
RSID	**Effect allele / other allele**	**Effect (95% CI), SD**	**Sample size**	**Effect allele frequency**
HMGCR				
rs12916	T/C	-0.068 (-0.073, -0.063)	316,731	0.599
rs12173076	T/G	-0.061 (-0.068, -0.053)	316,731	0.887
rs10515198	G/A	-0.055 (-0.063, -0.046)	316,731	0.906
rs3857388	T/C	-0.025 (-0.032, -0.017)	316,731	0.887
rs7711235	A/G	-0.022 (-0.028, -0.016)	316,731	0.772
NPC1L1				
rs2073547	A/G	-0.041 (-0.048, -0.035)	316,731	0.817
rs217386	A/G	-0.024 (-0.029, -0.019)	316,731	0.424

The effect allele is an Apo B-100 lowering allele. Effect represents the change in Apo B-100 levels in SD units per 1 copy of the effect allele. 1 SD unit for Apo B-100 is 0.24 g/L.

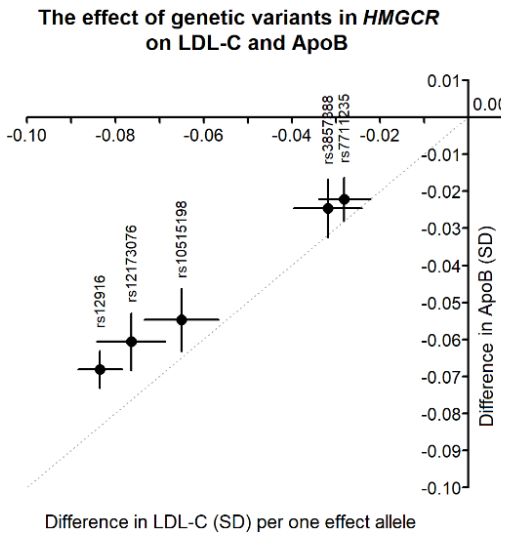


Figure 4:9 | Associations of genetic variants in *HMGCR* with LDL-C (SD) and Apo B-100 (SD) concentrations

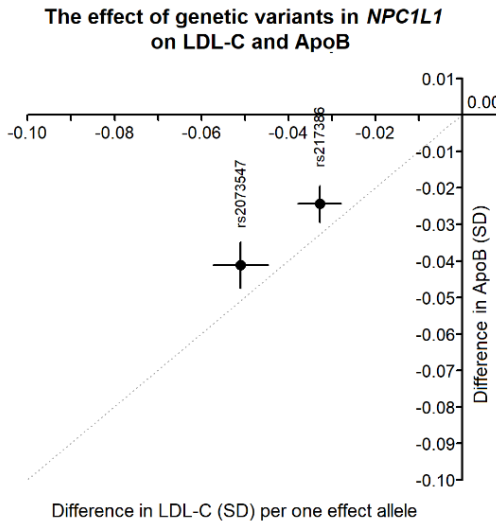


Figure 4:10 | Associations of genetic variants in *NPC1L1* with LDL-C (SD) and Apo B-100 (SD) concentrations

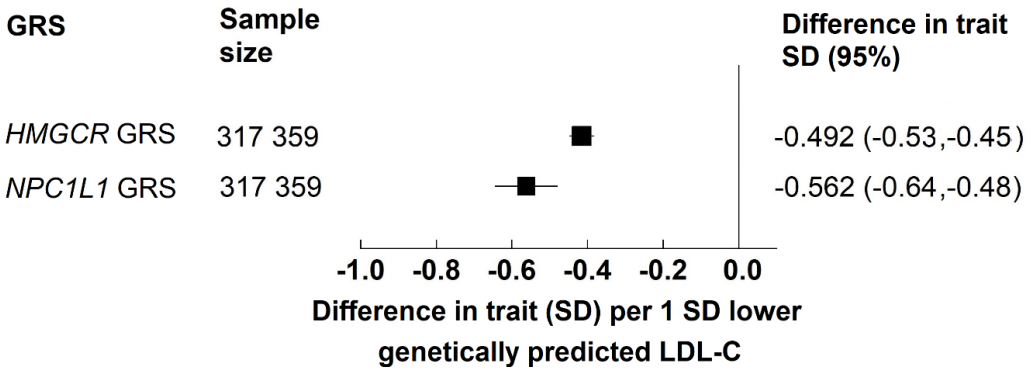


Figure 4:11 | Associations of HMGCR GRS and *NPC1L1* GRS with Apo B-100
The associations of GRS are shown in SD units per 1 SD lower genetically predicted LDL-C, Apo B-100 values were available for 317,359 participants of the UK Biobank.

4.4 Discussion

This chapter assessed the concordance of the effects of locus-specific index SNPs on LDL-C in two samples of Caucasian populations. GLGC is the largest GWAS of lipids, and the effect sizes reported by this consortium have been widely used in recent genetic studies of the effects of LDL-C on various health outcomes. The associations of uncorrelated SNPs in *HMGCR* and *NPC1L1* loci with LDL-C concentrations, as reported by GLGC, were compared with those in the UK Biobank study. There was no material difference in the population structure of both samples; the minor allele frequencies in UK Biobank and GLGC were very similar. UK Biobank was used as a confirmation sample, and an identical statistical model was applied to replicate the association of the same genetic variants with LDL-C, as have been done in the original sample (GLGC). Overall, the effect sizes estimated in UK Biobank were similar to the estimates in GLGC. The directional consistency of the effect of genetic variants on LDL-C observed in both samples implies that the SNPs selected for GRSs are reproducibly associated with the phenotype. Therefore, a GRS weighted on the effect on LDL-C, as reported by the GLGC study, is an appropriate instrumental variable for a single protein risk factor MR analysis.

There were some small differences observed between the two samples. 1 SD unit for LDL-C in GLGC was equal to 0.85 SD in the UK Biobank. The smaller SD values in the UK Biobank could be explained by a more homogeneous sample.

HMGCR GRS explained 0.36% of the variance in LDL-C in the UK Biobank, half the magnitude of the variance explained by the same SNPs in GLGC. *NPC1L1* GRS explained 0.14% of the variance in both datasets. Although both GRSs explain only a small proportion of variance in LDL-C, these associations were statistically significant

and were confirmed in the second sample, and previous studies using comparable GRSs have demonstrated effects of *HMGCR* and *NPC1L1* GRSs on cardiovascular outcomes [30,51,66,113].

The second objective of this chapter was to assess the effects of the index genetic variants in *HMGCR* and *NPC1L1* loci on Apo B-100 concentration. As expected, their effects on Apo B-100 were proportional to their effects on LDL-C. Each additional LDL-C lowering allele had an effect on Apo B-100 of a magnitude approximately 2/3 SD of the effect on LDL-C. There were no data on Apo B-100 available in GLGC.

There were no previous genetic studies reporting the effect estimates of the index *HMGCR* and *NPC1L1* genetic variants on Apo B-100, however several RCTs have reported the effects of statin therapy on LDL-C and Apo B-100. The Heart Protection Study[123], with 18,705 participants taking 40 mg of simvastatin for 4-6 weeks, reported a 42.5% reduction in LDL-C and a 32.8% reduction in Apo B-100 levels.

A meta-analysis of 9 RCTs of various doses and types of statins and 2 RCTs of combination therapy (statins a plus ezetimibe) with 17,035 participants showed an average on-treatment decrease in LDL-C concentration of 42.1% and in Apo B-100 concentration of 33.1%[115]. The study showed consistency in the proportional decreases in both lipids among RCTs with different doses of statins, sample sizes and untreated levels of lipids. In each RCT, the Apo B-100 concentration decreased less in relative terms than LDL-C. The comparison of the reduction in LDL-C and Apo B-100 in 11 RCT showed a linear relationship, with the reduction of Apo B-100 approximately 80% of the reduction in LDL-C[115].

The comparison of the effects of statins on LDL-C and Apo B-100 in the RCTs with the results of genetic analyses shows a similar proportional reduction in both

metabolites. The magnitude of the effects of *HMGCR* GRS and *NPC1L1* GRS on Apo B-100 was 81% and 78% of their effects on LDL-C, respectively.

Chapter 5 The associations of the *HMGCR* GRS with the risk of Type 2 Diabetes Mellitus, glycaemic traits, and anthropometric measures

5.1 Introduction

Genetic studies have used the principles of mendelian randomisation to examine a causal association of the LDL-C lowering alleles in *HMGCR*, as proxies for the pharmacological action of statins, with T2DM[51,126,127]. As discussed in Chapter 2, few genetic studies[51,127] have assessed the associations of variants in *HMGCR* with the risk of T2DM. Their results were directionally concordant with the results of RCTs of statins and showed that LDL-C lowering alleles within *HMGCR* were associated with a higher risk of T2DM.

A study by Swerdlow *et al*.[126], reviewed in Chapter 2.3.1, showed that two genetic variants in *HMGCR* (rs17238484 and rs12916) were associated with higher risk of T2DM (OR 1.02 95% CI 1.00, 1.05 and OR 1.06 95% CI 1.03, 1.09 respectively). Each LDL-C lowering allele in rs17238484 was also associated with 0.30 kg (95% CI 0.18, 0.43) higher bodyweight and 0.32 cm (95% CI 0.16, 0.47) larger waist circumference[126]. The authors concluded that the effect of *HMGCR* on the risk of T2DM can be partially explained by an increase in central adiposity. However, this conclusion was drawn from a meta-analysis of studies with varying definitions of T2DM. As T2DM was not specified as a primary outcome in many studies contributing individual level data, the definition of T2DM ranged from self-reported to biochemically validated or derived from medical records. Some studies did not distinguish between

type 1 and type 2 diabetes. Overall, there were 14,976 cases of T2DM available for rs12916 and 26,236 cases for rs17238484. Anthropometric measures (body weight and waist circumference) were available for 143,113 and 69,163 participants for rs17238484 and 65,871 and 39,591 participants for rs12916. The genetic data came from ethnically diverse population, including participants from European and North American regions.

A major advantage of using UK Biobank to elucidate the effect of *HMGCR* GRS on T2DM is the availability of medical data captured at the baseline assessment for a consistent outcome definition. To identify participants with T2DM and distinguish between type 1 and type 2 diabetes, we supplemented the algorithmic definition of T2DM developed by Eastwood *et al.*[104] (see Chapter 3.3 for details) with additional information from hospital admission data. The second advantage of using UK Biobank is the availability of extensive and detailed measurements, including physical measurements by bioimpedance, genetic data, and blood biomarkers of glycaemia. These data overcome some limitations of the previous studies, providing coherent and reliable data on T2DM, anthropometric measures, and glycaemic measurements, which have not been previously used for the analysis of the effect of *HMGCR* GRS on risk of T2DM.

There is increasing evidence that sex is an important factor in the pathophysiology of many diseases, including CVD[128] and T2DM[129-132]. Previous studies have shown that the age of onset of T2DM in men is younger than in women[132].

Men are also diagnosed with T2DM at lower levels of BMI due to greater amounts of trunk and visceral fat compared with premenopausal women with greater gluteofemoral fat distribution[131]. Certain biological differences in body composition and

glucose and fat metabolism underline sex-dimorphic mechanisms in T2DM aetiology[130]. To account for a potential sex-dimorphic effects of GRSs on adiposity and therefore on the risk of T2DM, sex-stratified associations with anthropometric measures, and the risks of T2DM were included in the analyses.

The aims of this chapter are, firstly, to investigate the association between genetic variation in the *HMGCR* gene and the risk of overall and incident T2DM and, secondly, to examine the effect of *HMGCR* GRS on glycaemic traits and measures of central and general adiposity in men and women.

5.2 Study design and method

5.2.1 Study population

Genetic analyses were conducted among 334,516 unrelated white British UK Biobank participants to examine the effect of *HMGCR* GRS on total T2DM, glycaemic traits, and anthropometric measures. A subset of 330,622 participants free from T2DM at baseline was used to assess the effect of *HMGCR* GRS on incident T2DM (see Chapter 3.3 for more details).

5.2.2 Baseline assessment

Extensive information on participant characteristics, medical history and medications was collected at the baseline assessment via a touch-screen questionnaire and a nurse-led interview (see Chapter 0 for more details).

Two blood biomarkers of glycaemia (HbA1c and non-fasting plasma glucose) were measured in samples taken at baseline in all 334,516 participants, however those who reported the use of glucose-lowering medication or had values outside of the reportable range were excluded from the analysis of the effect of *HMGCR* GRS on glycaemic traits (see Chapter 3.2.4 for more details). After these exclusions, HbA1c levels were available for 309,450 participants and non-fasting glucose levels for 283,305 participants. The UK Biobank blood biochemistry protocol is described in detail in Chapter 3.2.4.

Measures of central and overall adiposity were recorded during the baseline assessments. Measures of adiposity were measured, including body weight (kg), standing height (cm), waist circumference (cm), hip circumference (cm), waist-to-hip ratio, BMI (kg/m^2), body fat percentage (%), lean body mass (kg). The UK Biobank anthropometric measurement protocol is described in detail in Chapter 3.2.3

5.2.3 Total and incident T2DM

Among the 334,516 individuals included in the current study, there were 20,361 cases of total T2DM, including 6,467 cases of incident T2DM and 13,894 cases of prevalent T2DM. The algorithmic definition of T2DM is described in detail in Chapter 3.3.

5.2.4 Statistical methods

Statistical methods were described in Chapter 3.6.1. *HMGCR* GRS was used as a genetic proxy of statins. The associations of *HMGCR* with total and incident T2DM were assessed with a logistic regression model, and associations with glycaemic traits (HbA1c and non-fasting glucose) and anthropometric measures (body weight, height, waist and hip circumferences, waist-to-hip ratio, BMI, body fat percentage and body lean mass) were assessed with linear regression models and reported in SD units.

5.3 Results

5.3.1 Baseline characteristics of the study population

The baseline characteristics of UKB participants are summarised in **Table 5:1**. Due to sex differences in body anthropometry and composition, baseline characteristics were stratified by sex.

Participants with total T2DM were more likely to be men (62.4%) than women (37.6%). They were older (the mean age (SD) at the baseline assessment was 60.4 (6.7) years compared to non-diabetic participants with mean age of 56.7 (8.0) years) and had a higher Townsend index score (a higher Townsend score indicates more deprived areas). All available anthropometric measures of obesity, except waist-to-hip ratio, were higher in participants with T2DM. The percentage of current non-smokers among diabetic participants was smaller (42.0%) than in non-diabetic participants (55.7%), whereas the percentage of heavy smokers was higher (4.3%) compared to non-diabetic participants (2.5%). Participants with T2DM had a higher prevalence of co-morbidities; 62.0% had treated hypertension and 20.3% had cardiovascular disease (self-reported at touch screen questionnaire or during nurse led interview). 67.8% of participants with T2DM were on statin therapy. In those without T2DM, 14.3% were using statins, 19.4% reported treated hypertension and 5.4% - prevalent cardiovascular disease. Other medications that affect lipid levels were not included in the baseline table, as they were not included in the QDiabetes-2018 risk calculator.

Table 5:1 | Baseline characteristics of 334,516 European UK Biobank participants by T2DM status at baseline assessment and sex

Risk factors	Non-diabetic participants n = 314 155 (93.9%)		Total T2DM n = 20 361 (6.1%)		Incident T2DM n = 6 467 (1.9%)	
	Women n=172 056 (54.8%)	Men n=142 099 (45.2%)	Women n=7 664 (37.6%)	Men n=12 697 (62.4%)	Women n=2 699 (41.7%)	Men n=3 768 (58.3%)
Patient characteristics:						
Age at the baseline, (yr)	56.5 ± 7.9	56.8 ± 8.1	60.1 ± 6.8	60.6 ± 6.6	59.6 ± 7.2	59.9 ± 7.1
Townsend score (points)	-1.7 ± 2.9	-1.6 ± 2.9	-0.7 ± 3.2	-0.8 ± 3.3	-0.7 ± 3.2	-0.9 ± 3.3
Anthropometric measures:						
Bodyweight (kg)	70.7 ± 13.8	85.1 ± 14.3	84.7 ± 19.0	95.3 ± 18.9	83.9 ± 18.5	95.4 ± 18.7
Height (cm)	162.8 ± 6.2	176.1 ± 6.7	161.3 ± 6.3	174.8 ± 6.8	161.5 ± 6.2	174.8 ± 6.9
Waist circumference (cm)	83.7 ± 12.4	96.0 ± 11.3	99.4 ± 15.7	106.4 ± 13.9	98.3 ± 15.3	106.2 ± 13.7
Hip circumference (cm)	102.8 ± 10.7	102.9 ± 8.3	111.7 ± 15.7	107.8 ± 11.1	111.4 ± 15.1	107.8 ± 10.9
Waist-to-hip ratio	0.8 ± 0.1	0.9 ± 0.1	0.9 ± 0.1	1.0 ± 0.1	0.9 ± 0.1	1.0 ± 0.1
BMI (kg/m^2)	26.7 ± 5.1	27.4 ± 4.2	32.5 ± 7.0	31.1 ± 5.8	32.1 ± 6.9	31.1 ± 5.6
Body fat percentage (%)	35.8 ± 8.1	24.4 ± 6.5	41.0 ± 9.5	28.5 ± 7.8	41.0 ± 9.4	28.8 ± 7.4
Body lean mass (kg)	44.4 ± 10.1	63.6 ± 8.3	48.2 ± 13.7	66.8 ± 11.3	47.9 ± 13.3	66.8 ± 10.9
Glycaemic traits:						
Non-fasting glucose (mmol/L)	5.0 ± 0.7	5.0 ± 0.7	6.8 ± 2.8	7.1 ± 3.1	6.0 ± 2.2	6.3 ± 2.8

Risk factors	Non-diabetic participants n = 314 155 (93.9%)		Total T2DM n = 20 361 (6.1%)		Incident T2DM n = 6 467 (1.9%)	
	Women n=172 056 (54.8%)	Men n=142 099 (45.2%)	Women n=7 664 (37.6%)	Men n=12 697 (62.4%)	Women n=2 699 (41.7%)	Men n=3 768 (58.3%)
HbA1c (mmol/mol)	35.0 ± 3.8	34.9 ± 4.4	48.5 ± 12.8	49.4 ± 13.4	43.9 ± 11.7	44.5 ± 12.2
Lipids levels:						
Direct LDL-C (mmol/L)	3.7 ± 0.9	3.6 ± 0.8	3.1 ± 0.9	2.8 ± 0.8	3.7 ± 1.0	3.3 ± 0.9
HDL-C (mmol/L)	1.6 ± 0.4	1.3 ± 0.3	1.3 ± 0.3	1.1 ± 0.3	1.3 ± 0.3	1.1 ± 0.3
Triglycerides (mmol/L)	1.5 ± 0.8	2.0 ± 1.1	2.2 ± 1.2	2.3 ± 1.3	2.3 ± 1.2	2.5 ± 1.4
Apo A-I (g/L)	1.6 ± 0.3	1.4 ± 0.2	1.5 ± 0.2	1.3 ± 0.2	1.5 ± 0.3	1.4 ± 0.2
Apo B-100 (g/L)	1.0 ± 0.2	1.0 ± 0.2	1.0 ± 0.3	0.9 ± 0.2	1.1 ± 0.3	1.0 ± 0.3
Prior disease:						
Family history of diabetes	34 505 (20.1%)	24 857 (17.5%)	3 259 (42.5%)	4 649 (36.6%)	961 (35.6%)	1 125 (29.9%)
Type 2 diabetes	0 (0.0%)	0 (0.0%)	7 664 (100%)	12 697 (100%)	0 (0.0%)	0 (0.0%)
Treated hypertension	28 955 (16.8%)	32 089 (22.6%)	4 454 (58.1%)	8 174 (64.4%)	1 181 (43.8%)	1 873 (49.7%)
Cardiovascular disease	5 740 (3.3%)	11 366 (8.0%)	1 066 (13.9%)	3 059 (24.1%)	315 (11.7%)	792 (21.0%)
Schizophrenia or bipolar affective disorder	544 (0.3%)	511 (0.4%)	61 (0.8%)	91 (0.7%)	31 (1.1%)	32 (0.8%)
Gestational diabetes	0 (0.0%)	2 (0.0%)	43 (0.6%)	3 (0.0%)	37 (1.4%)	2 (0.1%)
Polycystic ovary syndrome	344 (0.2%)	0 (0.0%)	38 (0.5%)	0 (0.0%)	17 (0.6%)	0 (0.0%)

Risk factors	Non-diabetic participants n = 314 155 (93.9%)		Total T2DM n = 20 361 (6.1%)		Incident T2DM n = 6 467 (1.9%)	
	Women n=172 056 (54.8%)	Men n=142 099 (45.2%)	Women n=7 664 (37.6%)	Men n=12 697 (62.4%)	Women n=2 699 (41.7%)	Men n=3 768 (58.3%)
Medication:						
Statins	17 683 (10.3%)	27 216 (19.2%)	4 802 (62.7%)	9 001 (70.9%)	873 (32.3%)	1 686 (44.7%)
Atypical antipsychotics	374 (0.2%)	382 (0.3%)	61 (0.8%)	78 (0.6%)	30 (1.1%)	24 (0.6%)
Corticosteroids	1747 (1.0%)	1435 (1.0%)	186 (2.4%)	214 (1.7%)	64 (2.4%)	76 (2.0%)
Smoking:						
Non-smoker	103 142 (59.9%)	71 793 (50.5%)	3 999 (52.2%)	4 555 (35.9%)	1 375 (50.9%)	1 366 (36.3%)
Former smoker	54 329 (31.6%)	53 737 (37.8%)	2 816 (36.7%)	6 528 (51.4%)	940 (34.8%)	1 823 (48.4%)
Light smoker	2 600 (1.5%)	1 525 (1.1%)	95 (1.2%)	106 (0.8%)	37 (1.4%)	40 (1.1%)
Moderate smoker	8 645 (5.0%)	10 434 (7.3%)	442 (5.8%)	945 (7.4%)	203 (7.5%)	309 (8.2%)
Heavy smoker	3 340 (1.9%)	4 610 (3.2%)	312 (4.1%)	563 (4.4%)	144 (5.3%)	230 (6.1%)

Percentage (%) was used for binary traits and +/- SD for continuous traits. Participants with incident T2DM were free from T2DM at baseline but developed the condition after attending the baseline assessment centre. The total T2DM group had T2DM diagnosed at any time point and includes participants with incident T2DM.

5.3.2 *HMGCR* GRS and risk of T2DM

5.3.2.1 Individual SNPs and risk of T2DM

The associations of the five index SNPs within the *HMGCR* gene with total T2DM were positive and directionally concordant (**Table 5:2, Figure 5:1**). One SNP's (rs12916) association individually reached p=0.05 level of statistical significance. The rs12916 variant, associated with 0.073 SD (95% CI 0.066, 0.081) lower LDL-C, was associated with 3.0% (OR 1.03, 95% CI 1.01, 1.05) higher risk of total T2DM, per copy of the LDL-C lowering allele (T). Another SNP (rs12173076) was marginally associated with T2DM. Each additional LDL-C lowering rs12173076 allele (T), associated with 0.065 mmol/L (95% CI 0.054, 0.076) lower LDL-C, was associated with a 2.6% (OR 1.026, 95% CI 0.99, 1.07) higher risk of total T2DM.

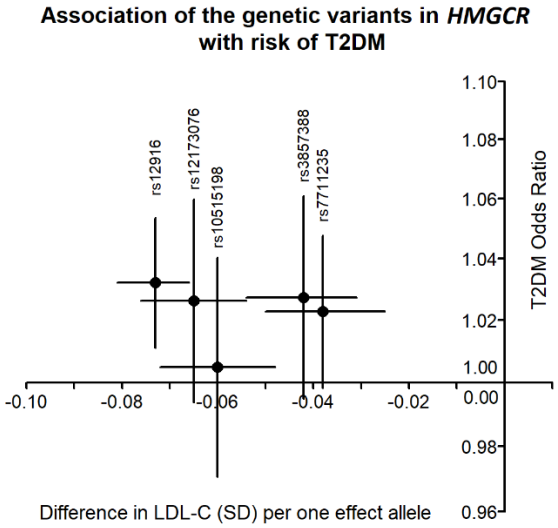


Figure 5:1 | Associations of the *HMGCR* index SNPs with the risk of total T2DM and LDL-C concentrations

The effect on total T2DM (OR, 95% CI) and on LDL-C (estimate, 95% CI) (SD) is shown per one copy of the LDL-C lowering allele, using the effect on LDL-C as reported by GLGC. Genotype data were available for 334,669 participants including 20,361 overall cases of T2DM and 314,155 controls.

Table 5:2 | Associations of the *HMGCR* index SNPs with total T2DM and LDL-C concentrations

RSID	Effect allele / other allele	Effect (95% CI) on LDL-C	Effect on T2DM OR (95% CI)
rs12916	T/C	-0.073 (-0.081, -0.066)	1.032 (1.011, 1.054)
rs12173076	T/G	-0.065 (-0.076, -0.054)	1.026 (0.994, 1.060)
rs10515198	G/A	-0.060 (-0.072, -0.048)	1.005 (0.971, 1.040)
rs3857388	T/C	-0.042 (-0.054, -0.031)	1.027 (0.999, 1.061)
rs7711235	A/G	-0.038 (-0.050, -0.025)	1.023 (0.998, 1.048)

The effects on T2DM and LDL-C are shown per copy of the LDL-C lowering allele, the effect on LDL-C in SD units was reported by GLGC, the effect on T2DM is an Odds Ratio derived from the UK Biobank.

5.3.2.2 *HMGCR* GRS and risk of T2DM

The *HMGCR* GRS was associated with a 26% (95% CI 1.08, 1.47, p= 0.003) higher risk of total T2DM per 1 SD lower genetically predicted LDL-C (**Figure 5:2**).

Among participants free from T2DM at baseline, the association between *HMGCR* and incident T2DM was consistent with the total T2DM result (OR 1.19, 95% CI 0.91, 1.55, p=0.203) though it was not statistically significant.

There was no heterogeneity in the effect of *HMGCR* on the risk of total T2DM between men and women (P_{het}= 0.592) (**Figure 5:2**).

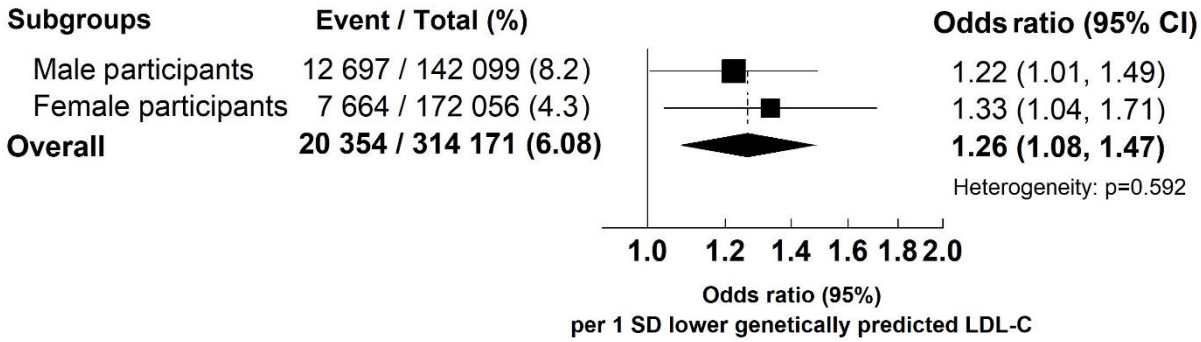


Figure 5:2 | Associations of *HMGCR* GRS with the risk of total T2DM stratified by sex

5.3.2.3 External validation analysis of the effect of *HMGCR* on T2DM

As an external validation analysis, the effect of *HMGCR* GRS on the risk of T2DM was also assessed in the DIAGRAM[47] consortium data set, which included 26,676 cases of T2DM. The SNP-by-SNP comparison is shown in the **Table 5:3.**

Table 5:3 | Associations of *HMGCR* genetic variants with risk of T2DM in the DIAGRAM consortium and UK Biobank

RSID	Effect allele / other allele	Risk of T2DM for the effect allele in DIAGRAM OR (95% CI)	Risk of T2DM for the effect allele in UK Biobank OR (95% CI)	LDL-C difference (from GLGC)
rs12916	T/C	1.013 (0.990, 1.037)	1.032 (1.011, 1.054)	-0.073 (-0.081, -0.066)
rs12173076	T/G	1.017 (0.980, 1.056)	1.026 (0.994, 1.060)	-0.065 (-0.076, -0.054)
rs10515198	G/A	1.006 (0.968, 1.047)	1.005 (0.971, 1.040)	-0.060 (-0.072, -0.048)
rs3857388	T/C	0.999 (0.961, 1.039)	1.027 (0.999, 1.061)	-0.042 (-0.054, -0.031)
rs7711235	A/G	1.002 (0.973, 1.032)	1.023 (0.998, 1.048)	-0.038 (-0.050, -0.025)

The effect is shown in OR (95% CI) units per 1 copy of the effect allele.

The effect of the *HMGCR* GRS in the DIAGRAM consortium on the risk of T2DM was consistent with the result from the UK Biobank study (OR 1.17, 95% CI 0.93, 1.48)

(**Figure 5:3**). A meta-analysis of the results from the current UK Biobank study and the DIAGRAM consortium showed that the *HMGCR* GRS was associated with 24% (95% CI 1.09, 1.41) higher risk of total T2DM per 1 SD lower genetically predicted LDL-C (**Figure 5:3**). There was no heterogeneity between the two studies (P_{het}=0.280, I^2=0%).

There was no material difference in the effect of *HMGCR* GRS on the risk of total T2DM in men and women in the UK Biobank nor the DIAGRAM consortium. (**Figure 5:4**).

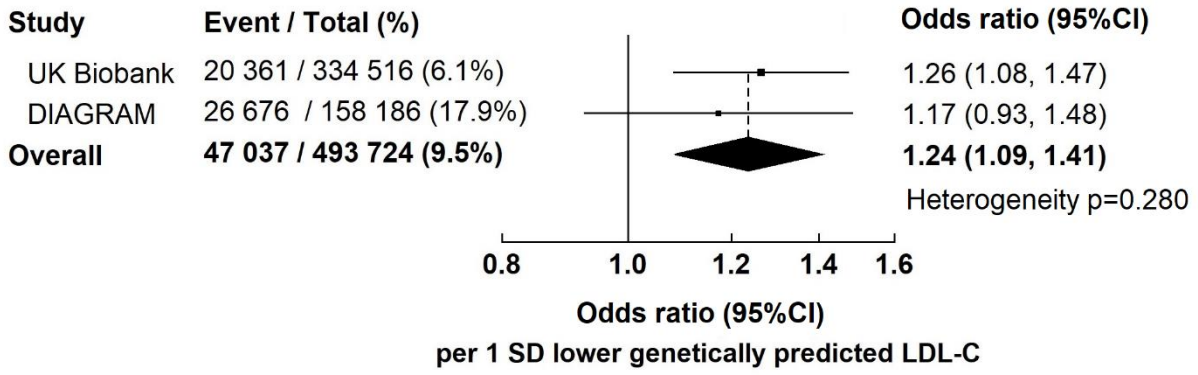


Figure 5:3 | Meta-analysis of the effects of *HMGCR* GRS on risk of T2DM in UK Biobank and the DIAGRAM consortium

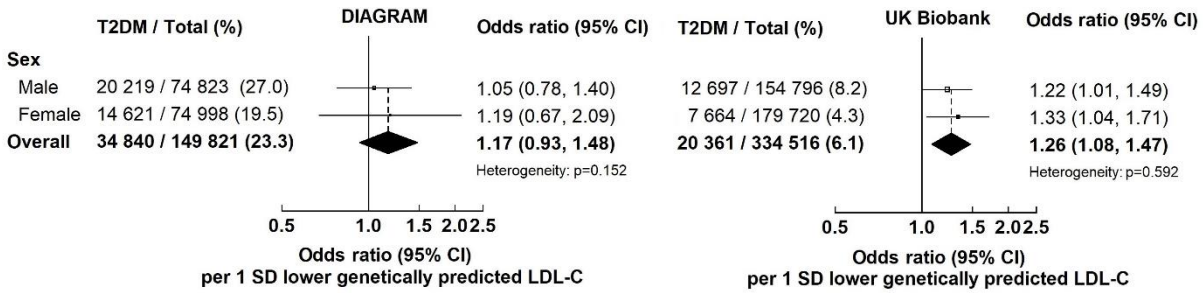


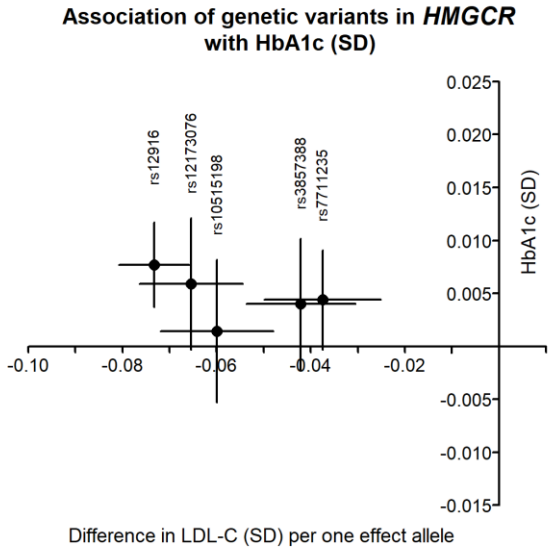
Figure 5:4 | Effects of *HMGCR* GRS on risk of T2DM in UK Biobank and the DIAGRAM consortium stratified by sex

5.3.3 *HMGCR* GRS and glycaemic traits

5.3.3.1 The effect of *HMGCR* on glycaemic traits

The rs12916 variant was associated with a higher level of HbA1c (0.01 SD 95% CI 0.00, 0.01) per 1 copy of the LDL-C lowering allele (**Figure 5:5 A**). The other SNPs were directionally consistent though, individually, not statistically significant. There were no associations between *HMGCR* SNPs and non-fasting blood glucose (**Figure 5:5 B)**.

When combined into a genetic risk score, *HMGCR* was associated with 0.05 SD higher HbA1c (95% CI 0.02, 0.08) (0.39 mmol/mol 95 CI 0.10, 0.69) per 1 SD lower genetically predicted LDL-C but there was no association with non-fasting blood glucose (0.00 SD, 95% CI-0.03, 0.03) (**Figure 5:6**).



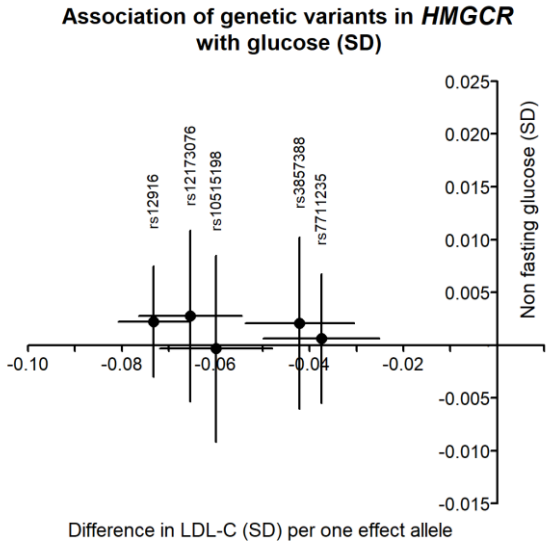


Figure 5.5A | Associations of the *HMGCR* SNPs with HbA1c (SD) and LDL-C (SD). The effect is shown per 1 copy of the LDL-C lowering allele. 1 SD of HbA1c is equivalent to 6.2mmol/mol, 1 SD of LDL-C is equivalent to 1.0 mmol/L.

Figure 5.5B | Associations of the *HMGCR* SNPs with non-fasting glucose (SD) and LDL-C (SD). The effect is shown per 1 copy of the LDL-C lowering allele. 1 SD of HbA1c is equivalent to 1.1mmol/L, 1 SD of LDL-C is equivalent to 1.0 mmol/L.

Figure 5:5 | Associations of the genetic variants in *HMGCR* with glycaemic traits

The figure shows glycaemic and lipid measures in SD units. X-axis shows point estimate and 95% CI for the association with LDL-C, Y-axis shows point estimate and 95% CI for the association with a trait.

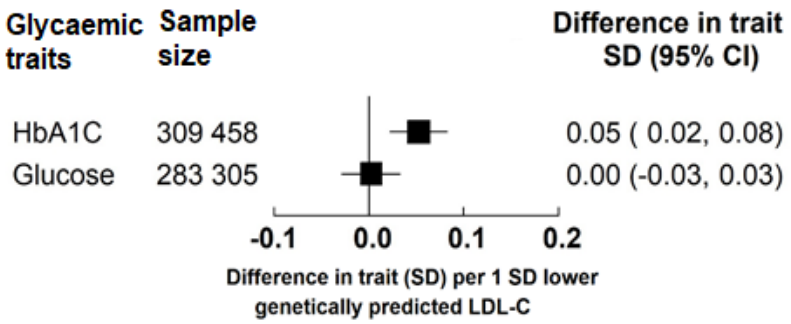


Figure 5:6 | The association of *HMGCR* GRS with glycaemic traits per 1 SD lower genetically predicted LDL-C

The effects on glycaemic traits are shown in SD units per 1 SD lower genetically predicted LDL-C. Participants who reported use of glucose-lowering medication were excluded from the analysis.

5.3.3.2 External validation analysis of the effect of *HMGCR* on HbA1c

The external validation analysis was carried out in an additional 281,416 individuals from the MAGIC consortium. The associations of individual SNPs with HbA1c in the validation dataset were directionally concordant and quantitatively consistent with the results obtained from the UK Biobank dataset (**Table 5:4**). The absence of locus specific heterogeneity in effect-size estimates is shown in a forest plot of the fixed-effect meta-analysis (Appendix 5.3). There was no evidence of heterogeneity in the associations between *HMGCR* GRS and HbA1c in UK Biobank and the set of *HMGCR* SNPs in the MAGIC consortium (P_{het}=0.696) (**Figure 5:7)**.

Table 5:4 | Effects of the *HMGCR* SNPs on HbA1c in UK Biobank and the MAGIC consortium

	Effect allele/ other allele	MAGIC difference in trait, SD (95% CI)	UK Biobank difference in trait, SD (95% CI)	LDL-C difference (from GLGC)
HbA1c				
rs12916	T/C	0.002 (0.000, 0.005)	0.008 (0.004, 0.012)	-0.073 (-0.081, -0.066)
rs12173076	T/G	0.001 (-0.002, 0.005)	0.006 (0.000, 0.012)	-0.065 (-0.076, -0.054)
rs10515198	G/A	0.003 (-0.001, 0.007)	0.001 (-0.005, 0.008)	-0.060 (-0.072, -0.048)
rs3857388	T/C	0.001 (-0.003, 0.005)	0.004 (-0.002, 0.010)	-0.042 (-0.054, -0.031)
rs7711235	A/G	0.001 (-0.002, 0.004)	0.004 (0.000, 0.009)	-0.038 (-0.050, -0.025)

The effect is shown in SD (95% CI) units per 1 copy of effect allele.

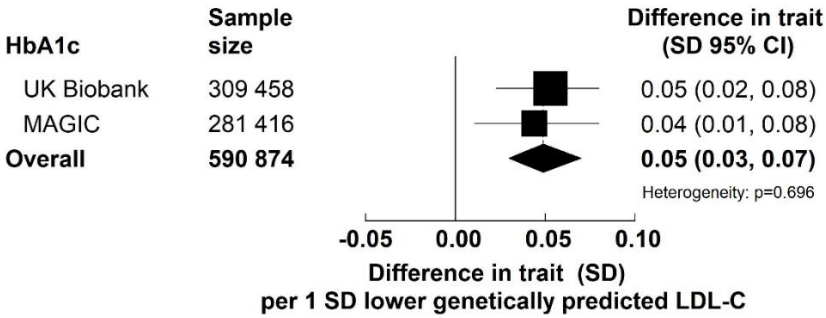


Figure 5:7 | Meta-analysis of the effect of *HMGCR* GRS on HbA1c in UK Biobank and the MAGIC consortium

The effects on glycaemic traits are shown in SD (95% CI) units per 1 SD lower genetically predicted LDL-C. Participants who reported use of glucose-lowering medication were excluded from the UK Biobank analysis. The MAGIC consortium data does not include participants with T2DM.

5.3.4 *HMGCR* GRS and measures of central and general adiposity

5.3.4.1 Individual SNPs and measures of central and general adiposity

Four *HMGCR* SNPs (rs12173076, rs3857388, rs7711235, rs12916) were associated with higher bodyweight (**Figure 5:8** A). They also had consistent effects on various measures of central and general adiposity (**Figure 5:8** C-H). One SNP (rs10515198) appeared to have no effect on the anthropometric measures and indices.

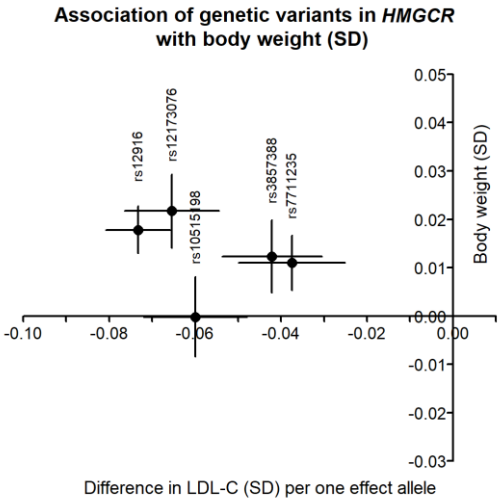


Figure 5:8A | Association of the *HMGCR* index SNPs with bodyweight (SD) and LDL-C (SD). The effect is shown per copy of the LDL-C lowering allele. 1 SD of body weight is equivalent to 15.9 kg, 1 SD of LDL-C is equivalent to 1.0 mmol/L.

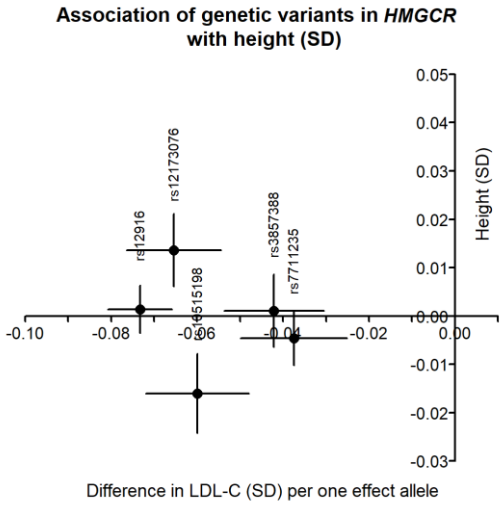


Figure 5:8B | Association of the *HMGCR* index SNPs with height (SD) and LDL-C (SD). The effect is shown per copy of the LDL-C lowering allele. 1 SD of height is equivalent to 9.2 cm, 1 SD of LDL-C is equivalent to 1.0 mmol/L.

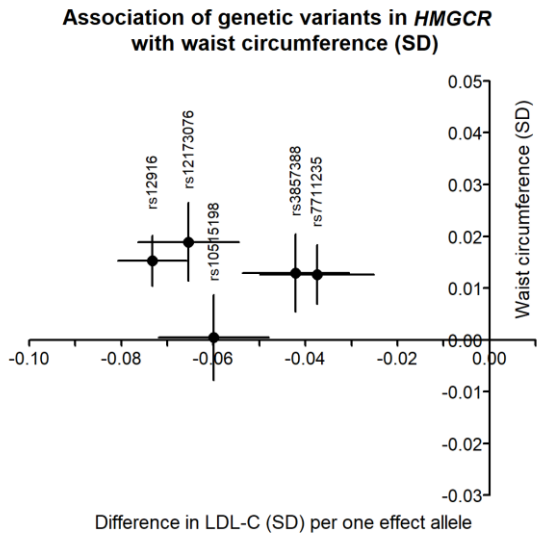


Figure 5:8C | Associations of the *HMGCR* index SNPs with waist circumference (SD) and LDL-C (SD). The effect is shown per copy of the LDL-C lowering allele. 1 SD of waist circumference is equivalent to 13.4 cm, 1 SD of LDL-C is equivalent to 1.0 mmol/L.

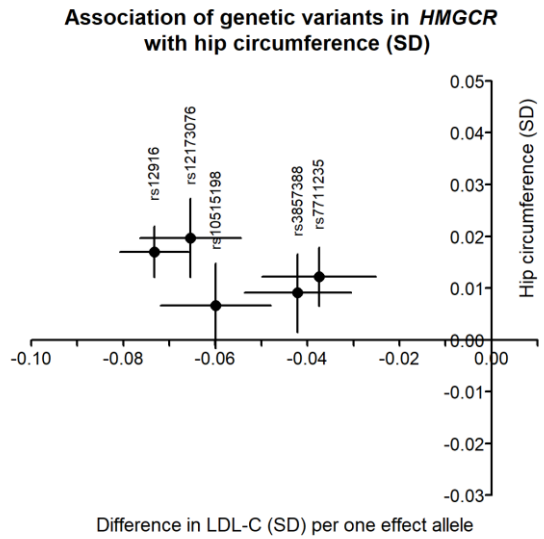


Figure 5:8D | Associations of the *HMGCR* index SNPs with hip circumference (SD) and LDL-C (SD). The effect is shown per copy of the LDL-C lowering allele. 1 SD of hip circumference is equivalent to 9.1 cm, 1 SD of LDL-C is equivalent to 1.0 mmol/L.

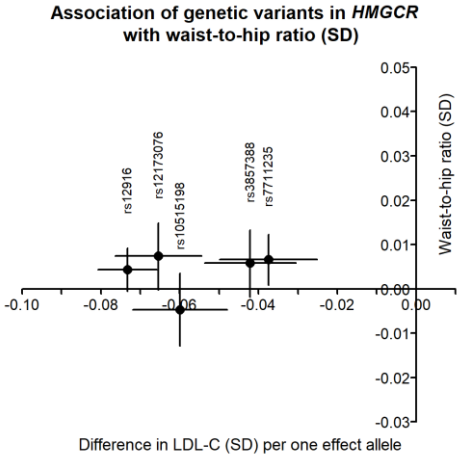


Figure 5:8E | Associations of the *HMGCR* index SNPs with waist-to-hip ratio (SD) and LDL-C (SD). The effect is shown per copy of the LDL-C lowering allele. 1 SD of waist-to-hip ratio is equivalent to 0.10, 1 SD of LDL-C is equivalent to 1.0 mmol/L.

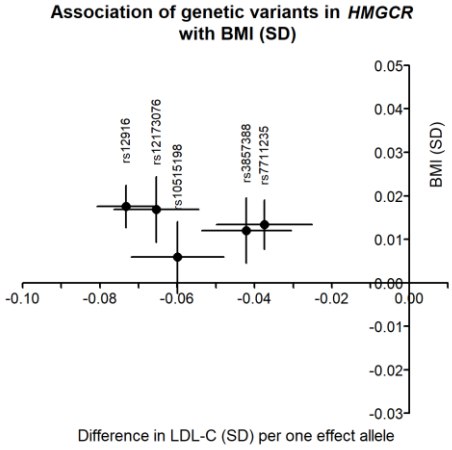


Figure 5:8F | Associations of the *HMGCR* index SNPs with BMI (SD) and LDL-C (SD). The effect is shown per copy of the LDL-C lowering allele. 1 SD of BMI is equivalent to 5.0, 1 SD of LDL-C is equivalent to 1.0 mmol/L.

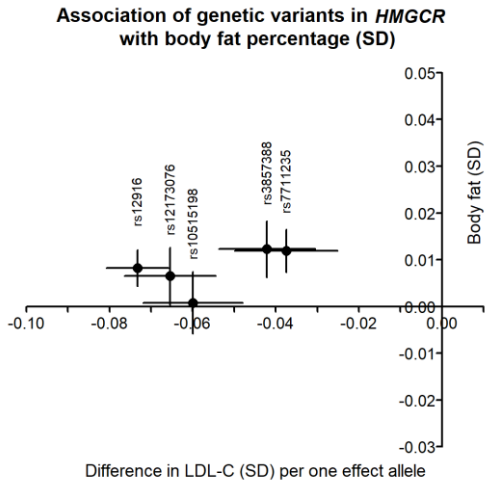


Figure 5:8G | Associations of the *HMGCR* index SNPs with body fat percentage (SD) and LDL-C (SD). The effect is shown per copy of the LDL-C lowering allele. 1 SD of body fat percentage is equivalent to 8.5%, 1 SD of LDL-C is equivalent to 1.0 mmol/L.

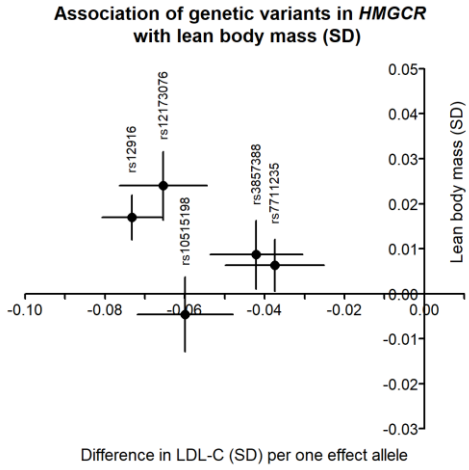


Figure 5:8H | Associations of the *HMGCR* index SNPs with lean body mass (SD) and LDL-C (SD). The effect is shown per copy of the LDL-C lowering allele. 1 SD of lean body mass is equivalent to 11.5 kg, 1 SD of LDL-C is equivalent to 1.0 mmol/L.

Figure 5:8 | Associations of genetic variants in *HMGCR* with anthropometric measures

The figure shows anthropometric measures and lipid in SD units. X-axis shows point estimate and 95% CI for the association with LDL-C, Y-axis shows point estimate and 95% CI for the association with a trait.

5.3.4.2 *HMGCR* GRS and measures of central and general adiposity

1 SD lower genetically predicted LDL-C was associated with 0.13 SD (95% CI 0.10, 0.17) (1.96 kg, 95% CI 1.43, 2.49) higher bodyweight (**Figure 5:9**). Similarly, it was associated with higher waist and hip circumference, BMI, body fat percentage, waist-to-hip ratio, and lean body mass (**Figure 5:9)**. There was no association between *HMGCR* GRS and height. There was no difference in the effects of *HMGCR* GRS on anthropometric measures in men and women.

5.3.4.3 External validation analysis of the effect of *HMGCR* GRS on anthropometric measures

Similar to the UK Biobank results, variants included in the *HMGCR* GRS were positively associated with adiposity measures in the GIANT study. Individual SNPs' effect sizes were approximately similar in UK Biobank and GIANT (Appendix 5.9).

In GIANT, 1 SD lower genetically predicted LDL-C was associated with higher BMI (0.21 SD 95% CI 0.11, 0.30), waist circumference (0.21 SD 95% CI 0.12, 0.30), hip circumference (0.19 SD 95% CI 0.12, 026) and waist-to-hip ratio (0.11 SD 95% CI 0.02, 0.20). There was no association with height and no evidence for heterogeneity between men and women **(Figure 5:10).**

There were directionally consistent sex-combined allelic effects between UK Biobank and GIANT with no evidence of cross-study heterogeneity in the effect on adiposity measures (Appendix 5.9).

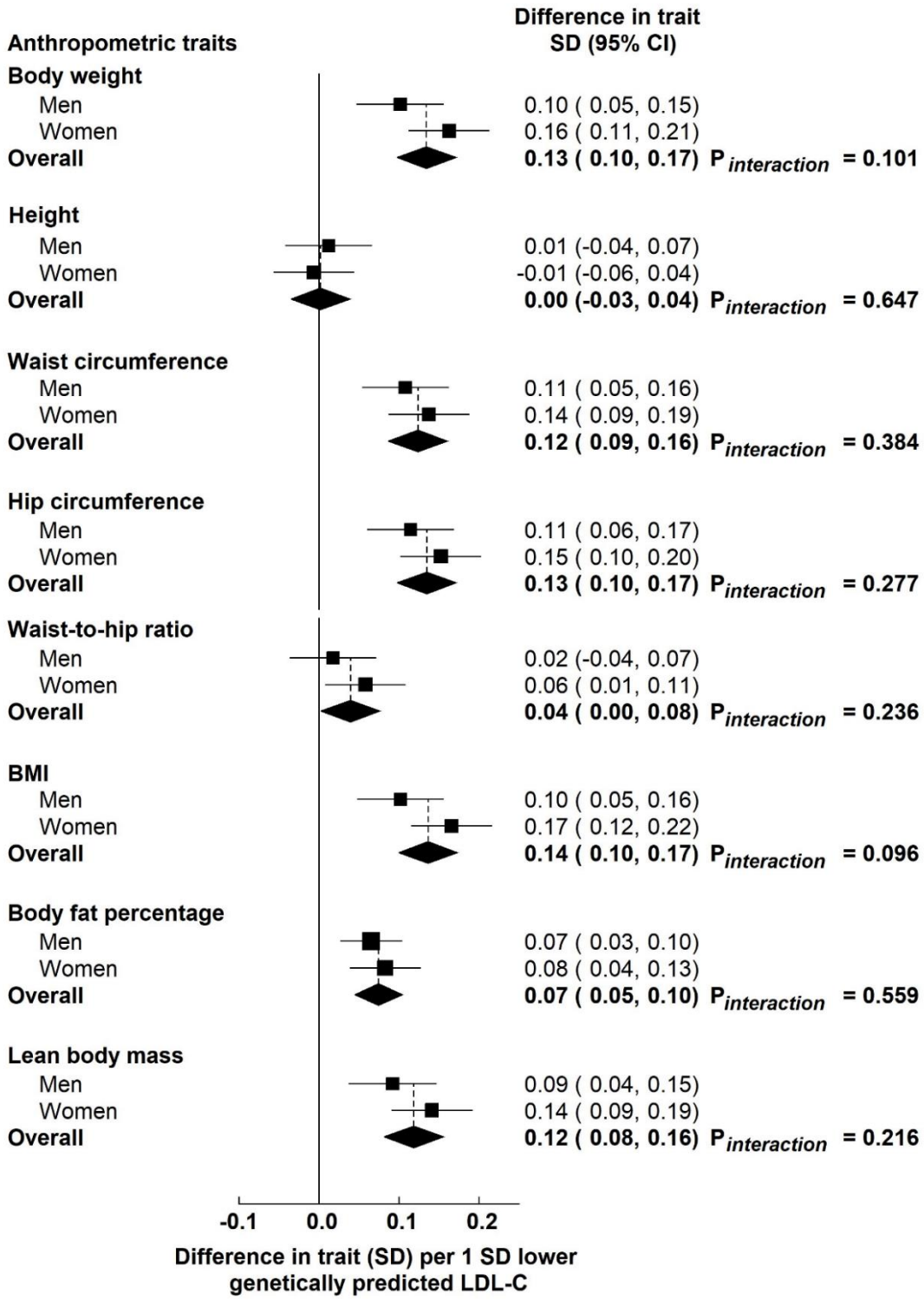


Figure 5:9 | Associations of the *HMGCR* GRS with anthropometric measures per one SD lower genetically predicted LDL-C

SD units (men/women): body weight 14.2/13.9 kg, height 6.8/6.2 cm, waist circumference 11.3/12.5 cm, BMI 4.2/5.1 kg/m^2 , waist-to-hip ratio 0.06/0.07, body fat percentage 7.5/10.3%, hip circumference 7.5/10.3 cm, body lean mass 7.7/5.0 kg.

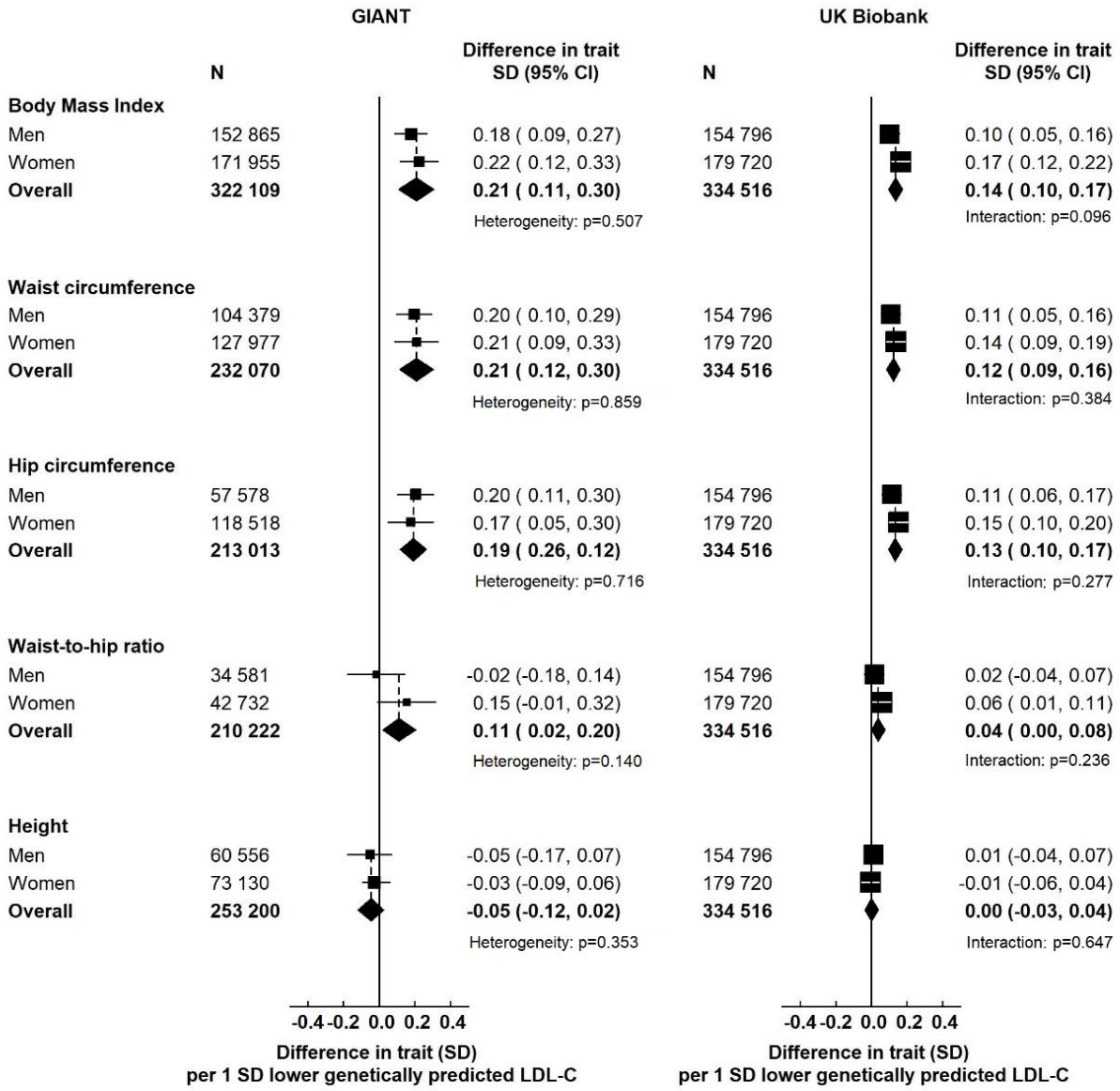


Figure 5:10 | Effects of HMGCR GRS on anthropometric measures in the GIANT consortium and UK Biobank stratified by sex

UK Biobank SD units (men/women): BMI 4.2/5.1 kg/m^{2,} waist circumference 11.3/12.5 cm, hip circumference 7.5/10.3 cm, waist-to-hip ratio 0.06/0.07, hieght 6.8/6.2 cm

GIANT SD units (men/women): BMI 3.6/4.7 kg/m^{2,} waist circumference 9.7/11.0 cm, hip circumference 6.9/9.7 cm, waist-to-hip ratio 0.05/0.06, height 6.1/5.7 cm

5.4 Discussion

5.4.1 The association of *HMGCR* GRS with T2DM

This chapter reports on the associations between genetic variation in *HMGCR*, reflecting the mechanism of action of statin therapy, and the risk of total and incident T2DM, glycaemic traits and anthropometric measures. In the present study the weighted *HMGCR* GRS (representing lower genetically predicted LDL-C) was positively associated with the risk of total T2DM. In UK Biobank, 1 SD lower genetically predicted LDL-C was associated with 26% (95% CI 8%, 47%) higher risk of total T2DM. Although the association between the *HMGCR* GRS and incident T2DM did not reach the 0.05 threshold of statistical significance, it was directionally concordant and quantitatively consistent with the association between the *HMGCR* GRS and total T2DM. There was no heterogeneity in the effect of *HMGCR* GRS on the risk of total T2DM in men and women.

The external validation analysis using summary level data from DIAGRAM found a quantitatively similar (but not statistically significant) association between the *HMGCR* GRS and the risk of T2DM (OR 1.17, 95% CI 0.93, 1.48 per 1 SD lower genetically predicted LDL-C). Consistency between the effect estimates in the current study and the result from the DIAGRAM consortium provides further evidence to support the causal role of lowering LDL-C with HMG-CoA reductase inhibition on the risk of T2DM.

These findings are consistent with previous MR studies, confirming the effect of *HMGCR* on T2DM. Ference *et al.*[51] examined the association between *HMGCR* GRS and risk of T2DM using five SNPs within 250 kb of the *HMGCR* gene in an independent

dataset[1]. Applying a similar methodology to the design of a GRS, but limiting the distance around the gene to 100 kb, a different set of SNPs was selected for the current study (based on the UK Biobank data), including one SNP (rs12916) common for both scores and one SNP (rs3857388) in high linkage disequilibrium (R^2=0.85) with rs2303152 included in Ference's GRS.

The effect of the *HMGCR* GRS on total T2DM in the current study was concordant, but slightly smaller in magnitude, compared to the results reported by Ference *et al*[51] (OR 1.26, 95% CI 1.08, 1.47 and OR 1.33, 95% CI 1.08, 1,61 per 1 SD lower genetically predicted LDL-C respectively). Similarly, the effect of *HMGCR* GRS on incident T2DM in the current study was directionally concordant but smaller in magnitude than reported by Ference *et al.*[51] (OR 1.19 95% CI 0.91, 1.55 and OR 1.55 95% CI 1.00, 2.37 per 1 SD lower genetically predicted LDL-C, respectively). Two other papers[126,127] have reported similar associations between rs12916, the genetic variant within *HMGCR* with the strongest association with LDL-C, and T2DM, however these previous studies were based on largely overlapping data[23]. Results from the

[1] External datasets used in Ference's analysis: ARIC, CARDIA, CHS, DIAGRAM, FHS, MESA, WHI, MIGen, WTCCC, GLGC, CARDIoGRAMplusC4D, MAGIC, GIANT.

[3] External dataset used in Swerdlow's analysis: AAA, ARIC, BRHS, BWHHS, CaPS, CARDIA, CCHS, CFS, CHS, EAS, ELSA, Ely, EPIC-NL, ET2DS, MRK Finland, FHS, HAPIEE-CZ, HAPIEE-LT, HAPIEE-PL, HAPIEE-RU, IMPROVE, InterAct, JUPITER, MESA, MRC NSHD, NPHS-II, PROSPER, SMART, TPT, UCP, WGHS, WHI, Whitehall II

External datasets used in Lotta's analysis: BWHHS, CaPS, CARDIA, CFS, DIAGRAM, ET2DS, InterAct, InterAct –Core-Exome, JUPITER, MESA, NPHS-II, UK Biobank, WHI, WHII, WGHS

current analysis of the UK Biobank data were consistent, confirming previous conclusions on the association between rs12916 and risk of T2DM.

The associations of three other SNPs with T2DM included in the present study (rs10515198, rs7711235, and rs12173076) have not previously been reported, nor did these genetic variants have other SNPs in high linkage disequilibrium which were used for the analyses in previous studies.

The results of the present study correspond to the effect of statins on T2DM observed in RCTs. A meta-analysis of 20 large RCTs of statin therapy demonstrated 0.92 mmol/L (95% CI 0.18, 1.67) reduction in LDL-C on statin (or intensive statin) vs. control (or moderate dose statin) at one year of follow up[67]. After an average of four years of follow-up, participants who were randomised to statin therapy (or intensive statin) had a 12% (95% CI 1.06, 1,18) higher relative risk of T2DM compared to those in the control group (or moderate intensity statin). The present study demonstrates the overall consistency between the risk of T2DM associated with pharmacological inhibition of HMG CoA reductase and genetically defined lower HMG CoA reductase activity. These genetic analyses showed that a life-long *HMGCR*-predicted 1 mmol/L lower plasma LDL-C concentration was associated with 26% (95% CI 1.07,1.49) higher risk of T2DM. This result provides unconfounded evidence that the association observed in randomised clinical trials between statins and T2DM is an on-target effect of HMG-CoA reductase inhibition.

5.4.2 The associations of *HMGCR* GRS with glycaemic traits

This study demonstrated a small but significant effect of *HMGCR* GRS on HbA1c. None of the previous genetic studies have systematically measured HbA1c levels. Use of the UK Biobank data allowed an assessment of the effect of *HMGCR* GRS on HbA1c in a large subset of 309,458 European participants who were not using glucose-lowering medication. This finding supports the conclusion of increased risk of T2DM associated with pharmacological inhibition of HMG-CoA reductase with statins, however it does not explain the biological mechanism of action linking statins with a rise in diabetes risk.

The external validation analysis was performed on a slightly smaller independent population from the MAGIC Consortium (281,416 participants). The directionally concordant and quantitatively consistent effect estimate from this external validation analysis (0.04 SD 95% CI 0.01, 0.07 per 1 SD lower genetically predicted LDL-C in MAGIC compared to 0.05 SD 95% CI 0.02, 0.08 in UK Biobank), validated the result from the current study and demonstrated its reproducibility in an independent population.

There were no effects of *HMGCR* GRS on non-fasting glucose levels. Evidence from previous genetic studies is inconsistent. Swerdlow et al.[126] showed a statistically significant association of *HMGCR* SNP (rs17238484) with non-fasting glucose concentration in 73,490 participants (0.23 mmol/L, 95% CI 0.02, 0.44), whereas Lotta et al.[127] found no effect of rs12916, the *HMGCR* SNP in close linkage disequilibrium with rs17238484 (R^2=0.451), on fasting glucose concentration (0.00, 95% CI -0.01, 001) in 133,010 participants. The present study, with 283,305 participants who were not on glucose-lowering medication and had non-fasting glucose measured, was in

line with Lotta et al.'s findings; none of the *HMGCR* SNPs, including rs12916, was associated with the non-fasting glucose concentrations. A single measure of non-fasting glucose is not a particularly robust biomarker to assess an individual's glycaemic status due to high intra-individual variation (due to recent dietary intake and measurement time[133]. HbA1c test, has a significant advantage over the non-fasting glucose test - it has a low intra-individual variability, it can be tested in a fasting or non-fasting state, and it represents average glucose exposure over the past two-to-three months[134].

5.4.3 The associations of *HMGCR* GRS with anthropometric measures

The effect of statin therapy on T2DM could be partly explained by an association of *HMGCR* GRS with increased adipose tissue or variations in body composition. *HMGCR* GRS was positively associated with the measures of general adiposity, such as BMI and body weight. The effects on waist circumference and hip circumferences were comparable suggesting an association with general adiposity rather than central (visceral) adiposity.

The external validation analysis used sex-combined and sex-specific data from the GIANT consortium to estimate the effects of *HMGCR* GRS on BMI, waist and hip circumference, waist-to-hip ratio, and height. The point estimates from the validation analysis were directionally concordant, and slightly larger in magnitude, than estimates from the present study, providing additional evidence of the effect of *HMGCR* on adiposity. Similar to the UK Biobank results, *HMGCR* had a comparable effect on waist circumference and hip circumference (0.21 and 0.19 SD respectively) in the GIANT

study. The effect on adiposity was likely not limited to the metabolically active visceral fat (as measured by waist circumference and waist-to-hip ratio) as an increase in hip circumference was also observed.

There was no statistically significant heterogeneity in the effects on adiposity between men and women in UK Biobank and GIANT studies. *HMGCR* GRS was associated with higher adiposity measures in both studies.

The results of the current study were concordant with the previously reported associations of individual genetic variants with adiposity measures. In the UK Biobank dataset, each rs12916 LDL-C lowering allele was associated with 0.26 (95% CI 0.19, 0.33) kg higher bodyweight, 0.19 (95% CI 0.13, 0.25) cm greater waist circumference, and 0.09 (95% CI 0.06, 0.11) kg/m² higher BMI. A previous genetic study[126] using different populations has demonstrated concordant results; each rs17238484 LDL-C lowering allele was associated with 0.30 (95% CI 0.18,0.43) kg higher bodyweight, 0.32 (95% CI 0.16, 0.47) cm greater waist circumference and 0.11 (95% CI 0,07, 0,14) kg/m² higher BMI. Each rs12619 LDL-C lowering allele showed concordant results of the same magnitude in that study.

Waist-to-hip ratio is a measure of abdominal adiposity that is associated with risk of T2DM. 1 SD increase in genetically instrumented WHR adjusted to BMI is associated with higher risk of T2DM (OR 1.77, 95% CI 1.57, 2.00)[135]. While establishing an effect of the *HMGCR* genetic variants on bodyweight and other measurements of general adiposity, this project found a weak association between *HMGCR* SNPs and waste-to-hip ratio (1 SD genetically determined LDL- is associated with 0.04 SD higher WHR). This analysis confirmed previous findings of Lotta *et al.*[127] and Swerdlow *et al.*[126] who reported consistent results for the association of rs12619 and rs17238484

with waste-to-hip ratio. The current study found no specific effect of *HMGCR* GRS on fat distribution patterns. Specifically, there was no evidence of a selective association with abdominal fat accumulation, which is a known risk factor for T2DM[136].

This genetic analysis is the first to suggest an association between genetic variants in *HMGCR* and an increase in lean body mass. Muscle-related symptoms are rare complications of statin therapy[8] and it has also been shown that statin therapy increases levels of creatine kinase[137], an enzyme derived from leakage from skeletal muscle. It is therefore plausible that statin therapy increases muscle mass as well as adiposity. A possible positive association of *HMGCR* with higher lean body mass is surprising, given that *HMGCR* GRS is also associated with a higher risk of T2DM and higher levels of HbA1c. Previous studies have demonstrated an inverse association between lean body mass and the risk of T2DM[138]; muscles also play an essential role in insulin and glucose regulation[139]. These insights suggest that the effect of statin therapy on glycaemia and risk of T2DM is not mediated by decreased muscle mass.

5.5 Summary

In the present Drug Target mendelian randomisation analysis, the weighted *HMGCR* GRS, constructed from five independent polymorphisms, was positively associated with the risk of total T2DM. The association of the *HMGCR* GRS with incident T2DM was quantitatively and directionally consistent.

The results of the present study also demonstrate an association between *HMGCR* GRS and HbA1c. Lower genetically predicted LDL-C, as a consequence of lower HMG CoA reductase activity due to SNPs in *HMGCR*, was associated with higher HbA1c. The analysis has also demonstrated strong evidence of an association between the *HMGCR* GRS and adiposity measures. Lower genetically predicted LDL-C was associated with higher body weight and waist and hip circumferences. The other adiposity measures had concordant results. There was no effect of *HMGCR* GRS on height.

The positive associations of *HMGCR* GRS with anthropometric measures, especially with bodyweight and waist circumference, suggests that the effect of *HMGCR* inhibition on the risk of T2DM may by partly explained by increased deposits of both visceral and subcutaneous fat. There was no heterogeneity between men and women in the effect of *HMGCR* GRS on anthropometric measures and the risk of total T2DM. Together with earlier findings demonstrating an increasing effect on adiposity, the *HMGCR* GRS association with HbA1c supports the causal role of *HMGCR* inhibition in the development of T2DM.

Chapter 6 The associations of the *HMGCR* GRS with the risk of T2DM, glycaemic traits and anthropometric measures by the baseline risk of T2DM

6.1 Introduction

Previous large-scale meta-analyses of RCTs of statins and genetic studies, including the present study, have reached a consensus on the effect of HMGCR inhibition on the risk of T2DM [19,67,140]. However, the mechanisms of the statin-induced transition to T2DM are poorly understood. A post-hoc analysis based on the JUPITER trial[87] examined the effect of treatment with rosuvastatin 20 mg therapy (compared to placebo) on the risk of incident T2DM in groups of participant without T2DM and with different T2DM risk factors (elevated fasting plasma glucose (>5.6 mmol/L), fasting triglycerides (> 1.69 mmol/L), higher BMI (>30 kg/m^2), and history of hypertension) at baseline. The subgroup of participants with 1-4 risk factors (N=11,508) had a 28% (HR 1.28; 95% CI 1.07, 1.54) higher proportional risk of statin-associated T2DM than placebo, whereas those without any risk factors (N=6,095) had no excess risk of the disease (HR 0.99; 95% CI 0.45, 2.21) on rosuvastatin compared to placebo. The authors concluded that statin therapy appears to exert a diabetogenic action only in individuals with pre-existing risk factors for T2DM.

Another analysis of 15,056 non-diabetic participants from the TNT and IDEAL trials[141] showed that pre-diabetic participants (fasting blood glucose between 100 and 126 mg/dl), randomised to high intensity statin, experienced a 29% (OR 1.29; 95% CI 1.04,

1.37) proportional increase in risk of incident T2DM compared to those randomised to moderate intensity statin. By contrast, there was no statistically significant increase in the risk of incident T2DM in participants with normal glucose control randomised to higher vs moderate intensity statins in these trials (HR 1.08; 95% CI 0.85, 1.38). The test for interaction, however, was not statistically significant ($P_{interaction}$=0.072), indicating that individual baseline risk of T2DM did not clearly modify the effect of treatment on the incident T2DM.

A Mendelian randomisation study which used an *HMGCR* GRS as a genetic proxy of pharmacological inhibition of HMGCR with statins also assessed the effect of the *HMGCR* GRS on incident T2DM [51]. Ference *et al.*[51] reported a borderline association between the *HMGCR* GRS and incident T2DM in people with impaired glucose tolerance (OR 1.96; 95% CI 1.00, 3.77 per 1 mmol/L lower *HMGCR* predicted LDL-C). No statistically significant association between *HMGCR* GRS and risk of incident T2DM was observed in participants with normal levels of fasting blood glucose at baseline (OR 1.04; 95% CI 0.89, 1.22).

These RCTs and genetic studies therefore raise the hypothesis that any diabetogenic effect of statins is limited to those already at high risk of developing T2DM. However, there are limitations in these analyses.

One of the major drawbacks of the post-hoc analysis of the JUPITER trial[87] was running a subgroup analysis in which one subgroup had very few events. The power in the two subgroups differed, as the high-risk group was twice as large as the low risk group (11,508 vs 6,095 participants) and contained 95% of the incident T2DM cases (462 vs 24). Another limitation of this analysis was not conducting a test for interaction but claiming a statistically significant subgroup effect in only one subgroup.

Similarly, the interpretation of the subgroup result for *HMGCR* GRS on the risk of incident T2DM in the genetic study[51] contains potential methodological flaws[142]. Ference *et al.*[51] did not assess the heterogeneity of the effect across the subgroups, despite claiming that the association of the *HMGCR* GRS with incident T2DM was limited to the group of patients with impaired glucose control. Moreover, the genetic study by Ference *et al.*[51] showed only a marginal association of *HMGCR* GRS with incident T2DM in the pre-diabetic subgroup (OR 1.001, 95% CI 1.000, 1001) and failed to reach 0.05 level of statistical significance even before consideration of multiple testing.

Given insufficient evidence of heterogeneity of the proportional effect of HMGCR inhibition on risk of T2DM in patients with normal and impaired glycaemic control has been presented to date, one of the specific objectives of this research was to examine the consistency of the association between *HMGCR* GRS and T2DM across participants with different risks of developing T2DM. This chapter therefore aimed to address the following research questions: (I) whether the effect of *HMGCR* on the risk of T2DM differs between subgroups of participants at different risks of T2DM; (II) whether the effects of *HMGCR* on blood biomarkers of glycaemic control and anthropometric measures differ between subgroups of participants at different risks of T2DM.

The *HMGCR* GRS was used as a genetic proxy of the on-target effect of statins to compare the risk of developing T2DM associated with HMGCR inhibition in the subgroups of participants at different risks of T2DM. Firstly, for this chapter's main analysis, subgroups were based on the genetic risk of total T2DM using a T2DM polygenic risk score, which represents an individual's genetic predisposition to the

disease. This polygenic T2DM risk score was used to stratify participants into low, medium, and high T2DM genetic risk groups. The use of a genetic risk score as a stratifying variable allowed the inclusion of the whole study population and the use of total T2DM as the outcome of interest. A potential limitation of this method is low discriminatory accuracy of a polygenic T2DM risk score, as it may only explain a small percentage of variance of T2DM risk.

To overcome this limitation, a secondary analysis was performed using the QDiabetes-2018 risk score as a stratifying variable. The QDiabetes-2018 risk score [143] estimates the 10-year absolute risk of developing T2DM in non-diabetic individuals. While allowing greater discriminatory accuracy, the QDiabetes-2018 risk score can only be applied as a stratifying variable to non-diabetic participants, thereby limiting the outcome of interest to incident cases of T2DM only.

6.2 Study design and method

6.2.1 Study population

For the main analysis, the study population with 20,361 cases of total T2DM and 314,155 controls was used to examine the associations between the *HMGCR* GRS and T2DM in subgroups of participants with different genetic risks of T2DM. The study population is described in detail in Chapter 5.3.1

For the secondary analysis, only participants who were free from T2DM at the baseline UKB assessment were included. Data from 6,467 cases of incident T2DM and 314,155 controls were used to examine the effect of the *HMGCR* GRS on incident T2DM in the subgroups with different estimated absolute risks of the disease as calculated by the QDiabetes-2018 risk score. This analysis excluded 13,894 participants with prevalent T2DM (i.e. diagnosed prior the baseline assessment date). The subpopulation used for the sensitivity analysis was described in detail in Chapter 5.3.1

As a sensitivity analysis, data from participants who were not diabetic at the baseline assessment were used to assess the effect of the *HMGCR* GRS on risk of incident T2DM according to HbA1c level. Stratified by pre-diabetes status (healthy and pre-diabetes subgroups had HbA1c levels of <42, and ≥ 42<48 mmol/l, respectively), subgroup analysis included participants who were free from T2DM at the baseline assessment and had a valid value of HbA1c. 15,780 participants with missing HbA1c values and 2,268 participants with HbA1c ≥ 48 mmol/L (the diagnostic threshold for T2DM) were excluded from the analysis.

6.2.2 T2DM risk scores

6.2.2.1 Polygenic T2DM risk score

Genetic susceptibility to T2DM is determined by more than 250 associated loci[144]. There are numerous intermediate phenotypes associated with the disease, including glucose-related and insulin-related measures, and adipose tissues quantity and distribution. The associations between these loci and T2DM were summarised by a large international meta-analysis of GWAS studies (the DIAGRAM Consortium[47,73,145]).

The T2DM GRS used in this study was based on the meta-analysis by Morris et al[47]. That study included 34,840 cases and 114,981 controls of European origin (data from DIAGRAMv3) and Pakistani descent (data from PROMIS)[47] in whom there was no evidence of heterogeneity between allelic effects in the two ethnic groups. Genotyping was carried out with the custom-designed array of 196,725 variants using the Metabochip. This GWAS identified 119 T2DM susceptibility loci, some of which were also associated with various T2DM related anthropometric measures and blood biomarkers. Combined, those loci accounted for 5.7% of variance in T2DM susceptibilities by a continuous liability scale.

To compute the polygenic T2DM risk score in the present study, 119 SNPs identified by Morris *et al*.[47] were extracted from the UKB genetic data. The GWAS summary associations were clumped using PLINK software and 1000 Genome Phase 3 European population dataset as a reference for linkage disequilibrium computation. Linkage disequilibrium threshold was set as r^2=0.2 and a significance threshold was set on $p<1.0E^{-8}$. Independent SNPs with the lowest p-value for the association with

T2DM were identified as index SNPs and were used to compute the polygenic T2DM risk score.

The clumping procedure retained 57 index SNPs, which were harmonised by the direction of the effect (higher risk of T2DM), and frequencies of the effect alleles were compared with the European population in the 1000 Genome Phase 3 European population dataset. Appendix 6.1 shows the basic information on the SNPs included into the polygenic T2DM risk score in UK Biobank and the DIAGRAM consortium databases.

Harmonised SNPs were weighted by their effect size on T2DM as reported in the DIAGRAM meta-analysis. The effect sizes were the natural logarithm of the reported odds ratios. The polygenic T2DM risk score was constructed using an additive model by summing the numbers of effect alleles weighted by their effect sizes. The algorithm of the computation of the polygenic T2DM risk score is shown in **Figure 6:1** and the distribution of the polygenic T2DM genetic risk score is shown in Appendix 6.2.

The polygenic T2DM risk score was used to classify the study population into three subgroups by genetic risk of the T2DM. Tertiles of the polygenic T2DM risk score were chosen as cut-off points for low/medium/high genetic susceptibility to T2DM.

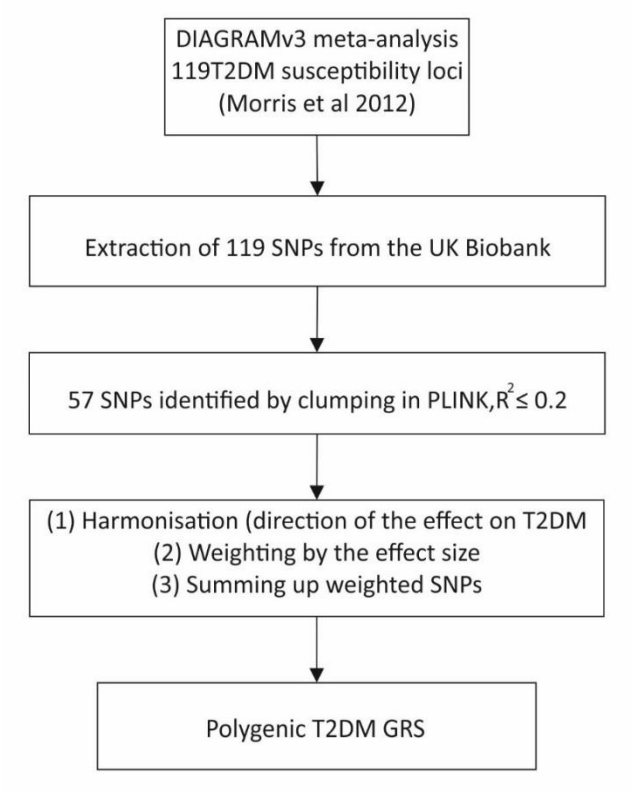


Figure 6:1 | Polygenic T2DM risk score computation algorithm
SNPs' weights were estimated by the DIAGRAM consortium (primary European population), R2 coefficients were estimated with the 1000 Genome Project reference panel (European population).

6.2.2.2 QDiabetes-2018 risk score

The QDiabetes-2018 risk score was developed by ClinRisk[143,146] for use in clinical practice, and it estimates the 10-year absolute risk of incident T2DM in people without T2DM aged 25-84[143]. The original algorithm was developed based on the medical data of the patients registered in 1,457 general practices across England. There were 178,314 cases of T2DM in the derivation cohort and 62,326 cases of T2DM in the validation cohort. Three models of the QDiabetes-2018 algorithm are available and each includes slightly different sets of risk factors. Model A includes demographic characteristics and anthropometric measurements, self-reported family history of

T2DM, comorbidities, use of medications which increase the risk of developing T2DM (including statins), an indicator of socio-economic status, and information about smoking. Model B adds fasting blood glucose concentration to these risk factors. Model C replaces the fasting blood glucose concentration with an HbA1c value.

In the current analysis Model A, which uses only self-reported risk factors, was used to split participants into low/moderate/high risk groups with cut-off points of <10%, 10—20%, and >20%. The threshold of 10% identified the top 20% of the participants with high risk of T2DM; the threshold of 20% identified the top 5% of the participants with very high risk of T2DM.

Appendix 6.3 shows the distribution of QDiabetes-2018 risk scores in the UKB study population. In this thesis the smoking status was reported the same way as it was used in the QDiabetes-2018 risk score[143], which is different from the original UKB format. The baseline characteristics table, stratified by QDiabetes-2018 risk groups (**Table 6:2**), includes 404 participants on hypoglycaemic medication. The algorithmic solution[104] used for the definition of prevalent T2DM classified these 404 individuals as possible T2DM due to lack of evidence for the probable disease status. Therefore, these cases were not excluded from the dataset prior to the analysis.

6.2.3 Statistical methods

General statistical methods were described in Chapter 3.6. To examine whether the effect of the *HMGCR* GRS on risk of total T2DM differs in the subgroups of participants with different genetic risks of T2DM, subgroup analyses were conducted. Tests for

trend were conducted to examine if the baseline risk of T2DM changed systematically across the groups of participants with different risks of T2DM.

Then, the associations of the *HMGCR* GRS with glycaemic traits and anthropometric measures were examined in each subgroup according to their underlying T2DM risk. Glycaemic traits and anthropometric measures were regressed on *HMGCR* GRS using linear regression. Each model was adjusted for sex, genotyping array and 40 principal components of ancestry. Participants who reported use of hypoglycaemic medications were excluded from the analysis of the effect of *HMGCR* GRS on glycaemic traits.

The group specific p-values were not reported due to the risk of over interpretation of the results of multiple testing. Instead, together with the P-values for interaction, the figures include P-values for trend (P_{trend}), which assesses a systematic change in the trait across ordered subgroups.

For the secondary analysis, the dataset was restricted to the participants who were not diabetic prior to the date of the baseline assessment. This study population was stratified into three risk groups by QDiabetes-2018 risk score and the logistic regression model, which was adjusted for sex, genotyping array and 40 principal components of ancestry, was used to estimate the odds ratio of incident T2DM in the subgroups.

Floating absolute risk methods were used to compute confidence intervals for the reference group and two non-reference groups.

The sensitivity analysis used HbA1c levels as a stratifying variable to split the study population into two subgroups of participants (with pre-diabetes [≥42<48 mmol/mol]

and no pre-diabetes [<42 mmol/mol]). HbA1c levels in such analyses act as a collider since it is influenced by both the exposure (*HMGCR* GRS) and the outcome (incident T2DM). Stratifying the population by HbA1c levels potentially introduces the risks of a collider-stratification bias, which may lead to spurious associations between the exposure and the outcome in one of the strata[147,148]. Collider-stratification bias could potentially invalidate the results of the subgroup analysis by introducing heterogeneity in the subgroup-specific estimates, even when the true estimates are homogeneous across the subgroups[147]. Such heterogeneity arises from the ability of the collider-stratification bias to alter the association between the instrumental variable and the outcome in one stratum, but not in another[147].

The possible threat of a collider-stratification bias was addressed by adding an MR-specific approach to the primary analysis. While the first model stratified the population by the collider (i.e. natural units of HbA1c (mmol/mol)), the second model used a new stratifying variable called a residual collider. The residual collider variable was computed by regressing the collider (HbA1c in mmol/mol) on the instrumental variable (*HMGCR* GRS). The model was adjusted by sex, 40 PCs, and genotyping array. The HbA1c residuals were not associated with the instrumental variable and, therefore, could not be a collider. The outcome was still influenced by the residual collider but not as a function of the instrumental variable, and therefore it did not violate the MR instrumental variable assumption.

The second model used HbA1c residuals as a stratifying variable; the non-diabetic population was split into 30 equal groups to generate a high risk group of similar size compared to the first model.

The sensitivity analysis presented estimates for the causal effect of *HMGCR* GRS on incident T2DM in subgroups of HbA1c mmol/mol and residual HbA1c.

6.3 Results

6.3.1 *HMGCR* GRS and total T2DM in participants with different genetic risks of T2DM

6.3.1.1 Characteristics of the study population based on polygenic T2DM risk

Baseline characteristics in the subgroups, based on the T2DM GRS, did not differ except for use of hypoglycaemic medications, and family history of T2DM and related comorbidities, such as cardiovascular disease and hypertension (**Table 6:1**). 16.9% of the participants in the low-risk group reported a family history of T2DM whereas, in the medium and high-risk groups, the percentages were 20.0% and 23.7% respectively. Similarly, there was a slightly higher prevalence of self-reported cardiovascular disease and treated hypertension in the high-risk group compared to the medium and low-risk groups.

Table 6:1 | Baseline characteristics of 334,515 participants of the UK Biobank in three subgroups by polygenic T2DM risk score

	Low T2DM genetic risk group	Medium T2DM genetic risk group	High T2DM genetic risk group
PATIENT CHARACTERISTICS			
Sample size	115 765 (34.6%)	111 888 (33.5%)	106 862 (31.9%)
Male sex	53 478 (46.2%)	51 735 (46.2%)	49 583 (46.4%)
Age at the baseline, (yr.)	56.9 ± 8.0	56.9 ± 8.0	56.9 ± 8.0
Townsend score (points)	-1.6 ± 2.9	-1.6 ± 2.9	-1.6 ± 2.9
GLYCAEMIC TRAIT			
HbA1c	35.1 ± 5.3	35.8 ± 6.0	36.6 ± 7.1
ANTHROPOMETRIC MEASURES			
Bodyweight (kg)	78.0 ± 16.5	78.1 ± 16.5	78.1 ± 16.1
BMI (kg/m^2)	27.3 ± 5.0	27.3 ± 5.0	27.3 ± 4.9
Waist circumference (cm)	90.1 ± 14.0	90.2 ± 14.0	90.2 ± 13.7
Hip circumference (cm)	103.3 ± 10.2	103.3 ± 10.2	103.2 ± 9.9
Waist to hip ratio	0.9 ± 0.1	0.9 ± 0.1	0.9 ± 0.1
Body fat percentage (%)	30.8 ± 9.4	30.8 ± 9.4	30.7 ± 9.4
Body fat mass (kg)	24.4 ± 10.1	24.4 ± 10.0	24.3 ± 9.8
SMOKING			
Non-smoker	63 593 (54.9%)	61 209 (54.7%)	58 686 (54.9%)
Former smoker	40 563 (35.0%)	39 258 (35.1%)	37 589 (35.2%)
Light smoker	1 483 (1.3%)	1 501 (1.3%)	1 342 (1.3%)
Moderate smoker	7 052 (6.1%)	6 965 (6.2%)	6 449 (6.0%)
Heavy smoker	3 074 (2.7%)	2 955 (2.6%)	2 796 (2.6%)
MEDICAL HISTORY			
Family history of T2DM	19 551 (16.9%)	22 360 (20.0%)	25 359 (23.7%)
Treated hypertension	24 321 (21.0%)	24 672 (22.1%)	24 679 (23.1%)
Cardiovascular disease	6 941 (6.0%)	7 074 (6.3%)	7 216 (6.8%)

Bipolar disorder	433 (0.4%)	396 (0.4%)	378 (0.4%)
Gestational T2DM	10 (0.0%)	15 (0.0%)	23 (0.0%)
Polycystic ovary syndrome	134 (0.1%)	128 (0.1%)	120 (0.1%)
MEDICATION			
Hypoglycaemic	1 920 (1.7%)	3 103 (2.8%)	4 667 (4.4%)
Statins	18 837 (16.3%)	19 557 (17.5%)	20 308 (19.0%)
Atypical antipsychotics	321 (0.3%)	292 (0.3%)	282 (0.3%)
Corticosteroids	1 240 (1.1%)	1 179 (1.1%)	1 162 (1.1%)

Percentage (%) was used for binary traits and mean +/- SD for continuous traits.

6.3.1.2 Stratification by the polygenic T2DM risk score

The number of participants with T2DM at baseline was 13,894. The polygenic T2DM risk score stratified the study population into three groups with different proportions of total T2DM. The overall prevalence of T2DM was 6.1%, the lowest prevalence of T2DM was in the low genetic risk group (3.8%), followed by medium and high genetic risk groups (6.0% and 8.7% respectively).

The medium genetic risk group was associated with a 60% (95% CI 1.56, 1.63) higher risk of T2DM compared to the low genetic risk group. The high genetic risk group had a twofold (OR 2.40; 95% CI 2.36, 2.45) increase in the risk of T2DM compared to the low genetic risk group. A test for trend demonstrated a strong linear relationship between the polygenic T2DM risk score and the risk of total T2DM (**Figure 6:2**) as expected.

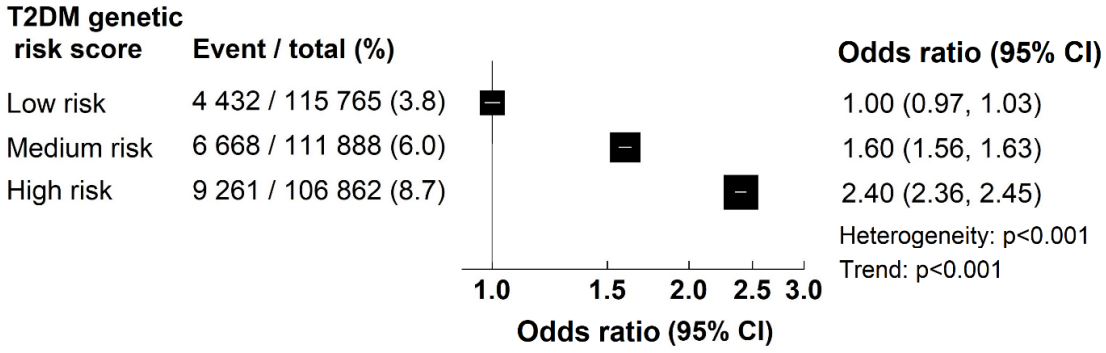


Figure 6:2 | Risk of T2DM in three risk-stratified subgroups (by tertiles of polygenic T2DM risk score)
T2DM genetic risk score is comprised of 56 SNPs identified by Morris et al. (2012). The model includes 20,361 cases of T2DM and 314,155 controls, all participants were white British who passed genetic QC. The confidence intervals were estimated using floating absolute risk method.

6.3.1.3 *HMGCR* GRS and risk of total T2DM by genetic risk of T2DM

The subgroup analysis comparing the effect of the *HMGCR* GRS on risk of total T2DM in the subgroups of participants stratified by the level of genetic susceptibility to T2DM found no evidence that the individual genetic risk of T2DM modifies the proportional effect of the *HMGCR* GRS on the disease (**Figure 6:3**) The test for heterogeneity found no material difference in the effect of *HMGCR* GRS on T2DM across the groups (P_{het}=0.411).

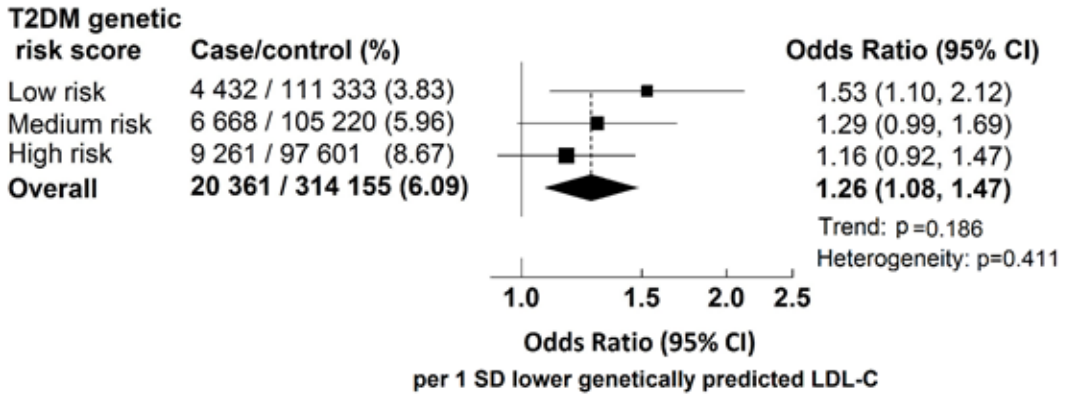


Figure 6:3 | Risk of total T2DM associated with 1 SD lower *HMGCR* predicted LDL-C (by tertiles of the polygenic T2DM risk score)
HMGCR GRS is a continuous genetic risk score (comprised of 5 HMGCR SNPs).

6.3.2 *HMGCR* GRS and incident T2DM in participants with different absolute risks of T2DM

6.3.2.1 Characteristics of the study population

A total of 320,621 participants who were free from T2DM at baseline were included in the secondary analysis in which the association between *HMGCR* GRS and incident diabetes was explored according to clinical risk of developing diabetes. Participants without T2DM at baseline were stratified by the QDiabetes-2018 estimated 10-year absolute risk of T2DM. There were 174,733 (52.2%) women, the mean (SD) age was 58.2 (8), 5.5% of the participants reported prevalent cardiovascular disease, 19.2% treated hypertension, and 18.4% participants had a family history of T2DM. 6,467 participants developed incident T2DM. The baseline characteristics for the subset of study population used for the sensitivity analysis stratified by QDiabetes-2018 risk score are summarised in **Table 6:2**.

The majority of participants had a low absolute risk of T2DM (79.1%) followed by a medium risk and high risks groups (15.0% and 5.9% respectively), although a similar number of incident cases of T2DM occurred in each group. The percentage of men in the low-risk group was the smallest (40.3%), notably increasing in the medium and high risks groups (62.7% and 71.2% respectively). Similarly, there was a higher prevalence of self-reported cardiovascular disease, treated hypertension and family history of T2DM in the high-risk group compared to the medium and low risk groups.

Table 6:2 | Baseline characteristics of 320,622 participants of the UK Biobank who were free from T2DM at baseline in three subgroups by QDiabetes-2018 risk score

	Low QDiabetes-2018 risk score (<10%)	Medium QDiabetes-2018 risk score (10-20%)	High QDiabetes-2018 risk score (>20%)
PATIENT CHARACTERISTICS			
Sample size	253 593 (79.1%)	48 220 (15.0%)	18 809 (5.9%)
Male sex	102 238 (40.3%)	30 239 (62.7%)	13 390 (71.2%)
Age at the baseline, (yr.)	55.8 ± 8.1	60.0 ± 6.7	60.2 ± 6.3
Townsend score (points)	-1.8 ± 2.8	-1.1 ± 3.1	-0.2 ± 3.4
GLYCAEMIC TRAIT			
HbA1c	34.6 ± 4.2	36.6 ± 5.2	38.2 ± 6.3
ANTHROPOMETRIC MEASURES			
Bodyweight (kg)	73.1 ± 13.0	90.9 ± 13.4	102.7 ± 15.0
BMI (kg/m^2)	25.7 ± 3.7	31.5 ± 4.2	35.4 ± 4.4
Waist circumference (cm)	85.6 ± 11.3	101.7 ± 9.7	111.2 ± 10.9

	Low QDiabetes-2018 risk score (<10%)	Medium QDiabetes-2018 risk score (10-20%)	High QDiabetes-2018 risk score (>20%)
Hip circumference (cm)	100.8 ± 8.4	109.8 ± 9.7	115.6 ± 10.5
Waist-to-hip ratio	0.8 ± 0.1	0.9 ± 0.1	1.0 ± 0.1
Body fat percentage (%)	29.7 ± 9.0	33.8 ± 9.7	35.8 ± 9.6
Body fat mass (kg)	21.8 ± 7.9	30.9 ± 10.5	36.9 ± 11.6
SMOKING			
Non-smoker	150 741 (59.4%)	20 682 (42.9%)	6 253 (33.2%)
Former smoker	79 806 (31.5%)	21 496 (44.6%)	9 527 (50.7%)
Light smoker	3 444 (1.4%)	564 (1.2%)	194 (1.0%)
Moderate smoker	14 261 (5.6%)	3 635 (7.5%)	1 695 (9.0%)
Heavy smoker	5 341 (2.1%)	1 843 (3.8%)	1 140 (6.1%)
MEDICAL HISTORY			
Family history of T2DM	38 369 (15.1%)	14 666 (30.4%)	8 413 (44.7%)
Treated hypertension	29 018 (11.4%)	21 522 (44.6%)	13 558 (72.1%)
Cardiovascular disease	6 649 (2.6%)	6 330 (13.1%)	5 234 (27.8%)
Schizophrenia or bipolar disorder	647 (0.3%)	257 (0.5%)	214 (1.1%)
Gestational T2DM	6 (0.0%)	5 (0.0%)	30 (0.2%)
Polycystic ovary syndrome	255 (0.1%)	68 (0.1%)	38 (0.2%)
MEDICATION			
Hypoglycaemic	117 (0.1%)	124 (0.3%)	163 (0.9%)
Statins	20 493 (8.1%)	16 325 (33.9%)	10 640 (56.6%)
Atypical antipsychotics	425 (0.2%)	198 (0.4%)	187 (1.0%)
Corticosteroids	2 076 (0.8%)	758 (1.6%)	488 (2.6%)

Percentage (%) was used for binary traits and mean +/- SD for continuous traits.

6.3.2.2 Stratification by QDiabetes-2018 risk score

The risk of incident T2DM in the medium risk group, as identified by the QDiabetes-2018 risk score, was around five times higher (OR 4.89; 95% CI 4.72, 5.08) compared to the low risk group. The risk of T2DM in the high-risk group was ten times higher (OR 10.10; 95% CI 9.73, 10.50) than in the low-risk group. The test for trend demonstrated the expected strong relationship between QDiabetes-2018 risk and the risk of incident T2DM (**Figure 6:4)**. While the overall risk of incident T2DM was 2.02%, stratified by QDiabetes-2018 risk score, absolute 10-year risk of incident T2DM increased from 0.88% in the low-risk group to 4.50 in the medium risk group and 11.00% in high-risk group.

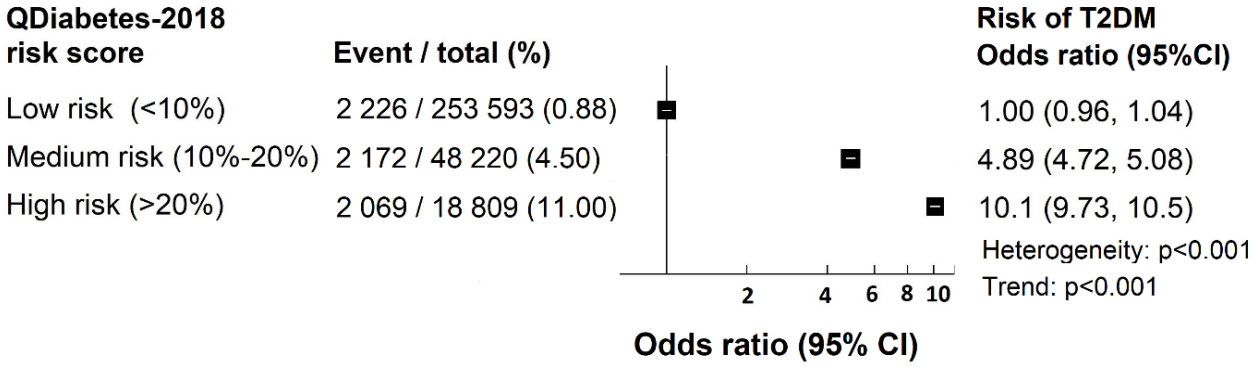


Figure 6:4 | Risk of incident T2DM in three risk-stratified subgroups (by QDiabetes-2018 risk score)

The model includes 6,467 cases of incident T2DM and 314,155 controls. All participants were white British and passed genetic QC. QDiabetes-2018 risk score is calculated with ClinRisk Ltd algorithm (Hippisley-Cox, Coupland 2017). The confidence intervals were estimated using a floating absolute risk method. The absolute 10-year risk of T2DM was estimated as 19%.

6.3.2.3 *HMGCR* GRS and the risk of incident T2DM by QDiabetes-2018 risk score

There was no difference in the effects of *HMGCR* on the risk of incident T2DM between the subgroups of participants with a different absolute risk of T2DM (**Figure 6:5)**. None of the associations was statistically significant. The effect estimates were directionally discordant, showing the highest risk of *HMGCR*-associated incident T2DM in the subgroup with the lowest absolute risk of T2DM. The risk further decreased with an increase in absolute risk, showing a non-significant protective effect in the participants at the highest absolute 10-year risk of incident T2DM. The test for trend was statistically significant (P_{trend}=0.032) and confirmed the systematic change in the *HMGCR*-associated risk of incident T2DM across the ordered QDiabetes-2018 subgroups.

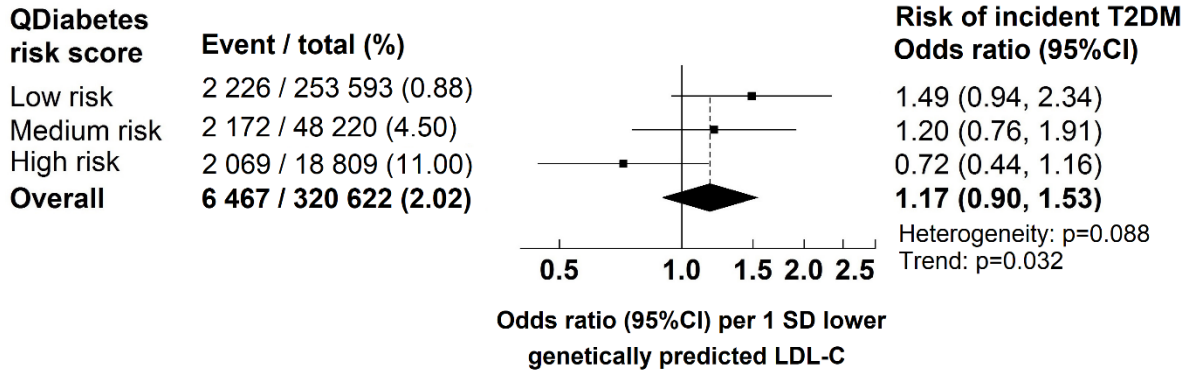


Figure 6:5 | Risk of incident T2DM associated with 1 SD lower genetically predicted lower LDL-C (by QDiabetes-2018 risk groups)
The model includes 6,467 cases of incident T2DM and 314,155 controls. All participants were white British who passed genetic QC. QDiabetes-2018 risk score was calculated with ClinRisk Ltd algorithm (Hippisley-Cox, Coupland 2017).

6.3.3 *HMGCR* GRS and the risk of incident T2DM by HbA1c levels

The comparison of the models stratified by HbA1c levels in natural units (mmol/mol), which is potentially prone to a collider-stratification bias, and a model stratified by the

residual collider, showed that in this particular case stratifying on a variable affected by the instrumental variable and by the outcome did not introduce collider-stratification bias. Estimates were not statistically significant in healthy and pre-diabetes subgroups, for both HbA1c and residual HbA1c. There was no heterogeneity in the effect between the groups. Participants in the pre-diabetic group did not have an excess risk of incident T2DM associated with the *HMGCR* GRS (**Figure 6:6 and**

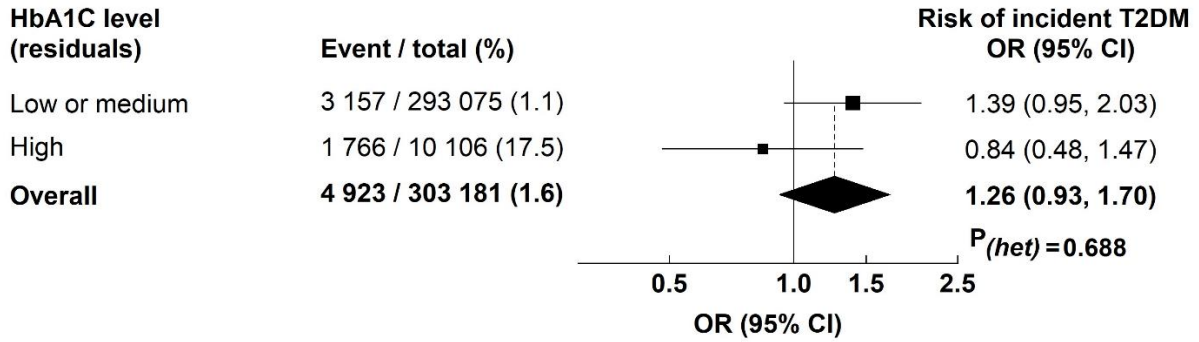


Figure 6:7).

The residual HbA1c was highly correlated with the HbA1c in mmol/mol (Appendix 6.5); both variables' distributions were the same shape (Appendix 6.4).

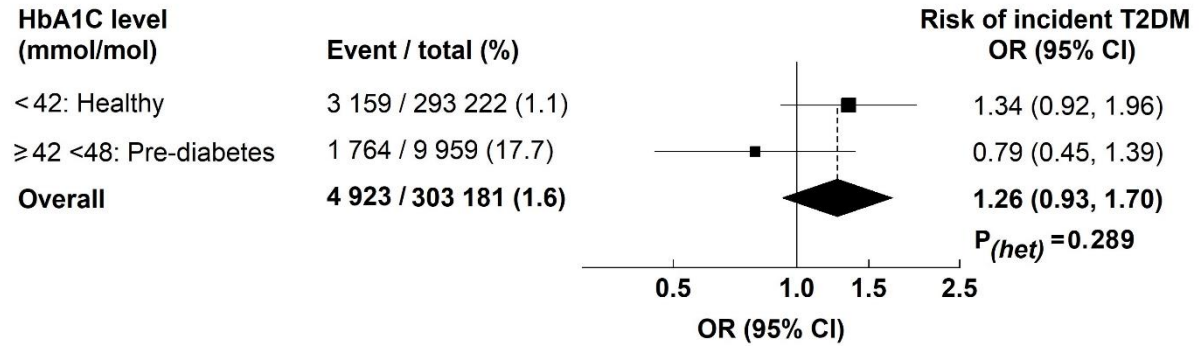


Figure 6:6 | Association of *HMGCR* GRS with incident T2DM by HbA1c (mmol/mol) level

A total of 6,467 cases of incident T2DM were identified in the UK Biobank. After the exclusion of participants with missing HbA1c levels (n=15,785), and those with HbA1c ≥48 (n=2,268, there were 4,923 incident T2DM available for the analysis. The subgroup analysis was

stratified according to HbA1c level (mmol/mol) including 290,063 healthy participants, and 8,195 participants with pre-diabetes.

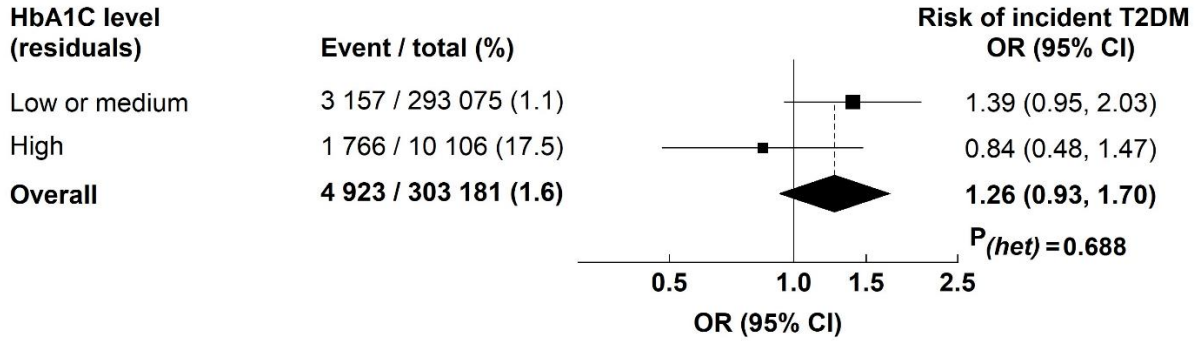


Figure 6:7 | Association of *HMGCR* GRS with incident T2DM by HbA1c (residuals adjusted for *HMGCR* GRS)
The subgroup analysis was stratified according to residual HbA1c, including 290,063 healthy participants, and 8,195 participants with pre-diabetes.

6.3.4 *HMGCR* GRS and glycaemic traits in participants with different risks of T2DM

6.3.4.1 *HMGCR* GRS and HbA1c

As already reported in Chapter 5, the association between *HMGCR* GRS and HbA1c was positive; 1 SD lower genetically predicted LDL-C was associated with 0.05 SD (95% CI 0.02, 0.08) higher HbA1c. These associations were directionally concordant in all three groups with different genetic risk of T2DM. There was no evidence for heterogeneity (P_{het}=0.378) or trend (P_{trend}=0.186), suggesting that the association between *HMGCR* GRS and HbA1c was not modified by the level of genetic susceptibility to T2DM (**Figure 6:8**).

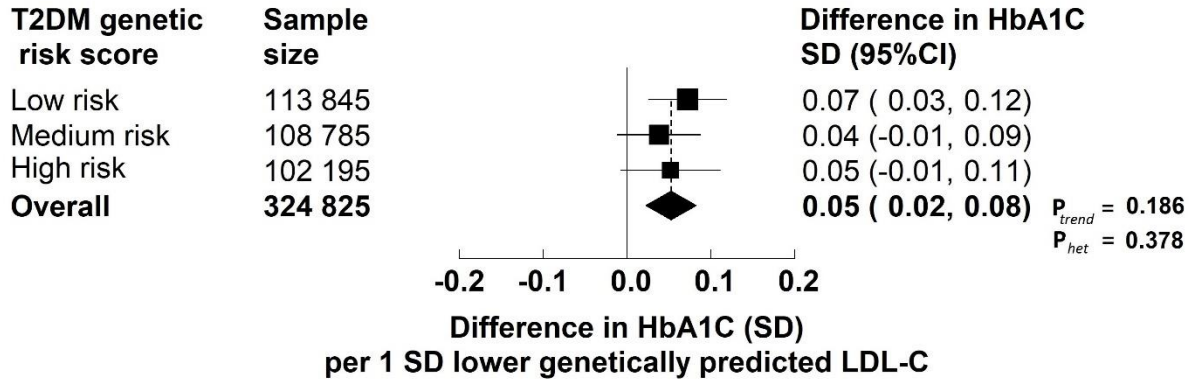


Figure 6:8 | Associations of *HMGCR* GRS with HbA1c concentration in the subgroups by tertiles of T2DM GRS

HMGCR GRS is a continuous genetic risk score (comprised of 5 HMGCR SNPs). A high HMGCR score is a proxy of statin therapy. T2DM genetic risk score is comprised of 56 SNPs identified by Morris et al. (2012). The estimate is shown in mmol/mol per 1 SD genetically predicted change in LDL-C. The model includes 10,738 cases of T2DM. 9,690 participants who reported use of any oral hypoglycaemic medications or insulin were excluded from the analysis.

6.3.4.2 *HMGCR* GRS and non-fasting glucose

There was no association between *HMGCR* GRS and the levels of non-fasting blood glucose (0.01 [95% CI -0.03, 0.04] SD) per 1 SD lower genetically predicted LDL-C. The subgroup analysis showed no heterogeneity between the effects across all three risk groups (P_{het}=0.234), implying that baseline risk of T2DM does not modify the effect of *HMGCR* on blood glucose (**Figure 6:9**).

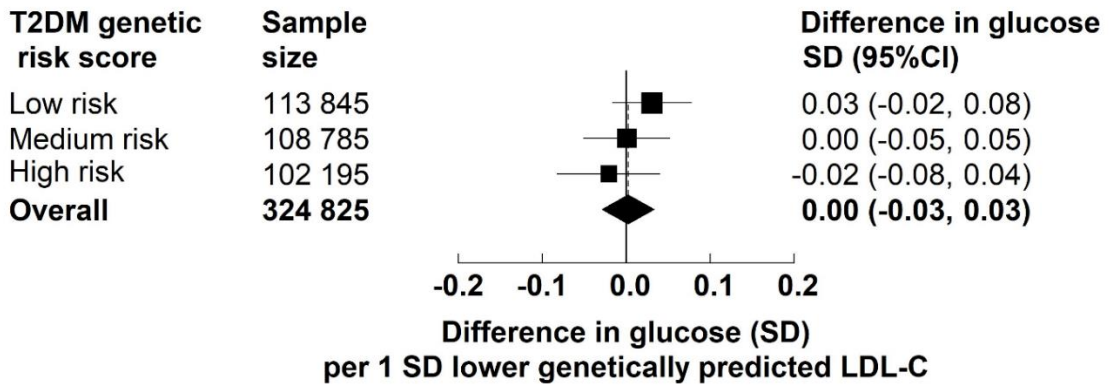


Figure 6:9 | Associations of *HMGCR* GRS with blood glucose level concentration in the subgroups by tertiles of T2DM GRS

HMGCR GRS is a continuous genetic risk score (comprised of 5 HMGCR SNPs). A high HMGCR score is a proxy of statin therapy. T2DM genetic risk score is comprised of 56 SNPs identified by Morris et al. (2012). The estimate is shown in mmol/L per one SD genetically predicted change in LDL-C. The model includes 10,738 cases of T2DM. 9,690 participants who reported use of any oral hypoglycaemic medications or insulin were excluded from the analysis.

6.3.5 *HMGCR* GRS and anthropometric measures in participants with different risks of T2DM

6.3.6 *HMGCR* GRS and anthropometric measures

As previously reported, the *HMGCR* GRS was associated with higher adiposity measures, 1 SD lower genetically predicted LDL-C was associated with 0.14 SD (95% CI 0.11, 0.18) higher body weight, 0.12 SD (95% CI 0.09, 0.16) greater waist circumference, and 0.15 SD (95% CI 0.11, 0.18) higher BMI (Chapter 5, section 5.3.4). The subgroup analysis showed no statistically significant interactions between the effects of *HMGCR* GRS on anthropometric measures across the levels of genetic susceptibility to T2DM (**Figure 6:10**).

6.3.6.1 *HMGCR* GRS and anthropometric measures by sex

There was no evidence for heterogeneity of the effect of *HMGCR* GRS on anthropometric measures among the levels of genetic susceptibility to T2DM in men (**Figure 6:11**) or in women (**Figure 6:12**). Tests for trend for each anthropometric trait failed to show any systematic difference in the effect of *HMGCR* GRS on anthropometric measures according to genetic susceptibility to T2DM. The baseline genetic risk of T2DM did not modify the effect of genetically predicted lower LDL-C on adiposity in men or in women.

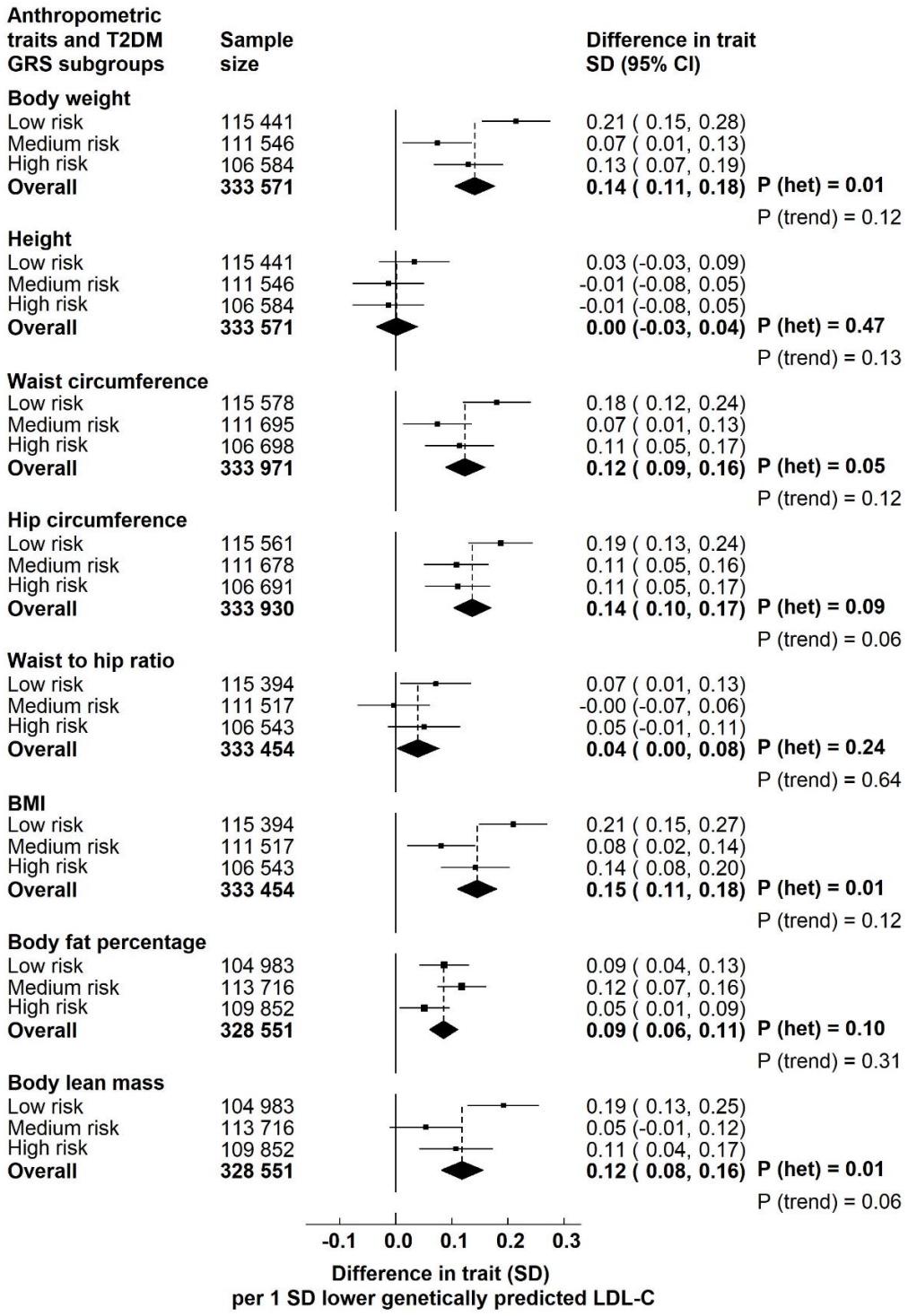


Figure 6:10 | Associations of *HMGCR* GRS with anthropometric measures stratified by T2DM genetic risk groups per 1 SD *HMGCR* predicted lower LDL-C.

SD units (men/women): body weight 14.2/13.9 kg, waist circumference 11.3/12.5 cm, BMI 4.2/5.1 kg/m^2 , waist-to-hip ratio 0.1/0.1, body fat percentage 5.8/6.9%, hip circumference 7.5/10.3 cm, body lean mass 7.7/5.0 kg.

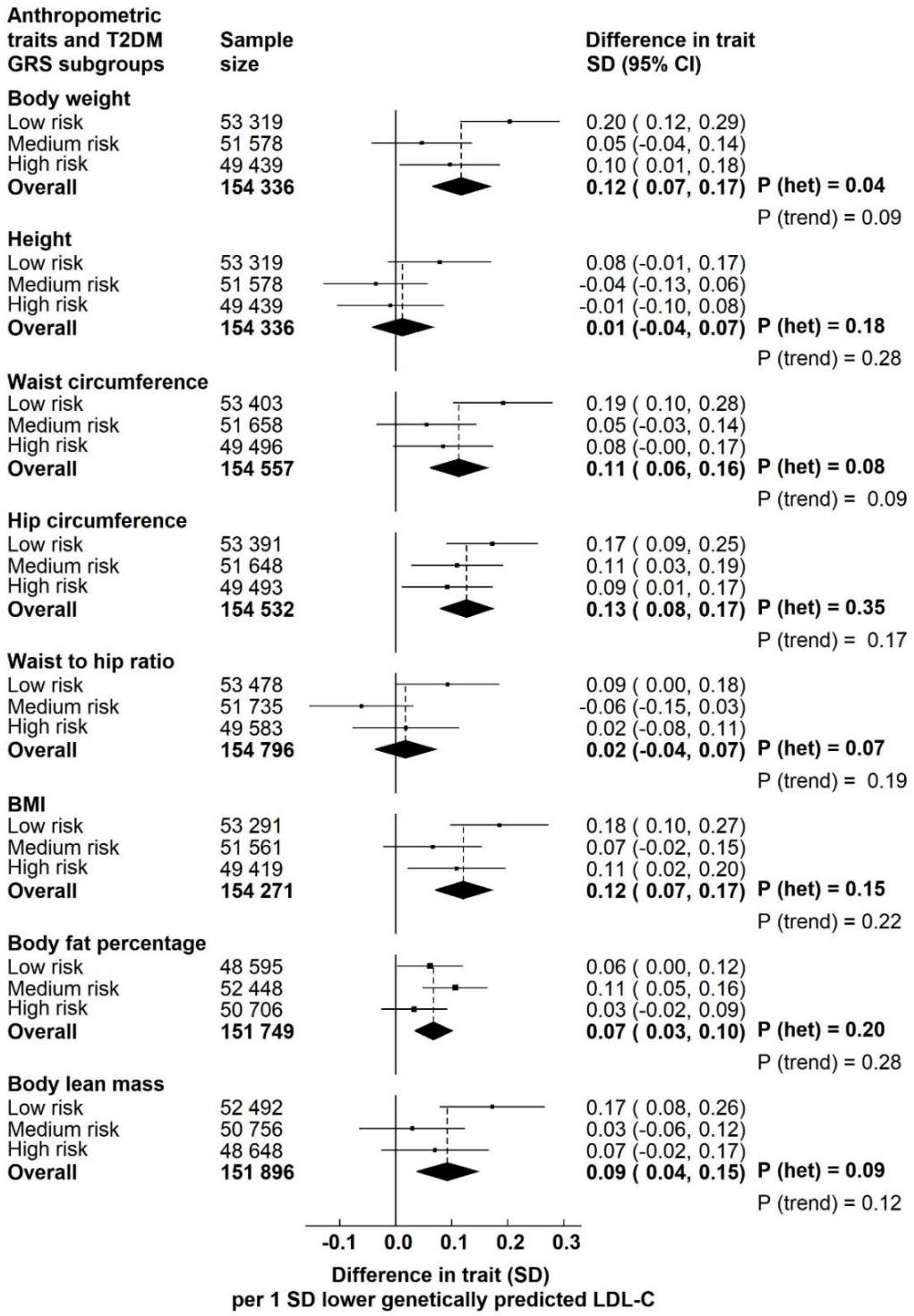


Figure 6:11 | Associations of *HMGCR* GRS with anthropometric measures stratified by T2DM genetic risk groups in MEN per 1 SD lower genetically predicted LDL-C

SD units for men: body weight 14.2 kg, waist circumference 11.3 cm, BMI 4.2 kg/m^2 , waist-to-hip ratio 0.1, body fat percentage 5.8%, hip circumference 7.5 cm, body lean mass 7.7 kg.

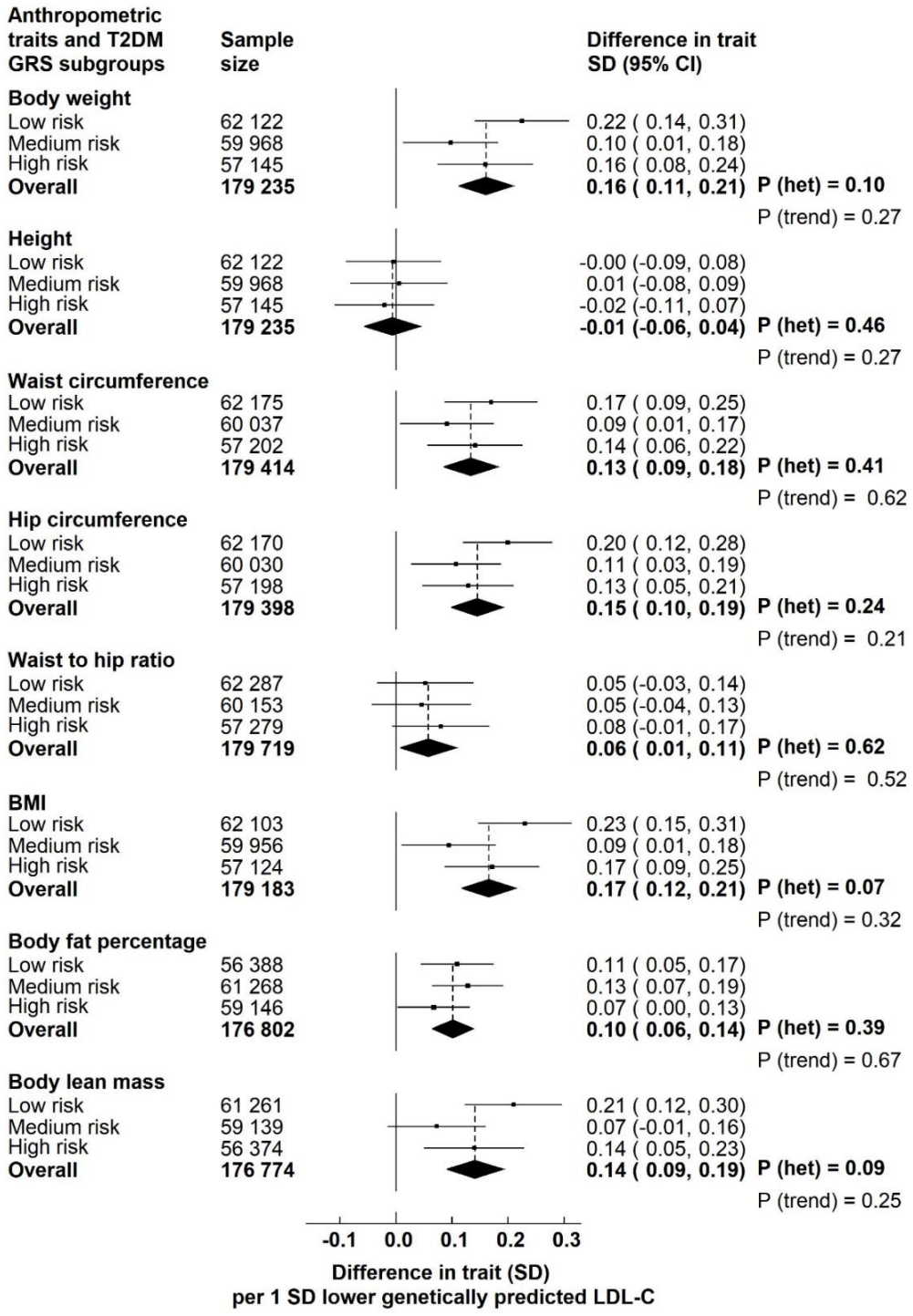


Figure 6:12 | Associations of *HMGCR* GRS with anthropometric measures stratified by T2DM genetic risk groups in WOMEN per 1 SD lower genetically predicted LDL-C

SD units for women: body weight 13.9 kg, waist circumference 12.5 cm, BMI 5.1 kg/m^2, waist-to-hip ratio 0.1, body fat percentage 6.9%, hip circumference 10.3 cm, body lean mass 5.0 kg.

6.4 Discussion

6.4.1 The effect of *HMGCR* GRS on T2DM

Results in the earlier chapters have shown that a *HMGCR* GRS associated with lower LDL-C led to an increased risk of T2DM, higher HbA1c and higher adiposity measurements. To investigate any heterogeneity of the effect of *HMGCR* GRS with varying susceptibility (both genetic and clinical) to T2DM, subgroup analyses were performed.

It has previously been reported in the literature that the risk of statin-associated incident T2DM was higher in participants with clinical risk factors for developing diabetes (e.g. fasting glucose (FG) > 5.6 mmol/L) compared to participants with normal glycaemic control (**Figure 6:13**). However, the current study showed no such relationship for subgroups defined based on genetic T2DM risk and HbA1c categories, or a weak trend in the opposite direction for subgroups based on QDiabetes-2018 risk score.

The effect of *HMGCR* GRS on the risk of T2DM across subgroups based on genetic risk of T2DM was consistent with the overall result. There was no statistically significant heterogeneity in the effect across these subgroups. The number of events in each subgroup was much higher than in any previous analyses[87]; the low genetic T2DM risk group had 4,432 events, while the medium and high risk groups had 6,668 and 9,261 events respectively. However, the polygenic T2DM risk score which was used in this study distinguished risk groups based on the individual genetic susceptibility to T2DM rather than on HbA1c or fasting glucose level, as has been

done in the previous studies[86,87]. The advantage of the polygenic T2DM risk score approach is the use of all participants' data, however the disadvantage is that the risk of T2DM across the subgroups is only modest.

In a secondary analysis, to assess whether heterogeneity was present according to clinical risk of incident T2DM, participants who were free from T2DM at baseline were split into three risk groups using the QDiabetes-2018 algorithm. The QDiabetes-2018 risk score as a factor variable has high discriminatory accuracy; participants in the high-risk group had a 10 times higher relative risk of incident T2DM than those in the low risk group (the absolute risks of developing incident T2DM in the low, medium and high risk groups were 0.8%, 4.5%, and 11.0% respectively). There was no material difference in the proportional effect of *HMGCR*-predicted lower LDL-C on the risk of incident T2DM between groups with different clinical risks of incident T2DM.

In contrast to the current results, post hoc analyses of three RCTs[87,141], (as one of the publications includes two RCTs), assessed the effect of statins on the risk of T2DM in the patients with and without risk factors for T2DM, and concluded that the risk of statin-induced T2DM is limited to those participants with risk factors.

The same studies also reported the overall effect of statins on incident T2DM by fasting glucose levels. In both studies, ORs for incident T2DM in participants with prediabetes (FG >5.6 mmol/L) were statistically significant and the point estimate was higher than in participants with normal glucose regulation. Although both studies concluded that the statin-induced risk of incident T2DM was limited to pre-diabetic participants, heterogeneity tests demonstrate no material difference between the subgroups (**Figure 6:13**).

Figure 6:13 below summarises the effects of statins on the risk of incident T2DM in subgroups of diabetes risk from these RCTs to show the extent to which the confidence intervals for the findings within subgroups actually overlap. Similarly, the current study found no difference in the effect of *HMGCR* GRS on incident T2DM in participants with normal glucose regulation (

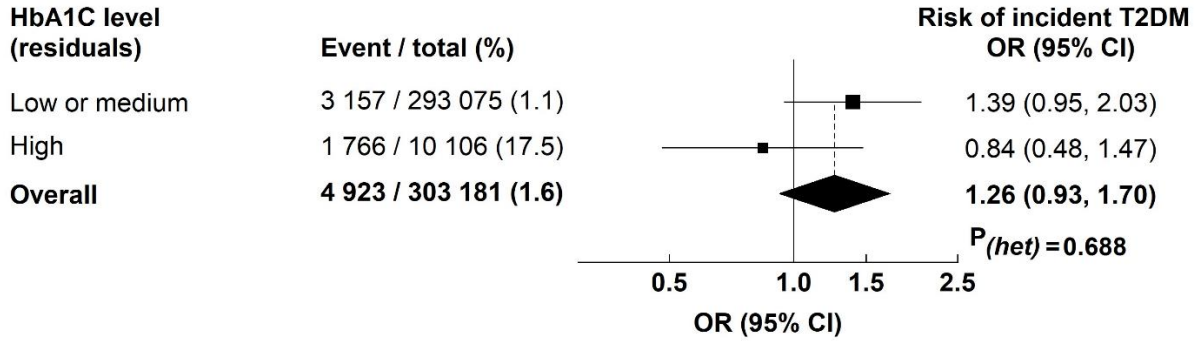


Figure 6:7). Unlike the previous studies, however, the subgroups in this study were defined by HbA1c rather than fasting glucose level. Comparing the subgroups based on fasting glucose and HbA1c is not ideal but still can be used, as HbA1c shows the average blood glucose level over the last 10-12 weeks[149].

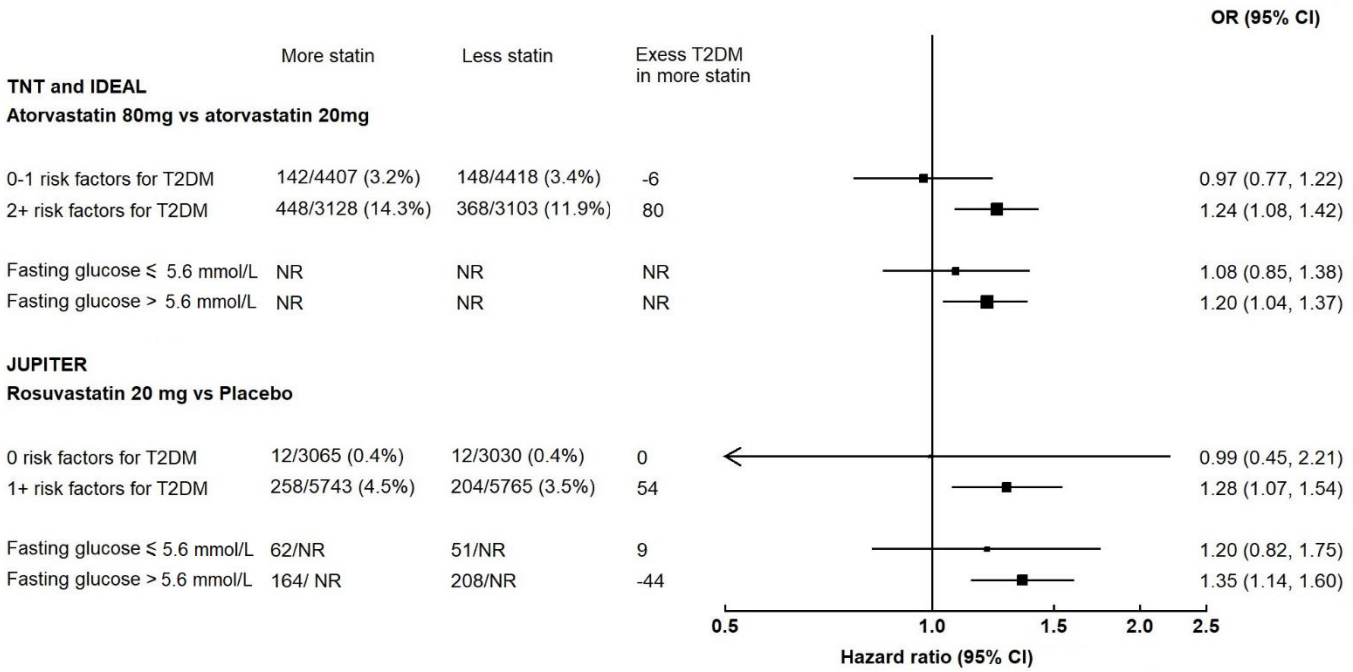


Figure 6:13 | Risk of incident T2DM in subgroups of participants with and without risk factors for T2DM from the TNT+IDEAL and JUPITER RCTs
NR – not reported. Pre-diabetes status in TNT and IDEAL[86] and JUPITER[87] RCTs was defined as fasting glucose level > 100 mg/dl.

When stratified by HbA1c level **(Figure 6:6**), the effect of *HMGCR* GRS on incident T2DM was consistent with the previous analysis (**Figure 6:5)**. Pre-diabetic participants with HbA1c levels over 42 mmol/mol did not have any excess proportional risk of T2DM associated with *HMGCR* GRS**.** On the contrary, the result suggested a trend towards lower proportional risk of incident T2DM associated with the *HMGCR* GRS in those with higher clinical risk of T2DM **(Figure 6:5**).

The results of the present investigation therefore do not support previous studies claiming that HMG CoA reductase inhibition increases the risk of T2DM only in patients with impaired glucose tolerance or other T2DM risk factors. In contrast to Ference's genetic study[51], which claimed that a higher proportional risk of *HMGCR*-induced incident T2DM is restricted to people with impaired fasting glucose (**Figure 6:14**),

there was no heterogeneity present in the current study (**Figure 6:5**). A trend was observed in the secondary analysis, demonstrating a decrease (not an increase) in the risk of incident T2DM associated with *HMGCR* GRS with an increase in QDiabetes-2018 risk score.

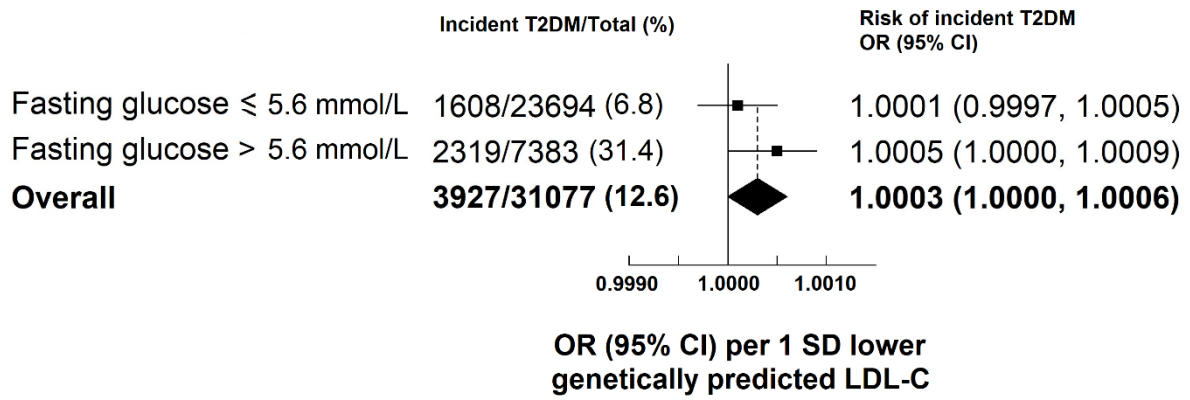


Figure 6:14 | Associations of *HMGCR* GRS with incident T2DM in participants with normal and high levels of fasting glucose
The figure is modified and adopted from Ference et al.[51]

The consistency in the patterns of associations in those at different risk of T2DM, considering both lifelong risk as estimated by a T2DM polygenic risk score and absolute risk of T2DM based on QDiabetes-2018, supports the robustness of the findings. Moreover, a further subgroup analysis by HbA1c levels showed no material difference in the proportional risk of T2DM associated with *HMGCR* GRS. There was no excess proportional risk of incident T2DM associated with *HMGCR* GRS in participants with pre-diabetic levels of HbA1c compared to those with normal levels.

There was a substantial difference in the data available for the present study and that of Ference[51]. In the subgroup analysis stratified by fasting plasma glucose levels, Ference *et al.*[51] analysed data for 31,077 participants without prevalent T2DM. There

were 2,319 incident cases of T2DM in the group with impaired fasting glucose (>100 mg/dl), and 1,608 cases in the group with normal fasting glucose (<100 mg/dl). By contrast, 320,622 participants without prevalent T2DM, including 6,467 cases of incident T2DM were available for the current analysis. The confidence intervals of the associations between the *HMGCR* GRS and incident T2DM in Ference's analysis were largely overlapping (OR 1.96 95% CI 1.00, 3.77 and OR 1.16, 95% CI 0.63, 2.12 in the pre-diabetes and healthy subgroups respectively) and, in the high-risk group, where the effect was observed, the lower limit of confidence interval nearly crossed 1 (OR 1.96; 95% CI 1.00, 3.77, p-value = 0.05). Had a test for heterogeneity been applied, there would not have been any statistical difference present between the subgroups. Nonetheless, despite similar proportional risks of T2DM due to *HMGCR* in people at differing risks of T2DM, those with pre-diabetes are at much greater risk of developing T2DM and so most of the statin-associated excess does arise in such people.

The main clinical implication of this study is to provide reliable evidence that statin therapy is unlikely to impose an extra proportional risk of T2DM only in patients who already have pre-diabetes or risk factors for T2DM. Instead, these results suggest that risk increases in all participants, and there may in fact be a slight trend toward a smaller proportional increase in risk in people with T2DM risk factors than in those without. Patients with pre-diabetes and diabetes are at higher risk of developing CVD; therefore, they should not refuse statin prescriptions because of its modest diabetogenic effect. Instead, given the current results, patients who are prescribed statins should be screened for diabetes and those for whom a slight increase in HbA1c means a transition towards diabetes should be regularly monitored using HbA1c levels and provided extra advice on glucose control and lifestyle modification.

6.4.2 The effects of the *HMGCR* GRS on glycaemic traits and anthropometric measures by risk of diabetes

There was no difference in the effect of *HMGCR* GRS on HbA1c and non-fasting glucose levels across the subgroups with different genetic risks of T2DM. These associations have not been reported in the existing literature.

One of the possible biological pathways explaining the effect of HMGCR inhibition on the risk of T2DM is through increasing ectopic and abdominal adipose tissue deposits and thereby impairing glucose homeostasis. Both BMI and waist circumference are strong predictors of T2DM. The European Prospective Investigation into Cancer and Nutrition (EPIC)[150] Potsdam study with 25,244 European participants, including 1008 cases of incident T2DM, showed that 1 kg/m^2 higher BMI was associated with 21% and 15% higher risk of T2DM in men and women respectively, and 1 cm higher waist circumference was associated with 8% higher risk of T2DM in both sexes[150]. The previous chapter has shown that *HMGCR* GRS is associated with higher measures of central and general adiposity. This chapter assessed the heterogeneity of this effect across subgroups of participants with different levels of genetic risk of T2DM. The effect on anthropometric measures, stratified by T2DM genetic risk and by sex, showed concordant results. No heterogeneity was observed among the subgroups, implying no effect modification by the baseline risk of T2DM. The current study showed that adipose tissue accumulation associated with the effect of *HMGCR* GRS does not differ in people with normal and impaired glucose homeostasis.

6.5 Summary

This subgroup analysis found no heterogeneity in the effect of *HMGCR* GRS on the proportional risk of total T2DM between participants at low, medium and high genetic risk of T2DM, as defined by a polygenic T2DM risk score. Similarly, the secondary analysis found no statistical heterogeneity in the effect of *HMGCR* GRS on incident T2DM across the low, medium and high clinical T2DM risk groups stratified by QDiabetes-2018 risk score. However, there was a weak trend towards lower proportional risk of *HMGCR*-associated T2DM in those at higher risk of T2DM.

The additional analysis of the participants without diabetes at baseline also found no evidence of excess proportional risk of incident T2DM associated with *HMGCR* GRS in the pre-diabetes subgroup as defined by HbA1c. Genetic risk of T2DM did not modify the effect of *HMGCR* GRS on HbA1c, non-fasting glucose concentrations and anthropometric measures.

In summary, none of the subgroup analyses including by *HMGCR* GRS, QDiabetes-2018 risk score and by HbA1c levels found evidence that patients at higher risk of T2DM experience a higher proportional risk of statin-induced transition to T2DM, or accumulate more adipose tissue, than those at lower risk of T2DM.

Chapter 7 The associations of the *NPC1L1* GRS with risk of T2DM, glycaemic traits and anthropometric measures

7.1 Introduction

The purpose of this chapter is to investigate the potential effects of a non-statin LDL-C lowering drug, ezetimibe, on total T2DM and related traits. Ezetimibe is the second most prescribed LDL-C lowering medication after statins[151]. It lowers LDL-C by inhibiting cholesterol absorption in the small intestine by binding to the Niemann-Pick C1-Like 1 protein. Genetic variants in the *NPC1L1* gene affect the function and expression of this protein and could be used as a proxy for examining the effect of pharmacological perturbation of the NPC1L1 drug target with ezetimibe on the risk of T2DM and related traits.

RCTs of ezetimibe showed evidence of the LDL-C lowering effects and CVD benefits of ezetimibe. The largest RCT of ezetimibe to date is IMPROVE-IT which randomised 18,144 participants, who had experienced an acute coronary syndrome event within 10 days of recruitment, to ezetimibe 10mg/simvastatin 40mg or placebo/simvastatin 40mg. At 1 year of treatment, the mean LDL-C level in the ezetimibe/simvastatin group was 1.8 mmol/L vs. 1.4 mmol/L in the placebo/simvastatin group. The primary outcome (cardiovascular death, myocardial infarction, or coronary revascularization) was significantly lower in the ezetimibe/simvastatin therapy group compared to the placebo/simvastatin group (HR 0.94, 95% CI 0.89, 0.99)[23].

The ENHANCE RCT[152] compared the effect of the addition of ezetimibe or placebo to simvastatin (80 mg) in 720 participants with familial hypercholesterolemia. The

addition of ezetimibe demonstrated a greater reduction in LDL-C levels compared to the placebo-controlled group. A between group difference in LDL-C levels after 24 months of therapy was 16.5%. A meta-analysis[153] of 14 RCTs with 2,007 participants randomised to high-intensity statin monotherapy or high-intensity statin plus ezetimibe found that ezetimibe lowered LDL-C by 14% (95% CI 10.22, 17.78).

Similarly, genetic studies of *NPC1L1* have demonstrated the associations between genetic variants in the *NPC1L1* gene with LDL-C and the risk of CVD. A study by the Myocardial Infarction Genetic Consortium[154] used genetic data from 22,092 participants from nine studies to identify 15 *NPC1L1* inactivating mutations. These inactivation mutations were genotyped in an independent sample of 91,002 mixed-race participants. Heterozygous carriers of one *NPC1L1* inactivating mutation had lower LDL-C levels than non-carriers (mean difference 0.31 mmol/L, P=0.04) and lower risk of coronary heart disease (OR 0.47 95% CI 0.25, 0.87). Another meta-analysis of genetic association studies[30] with 321,044 participants showed that a *NPC1L1* GRS based on two LDL-C lowering polymorphisms was associated with lower risk of coronary heart disease (OR 0.61 95% CI 0.42, 0.88 per 1 SD lower genetically predicted LDL-C).

While there has been a growing body of evidence of ezetimibe's effects on LDL-C and CVD, very little is known about its effect on the risk of T2DM and related traits. Genetic studies and RCTs were inconclusive about possible glycaemic effects of ezetimibe and its genetic proxy. The IMPROVE-IT RCT is the only RCT which has reported the effect of combination therapy (ezetimibe with statin) on T2DM compared to placebo with statin. Although the effect of co-administration of ezetimibe and simvastatin on the risk of incident T2DM was not pre-specified as one of the safety events, it was

reported at the American Heart Association Scientific Sessions in 2014[83]. There was no material difference in risk of incident T2DM between treatment and control arms (HR 1.04, 95% CI 0.94, 1.15). A small meta-analysis[155] of six studies with 342 participants analysed the effect of ezetimibe therapy on HbA1c levels. Pooled data from ezetimibe vs placebo RCT and a combination therapy with ezetimibe plus statin vs the same statin RCT showed no effect of ezetimibe on HbA1c (WMD 0.76 mmol/mol, 95% CI −0.77, 2.19).

A meta-analysis of genetic studies[30], however, showed a contrasting association - genetic variants in the *NPC1L1* gene were associated with a higher risk of T2DM. The *NPC1L1* GRS based on two LDL-C lowering polymorphisms was associated with a substantial elevation in the risk of T2DM (OR 2.42, 95% CI 1.70, 3.43 per 1 SD lower genetically predicted LDL-C). The same study showed that the *NPC1L1* GRS was not associated with measures of adiposity. One SD lower genetically predicted LDL-C was associated with 0.08 (95% CI -0.04, 0.20) SD difference in BMI, and -0.01 (95% CI -0.14, 0.12) SD difference in waist-to-hip ratio.

Considering the inconsistent evidence from RCTs and genetic studies, as well as the very limited number of studies examining the effects of ezetimibe and its genetic proxy on the risk of T2DM, this chapter aims to (1) investigate the effect of the *NPC1L1* GRS on total T2DM; (2) examine the effect of *NPC1L1* GRS on glycaemic traits and anthropometric measures in men and women.

7.2 Study design and method

7.2.1 Study population

Genetic analysis was conducted among 334,516 unrelated white British UK Biobank participants. The study population has been described in detail in Chapter 5.2.1. For the external validation analysis, data from the DIAGRAM, GIANT, and MAGIC consortia were used (see Chapter 3.4).

7.2.2 Baseline assessment

A detailed description of the baseline assessment, including participant characteristics and medical history, was presented in Chapter 3.1, and baseline characteristics were shown in Chapter 5.2.1. Similar to the *HMGCR* analysis, this chapter uses total T2DM (20,361 cases) and incident T2DM (6,467 cases) as primary outcomes, HbA1c levels (available for 309,450 participants who were not using hypoglycaemic medications) for analyses of the effects of *NPC1L1* on glycaemia, and body weight, height, waist circumference, hip circumference, waist-to-hip ratio, BMI, body fat percentage and lean body mass available for the full study population for analyses of the associations with anthropometric measures.

7.2.3 *NPC1L1* genetic risk score

For this analysis, the algorithm described in in chapter 3.5 was repeated. SNPs within and in close proximity to the *NPC1L1* gene were selected, as these variants are involved in the expression of the NPC1L1 protein[156], which is the drug target of ezetimibe.

Those genetic variants in the vicinity of the *NPC1L1* gene include common genetic variants within the *NPC1L1* encoding zone, which directly affect *NPC1L1* transcription, and intergenic variants located in proximal promoter region, which regulate the level of the *NPC1L1* gene expression.

As the association between genetic variants and the level of NPC1L1 protein has not been reported, the selection of genetic variants for *NPC1L1* GRS was based on their associations with a biomarker directly affected by the protein, namely LDL-C. The association between *NPC1L1* SNPs and LDL-C was reported in the GLGC GWAS[45].

The process of selecting SNPs for the *NPC1L1* GRS has been fully described in Chapter 3.

7.2.4 Statistical methods

The statistical tests used to assess the associations between *NPC1L1* genetic variants/ *NPC1L1* GRS and the risk of T2DM, glycaemic traits and anthropometric measures have been fully discussed in Chapter 3.

7.3 Results

7.3.1 *NPC1L1* GRS and risk of T2DM

7.3.1.1 Individual SNPs and risk of T2DM

Both index SNPs selected for *NPC1L1* GRS were associated with lower LDL-C as shown in **Figure 7:1**. These SNPs' associations with the risk of total T2DM were directionally concordant. Rs2073547 and rs217386 were both associated with 1.0% higher risk of total T2DM. However, individually, these associations did not reach statistical significance (**Figure 7:2**). The estimates of the effects of the SNPs on LDL-C concentrations (as reported by GLGC) and ORs for total T2DM are shown in **Table 7:1**.

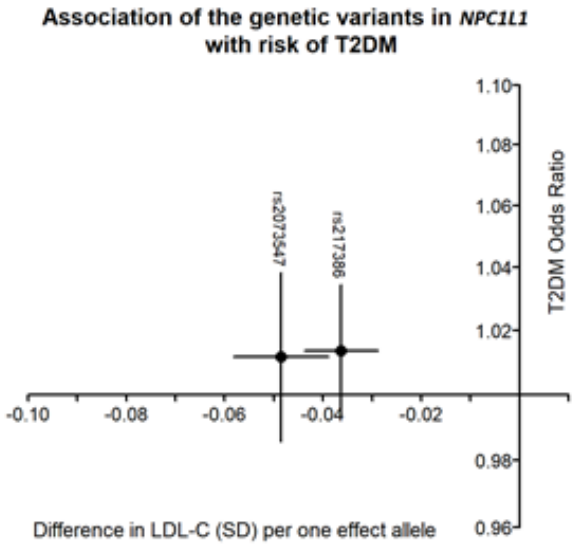


Figure 7:1 | Associations of the *NPC1L1* index SNPs with the risk of total T2DM and LDL-C concentrations
The effect on total T2DM (OR, 95% CI) and on LDL cholesterol (estimate, 95% CI) (SD) is shown per one copy of the LDL-C lowering allele, using the effect on LDL-C as reported by GLGC. Genotype data were available for 334,669 participants including 20,361 total cases of T2DM and 314,155 controls.

Table 7:1 | Associations of the *NPC1L1* index SNPs with total T2DM and LDL-C concentrations

RSID	Effect allele / other allele	Effect on T2DM OR (95% CI)	**LDL-C difference** (from GLGC) (95% CI)
rs2073547	A/G	1.012 (0.986, 1.038)	-0.049 (-0.058, -0.039)
rs217386	A/G	1.014 (0.993, 1.035)	-0.036 (-0.044, -0.294)

The effect on T2DM and LDL-C is shown per copy of the LDL-C lowering allele, the effect on LDL-C in SD units was reported by GLGC, the effect on T2DM is an Odds Ratio derived from the UK Biobank.

7.3.1.2 *NPC1L1* GRS and the risk of T2DM

There was no statistically significant effect of the *NPC1L1* GRS on risk of total T2DM (OR 1.25, 95% CI 0.90, 1.74) per 1 SD lower genetically predicted LDL-C. However, the magnitude of the effect (per 1 SD genetically predicted LDL-C) appeared consistent with *HMGCR (OR 1.26* 95% CI 1.08, 1.47). In the subgroup of participants who were not known to have T2DM at the time of recruitment, *NPC1L1* GRS was not associated with risk of incident T2DM (OR 0.98, 95% CI 0.55, 1.74) per 1 SD lower genetically predicted LDL-C.

There was no heterogeneity in the effects *NPC1L1* GRS on the risk of T2DM in men and women (P_{het}=0.764).

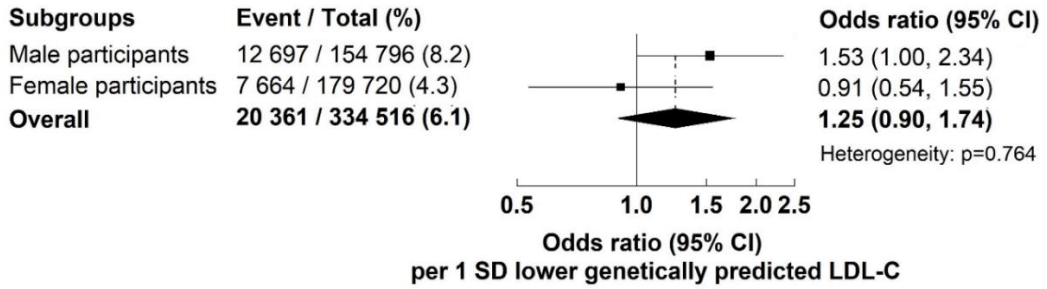


Figure 7:2 | Associations of the *NPC1L1* index SNPs with total T2DM stratified by sex

7.3.1.3 External validation analysis of the effect of *NPC1L1* on T2DM

The effect estimates of the individual *NPC1L1* SNPs on the risk of total T2DM as reported by the DIAGRAM consortium were directionally concordant with the effects observed in the UK Biobank (**Table 7:2** shows the associations of the *NPC1L1* SNPs with the risk of total T2DM in UK Biobank and the DIAGRAM consortium). In the DIAGRAM study, rs2073547 was associated with 5% (95% CI 1.02, 1.09) higher risk of total T2DM per one effect allele. The LDL-C lowering allele in rs217386 was associated with 3% (95% CI 1.00, 1.05) higher risk of total T2DM as reported by the DIAGRAM consortium (**Table 7:2**).

Table 7:2 | Associations of the *NPC1L1* index SNPs with total T2DM in UK Biobank and the DIAGRAM consortium

RSID	Effect allele /other allele	Risk of T2DM for the effect allele in DIAGRAM OR (95% CI)	Risk of T2DM for the effect allele in UK Biobank OR (95% CI)	LDL-C difference (from GLGC)
rs2073547	A/G	1.054 (1.022, 1.088)	1.012 (0.986, 1.038)	-0.049 (-0.058, -0.039)
rs217386	A/G	1.027 (1.003, 1.052)	1.014 (0.993, 1.035)	-0.036 (-0.044, -0.029)

The risk of total T2DM (odds ratio and 95% confidence intervals) is shown per one copy of the LDL-C lowering allele. The UK Biobank and the DIAGRAM consortium included 20,361 and 26,676 cases of total T2DM respectively.

In 2-sample inverse-variance weighted MR models conducted on summary level data from the DIAGRAM consortium, 1 SD lower genetically predicted LDL-C was associated with higher risk of total T2DM (OR 2.51, 95% CI 1.59, 3.96). When combined with the estimate computed in UK Biobank in a fixed effects meta-analysis, 1 SD lower genetically predicted LDL-C was associated with a 59% (95% CI 1.21, 2.08) higher risk of T2DM (**Figure 7:3**). There was heterogeneity between the results of the two studies (P_{het}=0.016), though the study-specific results were directionally concordant.

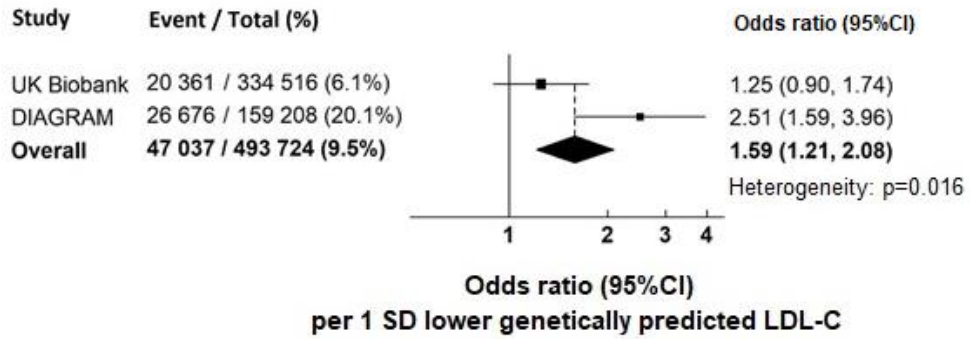


Figure 7:3 | Fixed-effects meta-analysis of the association between *NPC1L1* GRS with risk of total diabetes in a fixed effect model (UK Biobank and DIAGRAM consortium)

Analyses stratified by sex in the DIAGRAM consortium found no heterogeneity in the effect of *NPC1L1* GRS on the risk of T2DM in men and women (P_{het}=0.136). Unlike the UK Biobank result, the effect in the DIAGRAM consortium was both directionally concordant and statistically significant in both sexes (OR 2.00 95% CI 1.15, 3.50 and R 2.40 95% CI 1.28, 3.96 in men and women respectively) (**Figure 7:4**).

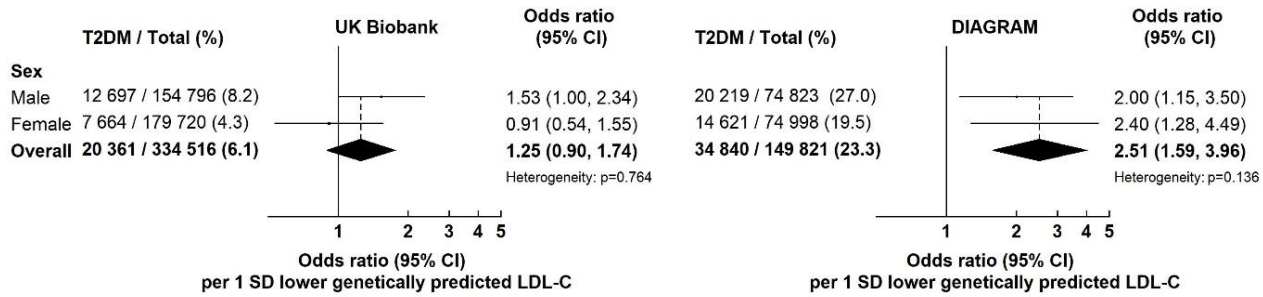


Figure 7:4 | Sex-specific effects of *NPC1L1* GRS on total T2DM in UK Biobank and the DIAGRAM consortium

7.3.2 *NPC1L1* GRS and glycaemic traits

7.3.2.1 The effect of *NPC1L1* on glycaemic traits

In UKB, both index SNPs in the *NPC1L1* GRS were associated with higher HbA1c levels (**Figure 7:5**). Each copy of the LDL-C lowering allele in rs2073547 was associated with 0.014 SD (95% CI 0.009, 0.019) higher HbA1c. 1 SD of HbA1c is equivalent to 6.2mmol/mol. Each copy of the LDL-C lowering allele in rs217386 was associated with 0.004 SD (95% CI 0.000, 0.008) higher HbA1c. There was no association between the index SNPs and non-fasting glucose concentrations (**Figure 7:6**). The effect size of each LDL-C lowering allele in rs2073547 and rs217386 was 0.003 SD (95% CI -0.002, 0.009) and -0.001 SD (95% CI -0.05, 0.003), respectively, for fasting glucose. 1 SD of non-fasting glucose is equivalent to 1.1mmol/L.

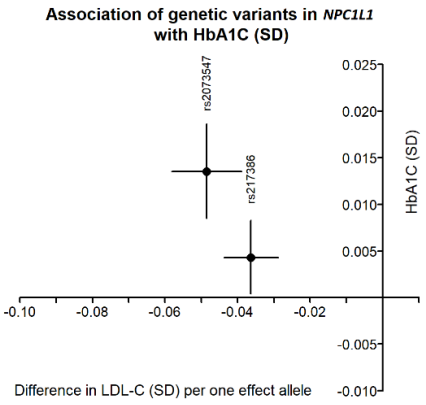


Figure 7:5 | Associations of the *NPC1L1* SNPs with HbA1c and LDL-C

Association is shown in SD units per 1 copy of the LDL-C lowering allele. 1 SD of HbA1c is equivalent to 6.2mmol/mol, 1 SD of LDL-C is equivalent to 1.0 mmol/L..

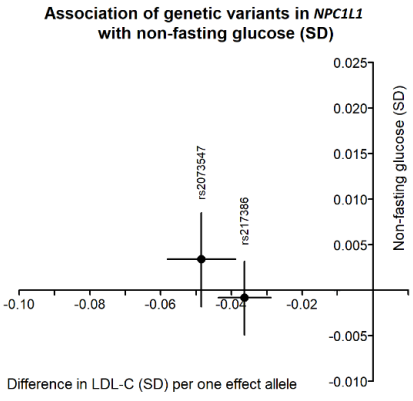


Figure 7:6 | Associations of the *NPC1L1* SNPs with non-fasting glucose and LDL-C.

Associations is shown in SD units per 1 copy of the LDL-C lowering allele. 1 SD of non-fasting glucose is equivalent to 1.1mmol/L, 1 SD of LDL-C is equivalent to 1.0 mmol/L..

The *NPC1L1* GRS was associated with higher HbA1c concentrations (0.15 SD 95% CI 0.09, 0.22 or 1.06 mmol/mol 95% CI 0.43,1.69 per 1 SD lower genetically predicted LDL-C) but was not associated with non-fasting glucose (**Figure 7:7**).

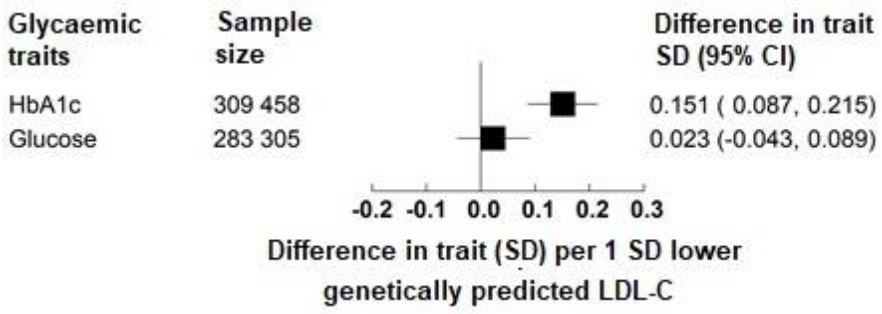


Figure 7:7 | Associations of *NPC1L1* GRS with glycaemic traits
The effect on glycaemic traits is shown in SD units per 1 SD lower genetically predicted LDL-C. Participants who reported use of diabetes medication were excluded from the analysis.

7.3.2.2 External validation analyses of the effect of *NPC1L1* on HbA1c

There was no heterogeneity in the effect of *NPC1L1* SNPs on HbA1c in UK Biobank and MAGIC. Estimates were quantitatively and directionally consistent (**Table 7:3**).

Table 7:3 | Effects of *NPC1L1* GRS on HbA1c in UK Biobank and the MAGIC consortium

	Effect allele/ other allele	MAGIC difference in trait, SD (95% CI)	UK Biobank difference in trait, SD (95% CI)	LDL-C difference (from GLGC)
HbA1c				
rs2073547	A/G	0.009 (0.004, 0.014)	0.014 (0.009, 0.019)	-0.049 (-0.058, -0.039)
rs217386	A/G	0.003 (-0.001, 0.007)	0.004 (0.000, 0.008)	-0.036 (-0.044, -0.029)

The effect is shown in SD (95% CI) units per 1 copy of effect allele.

In the MAGIC dataset, 1 SD lower genetically predicted LDL-C was associated with 0.14 SD (95% CI 0.06, 0.22) higher HbA1c. There was no statistical heterogeneity between MAGIC and UK Biobank results (0.14 [95% CI 0.06, 0.22] SD and 0.15 [95% CI 0.09, 0.22] SD respectively, P_{het}=0.645) (**Figure 7:8**).

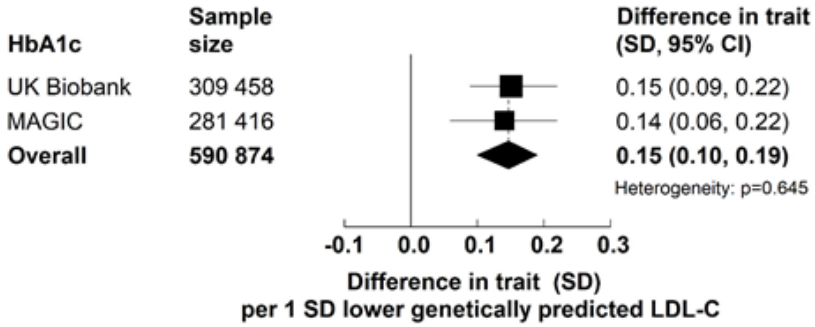


Figure 7:8 | Fixed-effect meta-analyses of the effect of *NPC1L1* GRS on HbA1c in UK Biobank and the MAGIC consortium

7.3.3 *NPC1L1* GRS and measures of central and general adiposity

7.3.3.1 Individual SNPs and measures of central and general adiposity

There was no clear heterogeneity between the SNPs effects on adiposity. Rs2073547 had no significant association with any trait, while rs217386 was associated with lower body weight, BMI, waist and hip circumferences, body fat, and lean body mass (**Figure 7:9**).

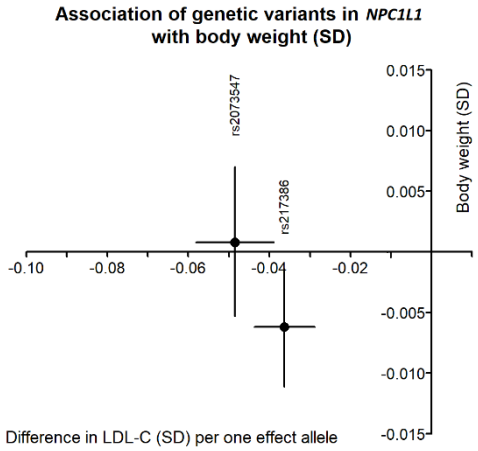


Figure 7:9A | Associations of the *NPC1L1* index SNPs with bodyweight (SD) and LDL-C (SD). The effect is shown per copy of the LDL-C lowering allele. 1 SD of body weight is equivalent to 15.9 kg, 1 SD of LDL-C is equivalent to 1.0 mmol/L.

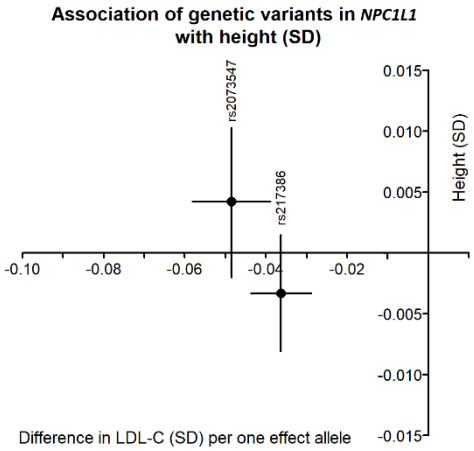


Figure 7:9B | Associations of the *NPC1L1* index SNPs with height (SD) and LDL-C (SD). The effect is shown per copy of the LDL-C lowering allele. 1 SD of height is equivalent to 9.2 cm, 1 SD of LDL-C is equivalent to 1.0 mmol/L.

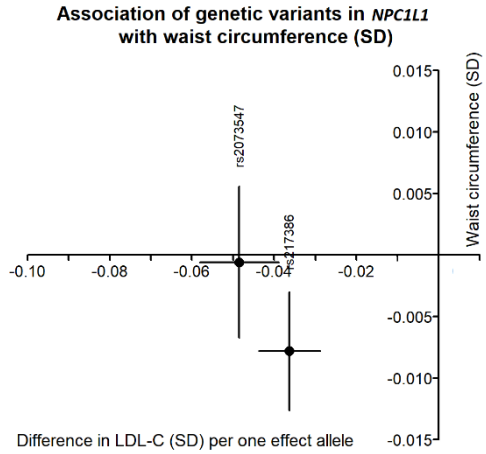


Figure 7:9C | Associations of the *NPC1L1* index SNPs with waist circumference (SD) and LDL-C (SD). The effect is shown per copy of the LDL-C lowering allele. 1 SD of waist circumference is equivalent to 13.4 cm, 1 SD of LDL-C is equivalent to 1.0 mmol/L.

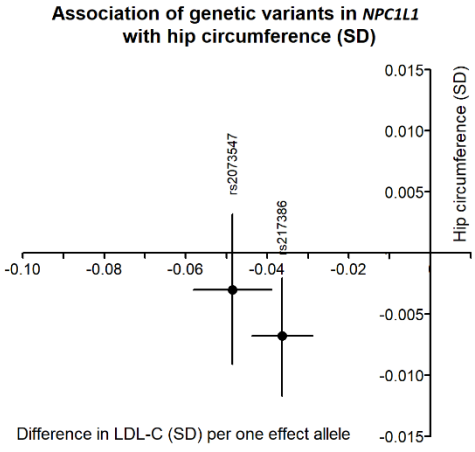


Figure 7:9D | Associations of the *NPC1L1* index SNPs with hip circumference (SD) and LDL-C (SD). The effect is shown per copy of the LDL-C lowering allele. 1 SD of hip circumference is equivalent to 9.1 cm, 1 SD of LDL-C is equivalent to 1.0 mmol/L.

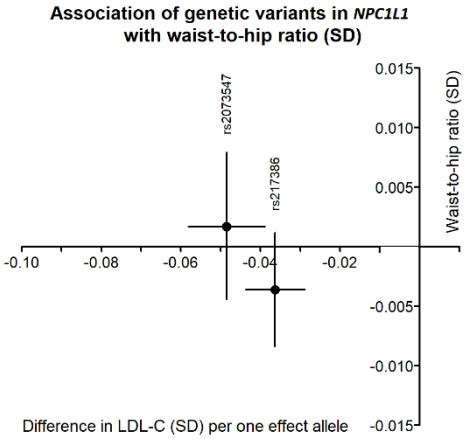


Figure 7:9E | Associations of the *NPC1L1* index SNPs with waist-to-hip ratio (SD) and LDL-C (SD). The effect is shown per copy of the LDL-C lowering allele. 1 SD of waist-to-hip ratio is equivalent to 0.10, 1 SD of LDL-C is equivalent to 1.0 mmol/L.

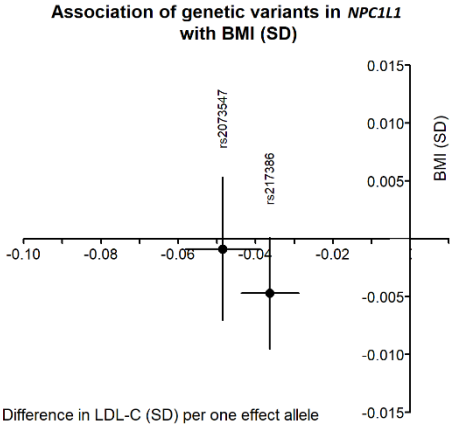


Figure 7:9F | Associations of the *NPC1L1* index SNPs with BMI (SD) and LDL-C (SD). The effect is shown per copy of the LDL-C lowering allele. 1 SD of BMI is equivalent to 5.0, 1 SD of LDL-C is equivalent to 1.0 mmol/L.

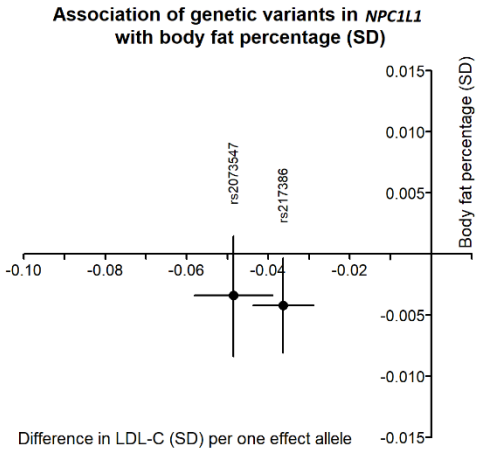


Figure 7:9G | Associations of the *NPC1L1* index SNPs with body fat percentage (SD) and LDL-C (SD). The effect is shown per copy of the LDL-C lowering allele. 1 SD of body fat percentage is equivalent to 8.5%, 1 SD of LDL-C is equivalent to 1.0 mmol/L.

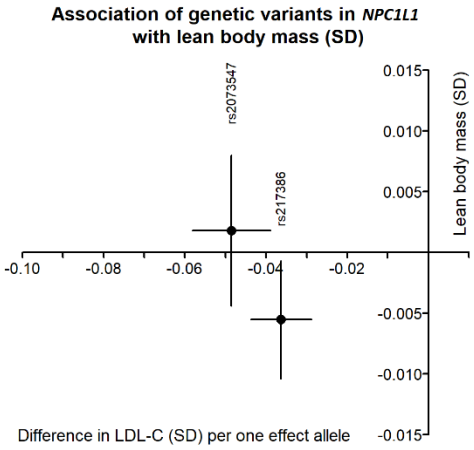


Figure 7:9H | Associations of the *NPC1L1* index SNPs with lean body mass (SD) and LDL-C (SD). The effect is shown per copy of the LDL-C lowering allele. 1 SD of lean body mass is equivalent to 11.5 kg, 1 SD of LDL-C is equivalent to 1.0 mmol/L.

Figure 7:9 | Associations of genetic variants in *NPC1L1* with anthropometric measures

The figure shows anthropometric measures and lipids in SD units. X-axis shows point estimate and 95% CI for the association with LDL-C, Y-axis shows point estimate and 95% CI for the association with a trait.

7.3.3.2 *NPC1L1* GRS and measures of central and general adiposity

Lower genetically predicted LDL-C was associated with lower levels of almost all adiposity measures. The associations were statistically significant for waist circumference (-0.08 SD 95% CI -0.16, -0.00) (-1.02 cm 95% CI -2.00, -0.04), hip circumference (-0.09 SD 95% CI -0.17, -0.01) (-1.00 cm, 95% CI -1.78, -0.19), and body fat percentage (-0.07 SD 95% CI -0.13, -0.00) (-0.64% 95% CI -1.23, -0.04) per 1 SD lower genetically predicted LDL-C (**Figure 7:10**).

Stratified by sex, 1 SD lower genetically predicted LDL-C in women was associated with 0.14 SD (95% CI 0.03-0.25) lower body weight, 0.13 SD (95% CI 0.02, 0.23) lower waist circumference, 0.16 SD (95% CI 0.05, 0.27) lower BMI, 0.15 SD (95% CI 0.05, 0.24) lower body fat percentage, and 0.16 (0.05, 0.27) lower hip circumference. There was no association with waist-to-hip ratio and lean body mass.

In men, *NPC1L1* GRS had no effect on measures of adiposity. The model with fitted interaction term showed that sex was an effect modifier in the association between *NPC1L1* GRS and body weight ($P_{interaction}$=0.020), BMI ($P_{interaction}$=0.005), and body fat percentage ($P_{interaction}$=0.008). However, the evidence that the effect on adiposity is stronger in women is weak, due to multiple testing.

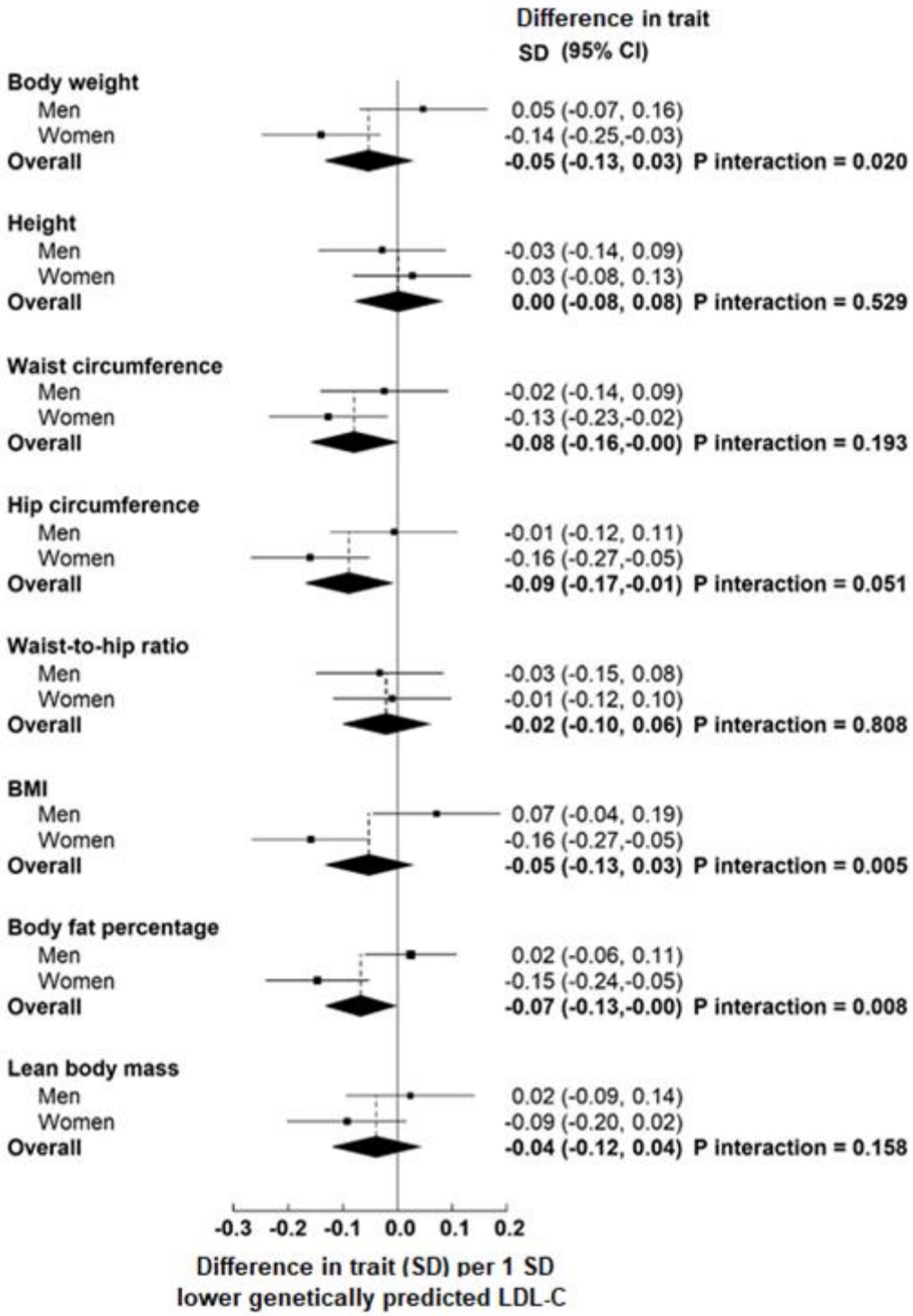


Figure 7:10 | Associations of the *NPC1L1* GRS with anthropometric measures per 1 SD lower genetically predicted LDL-C

SD units (men/women): body weight 14.2/13.9 kg, waist circumference 11.3/12.5 cm, BMI 4.2/5.1 kg/m^2 , waist-to-hip ratio 0.1/0.1, body fat percentage 5.8/6.9%, hip circumference 7.5/10.3 cm, body lean mass 7.7/5.0 kg.

7.3.3.3 External validation analysis of the effect of *NPC1L1* GRS on anthropometric measures

***NPC1L1* index SNPs**

There was no statistical heterogeneity between the effects of rs2073547 and rs217386 on anthropometric measures in UK Biobank and the GIANT consortium. None of the effects in the GIANT consortium on anthropometric measures was statistically significant (Appendix 7.4 and Appendix 7.8).

***NPC1L1* GRS**

The 2-sample inverse-variance weighted MR model using summary level data from the GIANT consortium showed no evidence of statistically significant associations between *NPC1L1* GRS and anthropometric measures (**Figure 7:11**).

There was no heterogeneity between the effects of *NPC1L1* GRS on adiposity estimated in the UK Biobank and in the GIANT consortium. Fixed effect meta-analyses showed associations with lower levels of all adiposity measures, which was statistically significant for waist circumference (-0.074 SD 95% CI -0.142), and hip circumference (-0.87 SD 95% CI -0.16, -0.02) per 1 SD lower genetically predicted LDL-C (**Figure 7:11**).

Analyses stratified by sex in the GIANT study found no heterogeneity between the effect of *NPC1L1* GRS on anthropometric measures in men and women (**Figure 7:12)**. The effect estimates were quantitatively and directionally consistent in both sexes. The directionally discordant results for men and women computed in the UK Biobank were not observed in the GIANT consortium (**Figure 7:12**)

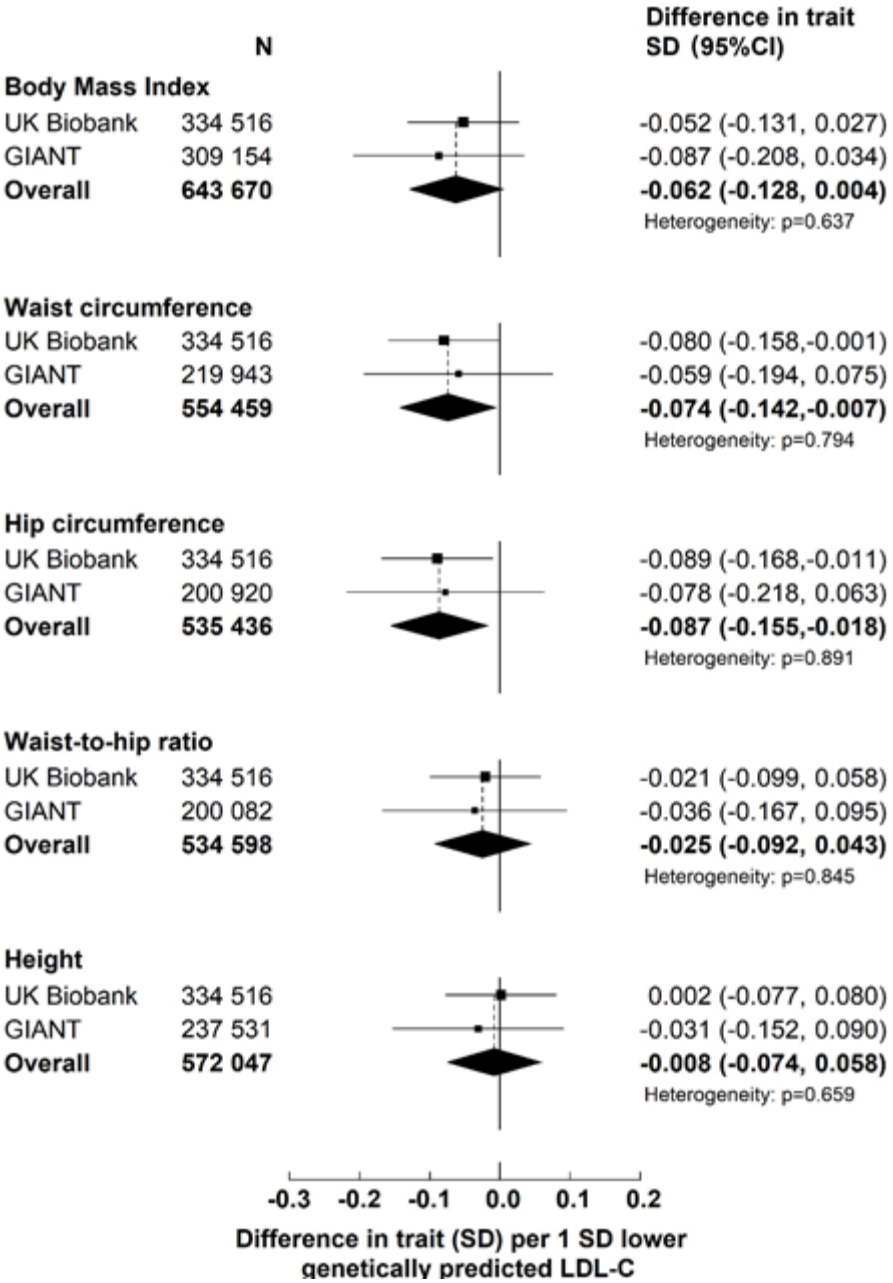


Figure 7:11 | Fixed-effects meta-analysis of the associations between *NPC1L1* GRS with anthropometric measures in UK Biobank and GIANT

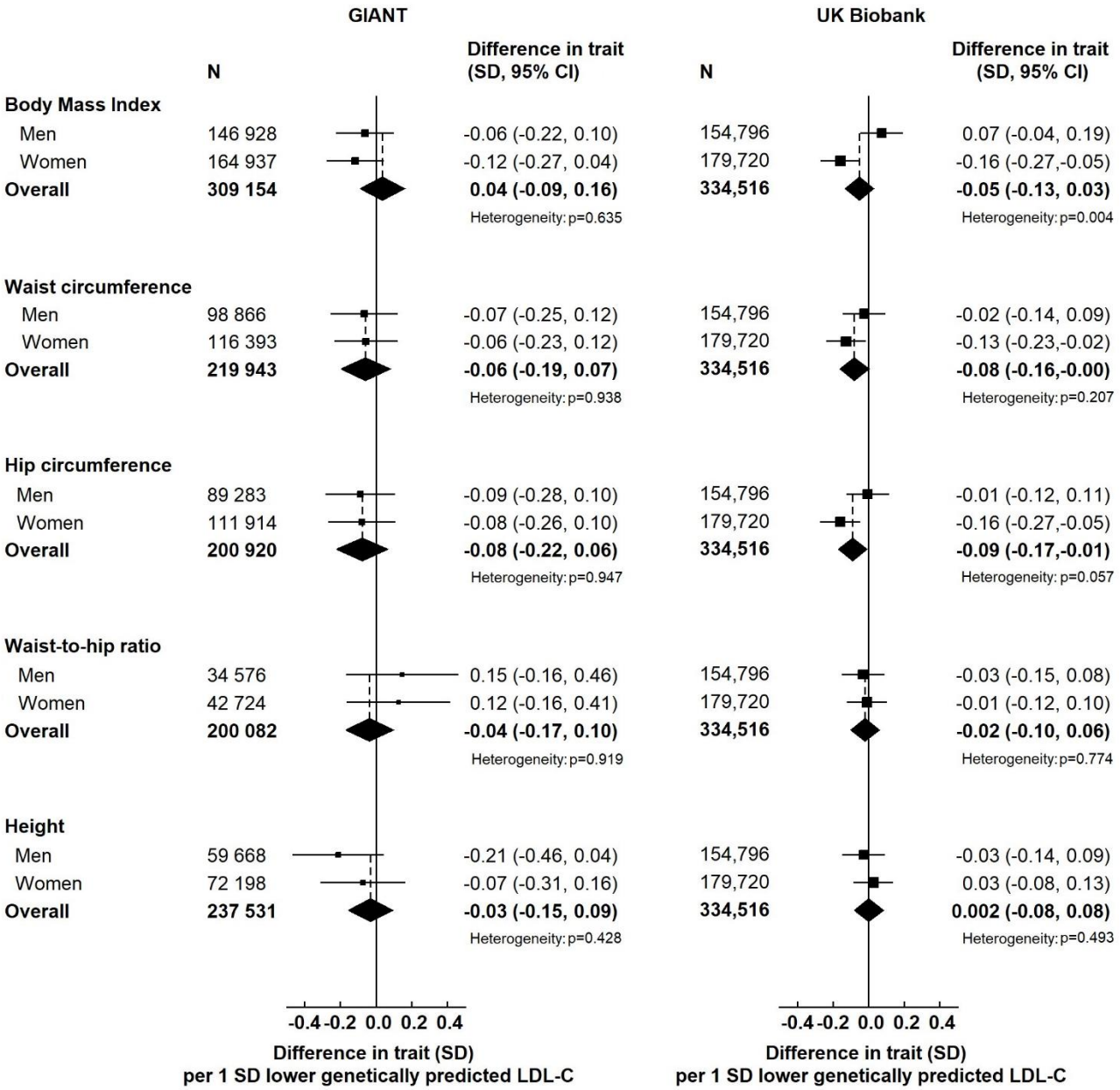


Figure 7:12 | Effects of *NPC1L1* GRS on anthropometric measures in the GIANT consortium and UK Biobank stratified by sex

UK Biobank SD units (men/women): BMI 4.2/5.1 kg/m^{2,} waist circumference 11.3/12.5 cm, hip circumference 7.5/10.3 cm, waist-to-hip ratio 0.06/0.07, height 6.8/6.2 cm

GIANT SD units (men/women): BMI 3.6/4.7 kg/m^{2,} waist circumference 9.7/11.0 cm, hip circumference 6.9/9.7 cm, waist-to-hip ratio 0.05/0.06, height 6.1/5.7 cm

7.4 Discussion

7.4.1 The association of *NPC1L1* GRS with T2DM

Analyses of the associations between *NPC1L1* GRS, a genetic proxy for ezetimibe, and risk of T2DM in UK Biobank found no significant effect of the *NPC1L1* GRS on either the risk of total T2DM, or on the risk of incident T2DM. Nonetheless, the effect on risk of total T2DM was similar in magnitude to that estimated for the *HMGCR* GRS, a genetic proxy for statin therapy (OR 1.25, 95% CI 0.90, 1.74 and OR 1.26 95% CI 1.08, 1.47 per 1 SD lower genetically predicted LDL-C respectively).

To date, evidence for the association between *NPC1L1* GRS and risk of T2DM from genetic studies has been mixed. Ference[66], in his analysis[4] of the effects of *NPC1L1* and *HMGCR* GRSs, found no association between a *NPC1L1* GRS and the risk of total T2DM, whereas Lotta *et al*.[30] reported a substantially different association. Lotta's[30] meta-analysis computed a *NPC1L1* GRS using the same SNPs as used in the present study and found more than a doubling in T2DM risk per 1 SD lower genetically predicted LDL-C. Lotta's meta-analysis of the association between *NPC1L1* GRS and risk of T2DM comprised three genetic studies, including the first release of UK Biobank genetic data (150,392 participants). Lotta's findings were largely driven by data from the EPIC-InterAct consortium and the DIAGRAM consortium.

[4] External datasets used in Ference's analysis: ARIC, CHS, FHS, MESA, CARDIA, WHI, MIGen, WTCCC, GLGC, CARDIoGRANplusC4D

It is not clear why the inverse variance-weighted MR analysis of the effect of the *NPC1L1* GRS on T2DM in the DIAGRAM consortium[157] (**Figure 7:4**) demonstrated a much larger association between genetically lower LDL-C and risk of T2DM than was observed in the UK Biobank. The DIAGRAM dataset comprised 18 smaller GWAS studies, including 26,676 cases of T2DM and 132,532 controls. The DIAGRAM study populations consisted of a diverse range of European ethnic groups. There was no unified definition of T2DM across these smaller studies, with outcome definitions including self-reported diabetes, use of hyperglycaemic medication, reports of T2DM in electronic medical records, and biochemical measurements (blood glucose and/or HbA1c). Genotyping was performed on different platforms, and study—specific quality controls were applied in each GWAS included to the meta-analysis. The analysis of the associations between genetic variants and the risk of T2DM was adjusted for BMI, which could potentially amplify this association.

The present study used one algorithmically derived definition of T2DM and provided a similar number of total T2DM cases and a larger number of controls compared to DIAGRAM. Genotyping of the majority of the UK Biobank participants was performed using one platform, therefore the result of the analysis run on UK Biobank dataset is considered highly reliable. The effect of *NPC1L1* GRS on T2DM in the current study was directionally consistent with the effect estimate from the IMPROVE-IT trial (HR 1.04, 95% CI 0.94, 1.15)[83], though IMPROVE-IT was underpowered to detect any effect on T2DM, especially for a drug with only a modest LDL-C lowering effect. In the case of statins, numerous large RCTs were needed to detect the modest increase in new-onset T2DM, therefore it is too early to conclude that ezetimibe has no effect on T2DM based on one large RCT only. The difference in magnitude of the effects in IMPROVE-IT trial and the current genetic study may be partly explained by the life-

long effect of genetically lower LDL-C compared to a short term pharmacological intervention with ezetimibe[68] (the median follow-up was 6 years). Another possible explanation comes from the comparison of *NPC1L1* with *HMGCR* and statins. *HMGCR* and *NPC1L1* have similar effects on T2DM per 1 SD change in LDL-C (26% and 25% respectively). Ezetimibe lowered LDL-C by 0.43 mmol/L in IMPROVE-IT and yielded a non-significant increase in T2DM of 4% while statins increased T2DM by 12% based on an approximate 1mmol/L decrease in LDL-C[126]. Scaled to the 1 mmol/L change in LDL-C, the effect of ezetimibe on T2DM may be similar to the effect of statin therapy.

There was no heterogeneity in the effect of the *NPC1L1* GRS on T2DM in men and women. Directionally discordant point estimates in UK Biobank were not confirmed in the validation analysis conducted in the independent population.

7.4.2 The associations of *NPC1L1* GRS with glycaemic traits

There was strong evidence of an association between *NPC1L1* GRS and HbA1c. 1 SD lower genetically predicted LDL-C was associated with higher HbA1c. The effect size of the *NPC1L1* GRS on HbA1c was larger than the effect size of the *HMGCR* GRS (0.15 [95% CI 0.09, 0.22] SD and 0.05 [95% CI 0.02, 0.08] SD HbA1c per 1 SD lower genetically predicted LDL-C respectively). The *NPC1L1* GRS had no association with non-fasting glucose.

The results of this study, showing that lower *NPC1L1*-predicted LDL-C is associated with higher HbA1c, potentially contradict previous findings from small RCTs of ezetimibe. A meta-analysis of RCTs of the effect of ezetimibe on glycaemic control[155]

did not detect any effect of a monotherapy with ezetimibe vs placebo, or a combination therapy with ezetimibe and statins vs monotherapy with the same statin on levels of HbA1c (weighted mean difference was 0.12 SD, 95% CI −0.12 to 0.33). The current study, therefore, has demonstrated that genetic proxies of both statin therapy and ezetimibe have directionally concordant effects on HbA1c levels, though the effect of *NPC1L1* GRS on HbA1c seems to be greater in magnitude than the effect of *HMGCR* GRS.

7.4.3 The associations of *NPC1L1* GRS with anthropometric measures

There was no strong evidence for an association between *NPC1L1* GRS and anthropometric measures in sex-combined analyses. *NPC1L1*-predicted lower LDL-C was marginally associated with smaller waist circumference, smaller hip circumference, and lower body fat percentage.

These results differ from the associations of *HMGCR* GRS with anthropometric measures. While *NPC1L1* analysis has been unable to demonstrate higher measures of adiposity associated with 1 SD lower genetically predicted LDL-C, *HMGCR* showed strong statistically significant associations with body weight (0.13 [95% CI 0.09, 0.16] SD) and other anthropometric measures. The differences in the effects on adiposity could be attributed to the different biological pathways underlying the diabetogenic risk associated with lower genetically predicted LDL-C[158], suggesting that adiposity may not be an important part of the diabetogenic process mediated via *HMGCR* or *NPC1L1*.

7.4.4 Limitations of the current study

The findings in this project are subject to at least two limitations. First, the analysis is affected by a weak instrument bias. *NPC1L1* GRS explained only a small proportion of variance in LDL-C (0.138%). The F-statistic from the regression of LDL-C levels on *NPC1L1* GRS was 10.42, nominally meeting the lowest threshold of 10 for a weak instrument bias[159].

Second, incident cases of T2DM which were diagnosed in primary care only could not be included. Patients with incident T2DM who were not admitted to the hospital were not identified in the current study.

7.5 Summary

In the present MR analysis, the weighted *NPC1L1* GRS, constructed from two independent polymorphisms, was not associated with significantly increased risks of total or incident T2DM, though the results for total T2DM were very similar to those from analyses of *HMGCR*. The magnitude and the direction of the effect on incident T2DM were discordant with the total T2DM. The results of the present study demonstrated an association between *NPC1L1* GRS and HbA1c. Lower *NPC1L1*-predicted LDL-C was associated with higher HbA1c, similar to analyses of *HMGCR*.

The associations with anthropometric measures were complex. There were marginally significant protective effects on waist circumference, hip circumference, and body fat percentage. In the sex-specific analysis, *NPC1L1* GRS was not associated with any adiposity measures in men but demonstrated consistent protective associations with adiposity measures in women.

The directional consistency of the underpowered association of *NPC1L1* GRS with the risk of total T2DM and a strong statistically significant association with higher HbA1c levels suggests a possible causal effect of ezetimibe on the risk of T2DM.

Chapter 8 Discussion

8.1 Introduction

This research project aimed to elucidate the effects of the genetic proxies of two lipid-modifying therapies on the risk of T2DM and T2DM-related measures of glycaemia and anthropometric traits. *HMGCR* and *NPC1L1* GRSs were used as genetic proxies of statins and ezetimibe, respectively, two drugs which reduce the circulating levels of LDL-C through different biological mechanisms of action. The first part of the analysis examined the associations of the *HMGCR* SNPs and a *HMGCR* GRS with the risks of total and incident T2DM, and effects on HbA1c levels and measures of adiposity.

Previous genetic studies[51,126] have shown the associations of genetic variants in *HMGCR* with the risk of T2DM and adiposity. Before UK Biobank data became available, genetic studies used aggregated from multiple sources genetic and phenotypic data. Between study heterogeneity and differences in outcome definitions potentially introduced the risk of bias. The definition of T2DM in contributing studies was inconsistent and varied from self-reported diagnosis to diabetes diagnosed by a health professional, occasionally without specification of the type of diabetes or diagnosis confirmed by one baseline or two consequent biochemical tests.

Moreover, studies which reported the associations with increased adiposity and biomarkers of glycaemia had to analyse results varying by analytical method and collected following different research protocols. The UK Biobank data used for the current study allowed us to define the primary outcome, total T2DM, systematically, following the same algorithm for all participants. A variety of data sources such as

hospital episode statistics, elevated glycaemic measures and use of hypoglycaemic medications were available to identify total and incident T2DM cases. The secondary outcomes, namely measures of glycaemia (including HbA1c levels) and measures of adiposity were collected and analysed following standard procedures and using the same equipment. The advantages of the data available for the current study allowed estimation of any effects with greater precision. The results of the first part of the analysis add robust genetic evidence to the growing body of research confirming the association of *HMGCR* genetic variants with the risk of T2DM, and supportive effects on glycaemia and adiposity.

The second part of the analysis presents novel evidence of the effect of *HMGCR* on total and incident T2DM and measures of glycaemia and adiposity in participants at different levels of risk of T2DM. Existing evidence from post-hoc analyses of a small number of statin RCTs has suggested that an increased statin-associated risk of T2DM is confined to patients who are already at high risk of developing T2DM. This is despite the apparent overlap of the results in subgroups (or participants with and without T2DM risk factors) presented in those studies.

The current study was well-positioned to explore the diabetogenic effects of statin's genetic proxy in people at differing risks of developing T2DM. In addition to genetic susceptibility to T2DM (calculated using a T2DM risk score), this study also used a calculation of 10-year risk (calculated using QDiabetes-2018) based on established risk factors for T2DM, and HbA1c levels to define subgroups of individuals at different levels of T2DM risk.

Finally, the last part of the project investigated the effects of *NPC1L1*, a genetic target for ezetimibe, on the risk of T2DM and related traits for two purposes: firstly, to assess

the effect of the genetic proxy of ezetimibe on T2DM due to limited knowledge on this potential adverse effect; and, secondly, to investigate whether the diabetogenic effects of these lipid modulating agents depend on a specific biological pathway or are executed through their effects on LDL-C.

8.2 Summary of findings

8.2.1 The associations of *HMGCR* and *NPC1L1* with T2DM

8.2.1.1 The associations of *HMGCR* and *NPC1L1* GRSs with T2DM and glycaemia

Genetic variants

This study used large-scale, long-term UK Biobank data to elucidate the effects of LDL-C lowering genetic variants in the *HMGCR* and *NPC1L1* genes on the risk of T2DM. Genetic variants in the *HMGCR* gene, used as a proxy for pharmacological inhibition of HMG-CoA reductase with statins, were associated with elevated risk of T2DM (**Figure 8:1)** and higher HbA1c concentrations (**Figure 8:2**). The associations of genetic variants with T2DM and HbA1c followed a similar pattern. The validation analyses conducted in DIAGRAM (for total T2DM) and MAGIC (for HbA1c) showed directionally concordant associations, with CI of the estimates in UK Biobank and validation dataset overlapping considerably (**Figure 8:3 - Figure 8:4)**. In contrast to the main analysis, one SNP (rs3857388) in the validation analysis was not associated with T2DM but was associated with higher HbA1c levels.

The associations of genetic variants within *NPC1L1*, used as a proxy of pharmacological inhibition of Niemann-Pick C1-Like 1 protein with ezetimibe, was qualitatively associated with higher risk of T2DM (**Figure 8:5**) of a similar magnitude to *HMGCR*, and also with higher HbA1c concentrations (**Figure 8:6)**.

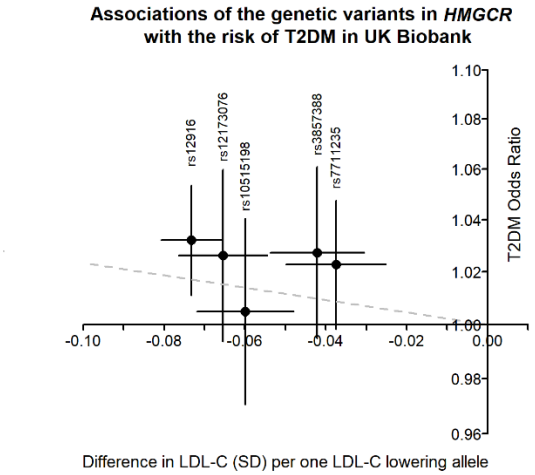


Figure 8:1 | Associations of the HMGCR SNPs with total T2DM (OR, 95% CI) in UK Biobank, and LDL-C (SD, 95 CI) in GLGC

The effect is shown per 1 copy of the LDL-C lowering allele. 1 SD LDL-C is equivalent to 1.0 mmol/L. The grey dotted line shows the effect of HMGCR GRS on total T2DM is 1.26 (95% CI 1.08, 1.47) per 1 SD lower genetically predicted LDL-C. Genotype data were available for 334,669 participants including 20,361 cases of total T2DM and 314,155 controls.

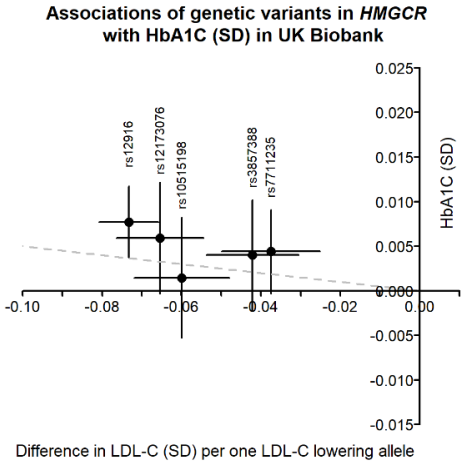


Figure 8:2 | Associations of the HMGCR SNPs with HbA1c (SD, 95% CI) in UK Biobank, and LDL-C (SD, 95% CI) in GLGC

The effect is shown per 1 copy of the LDL-C lowering allele. 1 SD of HbA1c is equivalent to 6.2mmol/mol, 1 SD LDL-C is equivalent to 1.0 mmol/L. The grey dotted line shows the effect of HMGCR GRS on HbA1c is 0.05 SD (95% CI 0.02, 0.08) per 1 SD lower genetically predicted LDL-C

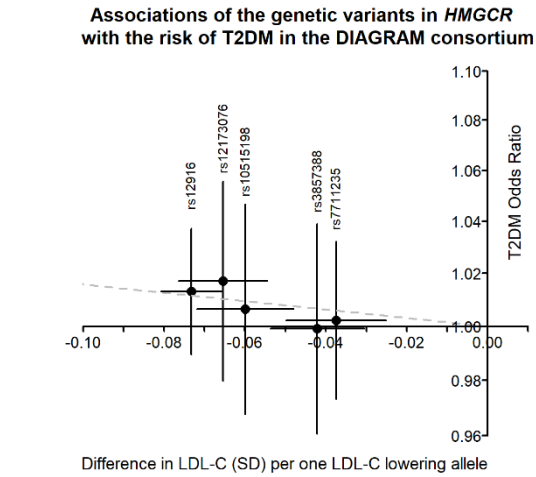


Figure 8:3 | Associations of the HMGCR SNPs with T2DM (OR, 95% CI) in the DIAGRAM consortium, and LDL-C (SD, 95 CI) in GLGC

The effect is shown per 1 copy of the LDL-C lowering allele. 1 SD LDL-C is equivalent to 1.0 mmol/L. The grey dotted line shows the effect of HMGCR GRS on total T2DM is 1.17 (95% CI 0.93, 1.48) per 1 SD lower genetically predicted LDL-C. Genotype data were available for 158,186 participants including 26,676 cases of T2DM.

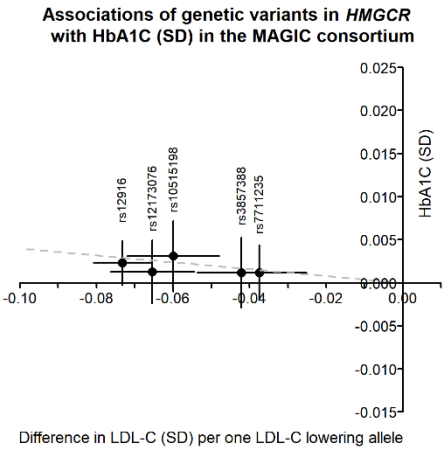


Figure 8:4 | Associations of the HMGCR SNPs with HbA1c (SD, 95% CI) in the MAGIC consortium, and LDL-C (SD, 95% CI) in GLGC

The effect is shown per 1 copy of the LDL-C lowering allele. 1 SD LDL-C is equivalent to 1.0 mmol/L. The grey dotted line shows the effect of HMGCR GRS on HbA1c is 0.04 SD (95% CI 0.01, 0.08) per 1 SD lower genetically predicted LDL-C. Genotype data were available for 281,216 participants.

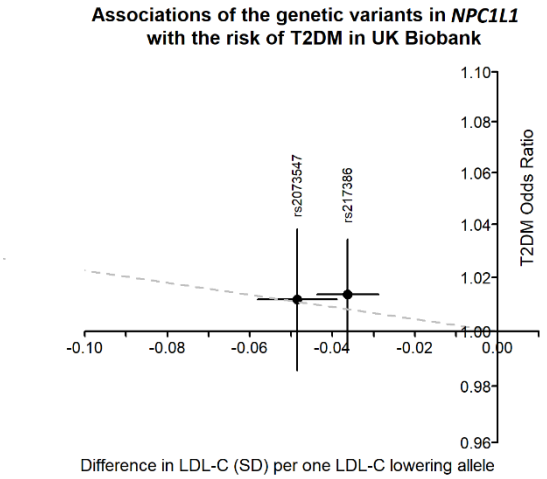


Figure 8:5 | Associations of the NPC1L1 SNPs with total T2DM (OR, 95% CI) in UK Biobank, and LDL-C (SD, 95 CI) in GLGC

The effect is shown per 1 copy of the LDL-C lowering allele. 1 SD LDL-C is equivalent to 1.0 mmol/L. The grey dotted line shows the effect of NPC11L1 GRS on total T2DM is 1.25 (95% CI 0.90, 1.74) per 1 SD lower genetically predicted LDL-C. Genotype data were available for 334,669 participants including 20,361 overall cases of T2DM and 314,155 controls.

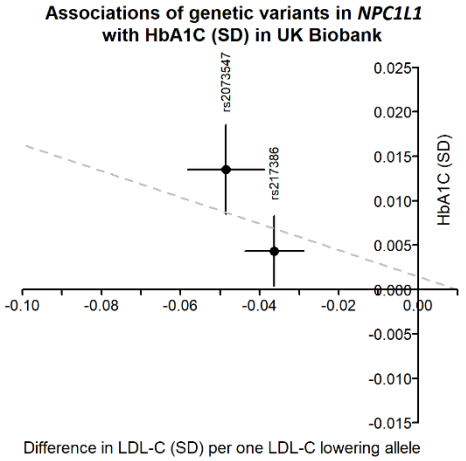


Figure 8:6 | Associations of the NPC1L1 SNPs with HbA1c (SD, 95% CI) in UK Biobank, and LDL-C (SD, 95% CI) in GLGC

The effect is shown per 1 copy of the LDL-C lowering allele. 1 SD of HbA1c is equivalent to 6.2mmol/mol, 1 SD LDL-C is equivalent to 1.0 mmol/L. The grey dotted line shows the effect of NPC1L1 GRS on HbA1c is 0.15 SD (95% CI 0.09, 0.22) per 1 SD lower genetically predicted LDL-C

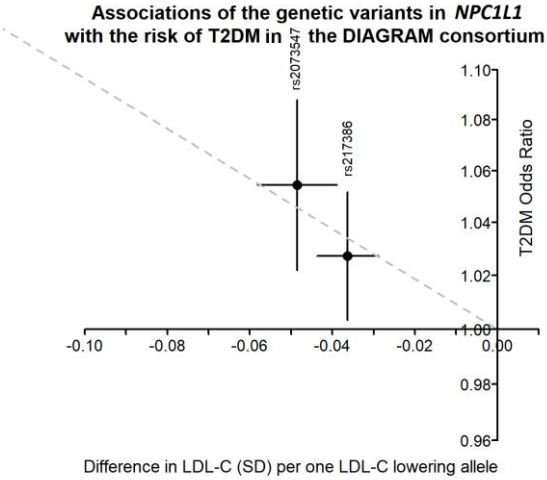


Figure 8:7 | Associations of the NPC1L1 SNPs with T2DM (OR, 95% CI) in the DIAGRAM consortium, and LDL-C (SD, 95 CI) in GLGC

The effect is shown per 1 copy of the LDL-C lowering allele. 1 SD LDL-C is equivalent to 1.0 mmol/L. The grey dotted line shows the effect of NPC1L1 GRS on T2DM is 2.51 (95% CI 1.59, 3.96) per 1 SD lower genetically predicted LDL-C. Genotype data were available for 159,208 participants including 26,676 cases of T2DM.

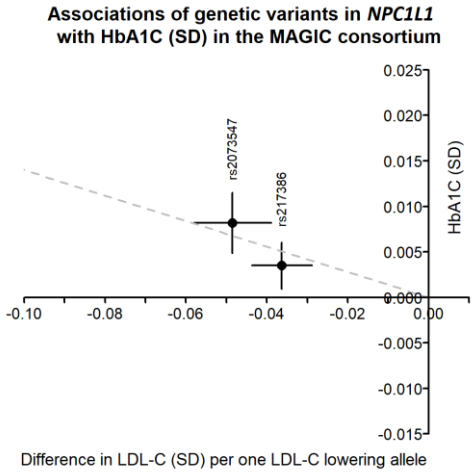


Figure 8:8 | Associations of the NPC1L1 SNPs with HbA1c (SD, 95% CI) in the MAGIC consortium, and LDL-C (SD, 95% CI) in GLGC

The effect is shown per 1 copy of the LDL-C lowering allele. 1 SD LDL-C is equivalent to 1.0 mmol/L. The grey dotted line shows the effect of NPC1L1 GRS on HbA1c is 0.14 SD (95% CI 0.06, 0.22) per 1 SD lower genetically predicted LDL-C. Genotype data were available for 281,216 participants.

GRS for *HMGCR* and *NPC1L1*

Five independent variants in the *HMGCR* gene and two independent variants in the *NPC1L1* gene were combined to generate a *HMGCR* GRS and a *NPC1L1* GRS, which explained 0.36% and 0.14% of the variance in LDL-C respectively. While modulating the levels of LDL-C through different biological mechanisms, both GRSs were associated with higher risk of T2DM in the UK Biobank (**Figure 8:1, Figure 8:3**). 1 SD lower *HMGCR*-predicted LDL-C was associated with 26% (95% CI 1.08, 1.47) higher risk of total T2DM. The association of *NPC1L1* GRS with the risk of total T2DM was quantitatively similar to the association estimated for *HMGCR* GRS (OR 1.25 95% CI 0.90, 1.74 per 1 SD lower *NPC1L1*-predicted LDL-C). Unlike the *HMGCR* GRS, which had greater statistical power, *NPC1L1* GRS did not reach the level of statistical significance of 0.05, but was significant when meta-analysed with the other study. A positive but underpowered association implies that *NPC1L1* GRS is likely to be associated with a higher risk of T2DM, consistent with two previous studies examining the effect of *NPC1L1* GRS on T2DM[30,158] .

These results contribute to the growing body of evidence suggesting the association between modulation of circulating levels of LDL-C with statins and increased risk of T2DM, which was first reported by meta-analyses of large, long-term RCTs of statins[19,67] and subsequently in observational studies[160]. Two previous genetic studies[30,51] have used genetic consortia data, which did not include UK Biobank, and reported directionally concordant associations with T2DM of larger magnitude. Ference *et at*[51] and Lotta *et al* [30]estimated that 1 SD lower *HMGCR*-determined LDL-C was associated with 59% (95% CI 1.25, 2.02) and 39% (95% CI 1.12, 1.73) higher risk of T2DM respectively.

The evidence of the effects of ezetimibe on glycaemia, adipose tissue and risk of T2DM from small studies is controversial. While some studies concluded that ezetimibe is associated with elevated blood glucose[161], other studies reported favourable effects on insulin resistance and visceral fat[162,163]. A single-centre clinical study examined the effects of combination therapy with ezetimibe and statin versus statin monotherapy on glucose metabolism in 133 participants[164]. The study reported a significant reduction of insulin resistance in the ezetimibe plus statin group compared to the statin monotherapy group; however, there was no effect on dysglycaemia[164].

The current study was the first to examine the association between *HMGCR* and *NPC1L1* in a large population with HbA1c measured using a single assay. This advantage is significant for a disease like T2DM, in which diagnosis is based on the threshold in a biomarker. A recent study has shown that biochemical tests in the UK Biobank are superior to the tests conducted on different test platforms[165]. Methodological differences such as sample storage, type of analyser, and quality assurance protocols impact the measurement values from different laboratories. All previous studies of the association between *HMGCR* GRS and HbA1c used genetic consortium summary level data, which is a combination of many GWASs (for example, MAGIC consortium[166] data for HbA1c consists of 52 GWASs).

The analysis of the associations between *HMGCR* GRS and HbA1c conducted on the UK Biobank data also presents an advantage over data received from RCTs of statins[82], as only the minority of RCTs pre-specified T2DM as a potential adverse event and, therefore, did not include regular HbA1c testing. Meta-analyses, which reported the effects on the HbA1c levels most likely conducted the tests on different platforms and may be biased due to assay methods and calibration differences. Data

for the association pooled from different RCTs are likely to be even less accurate. In contrast to the data from meta-analyses of RCTs of statins, this study estimated the association between *HMGCR* GRS and HbA1c with a high level of precision and identified participants receiving oral hypoglycaemic therapy/insulin and excluded them from the analysis.

While the effects of statin and its genetic proxy on the risk of T2DM have been well documented in recent years, the effects of ezetimibe and its genetic proxy on T2DM is less clear. The largest, long-term RCT of combination therapy with ezetimibe/statin vs placebo/statin has not reported any differences in the prespecified safety outcomes by treatment[23]. Although T2DM was not prespecified as a potential adverse effect in IMPROVE-IT RCT, a post-hoc analysis found no excess risk of T2DM in the ezetimibe/statin arm (HR 1.04, 95% CI 0.94, 1.15)[23,167] However, IMPROVE-IT [68] did not pre-specify new-onset T2DM as a safety outcome, and was statistically underpowered to detect any such effect of ezetimibe added to statin therapy. The lack of association between ezetimibe/statin treatment and the risk of T2DM in that RCT, therefore, should not be considered as a proof of clinical safety with regard to new-onset T2DM. As shown by the statin example, results of several RCTs may be needed to provide sufficient power to reliably estimate the risk of T2DM.

In contrast to the IMPROVE-IT results, however, genetic evidence from the current project suggests that biological mechanisms of action of *NPC1L1* are associated with elevated risk of T2DM. The magnitude of the effect of the genetic proxy of ezetimibe was similar to the effect of genetic proxy of statins, per 1mmol/L lower genetically determined LDL-C, therefore it seems premature to conclude that pharmacological inhibition of NPC1L1 does not exert a diabetogenic effect.

One genetic study by Lotta *et al*.[30] claimed that a *NPC1L1* GRS was associated with increased risk of T2DM and that the risk was of a higher magnitude (OR 2.42, (95% CI 1.70, 3.43) vs OR 1.25 (95% CI 0.90, 1.74) than the current study. The current study is the first to estimate the effects of individual genetic variants in *NPC1L1* on T2DM risk in the UK Biobank, a study which differs from DIAGRAM regarding T2DM diagnostic criteria and benefits from more systematic and coherent methods ascertainment T2DM phenotype. These results agree with the results based on the DIAGRAM consortia but suggest a more conservative effect magnitude.

The discrepancy in the magnitude of the associations between DIAGRAM and UK Biobank may be partly explained by the characteristics of the studies included in the DIAGRAM consortia. The DIAGRAM dataset included 18 smaller GWASs with no uniform definition of T2DM and diverse European populations, and the effect estimates were derived from a model adjusted for BMI. Due to the specifically designed Metabochip panel, which was enriched for T2DM and related traits, the GWAS is expected to have inflation in association signals[47]. Another study used overlapping data from UK Biobank, the DIAGRAM consortium and data from the database of Genotypes and Phenotypes studies, with 629,686 participants, including 53,145 cases of T2DM[50]. The association between *NPC1L1* GRS based on five genetic variants and T2DM was estimated as 1.44 (95% CI 1.16, 1.77) per 1 SD lower *NPC1L1* predicted LDL-C, which is closer to the association estimated in the current study.

Similar to *HMGCR* GRS, the current study investigated the association between *NPC1L1* GRS and HbA1c. 1 SD lower *NPC1L1*-predicted LDL-C was associated with 0.15 SD (95% CI 0.09, 0.22) higher HbA1c. The magnitude of the effect of *NPC1L1* GRS on HbA1c in UK Biobank was larger than the effect of *HMGCR* GRS, an

observation which remains unexplained but further supports the idea that ezetimibe may exert a diabetogenic effect.

This result adds to current knowledge of the possible effect of LDL-C lowering with ezetimibe on glucose control and the risk of T2DM. The only large, long-term RCT of ezetimibe[23] did not report its effect on glycaemic traits and the risk of incident T2DM (except one conference paper[167]). The current study supplements the absence of evidence for glycaemic measures from large RCTs and suggests that ezetimibe might have a comparable to statins.

8.2.1.2 The effects of *HMGCR* GRS on participants at different risks of T2DM

Evidence from the recent meta-analyses of RCTs of statins[84,140] and genetic studies[30,51,67] demonstrates that statins, as a class, increase the risk of T2DM. However, the question of whether this risk affects all patients or is limited to only those already at high risk of T2DM, has remained controversial (**Figure 6:13**). The limited evidence from a post-hoc subgroup analysis of the JUPITER trial[87], analyses of the TNT and IDEAL trials[86], and a genetic study[51] all claimed that an increased risk of statin-associated T2DM is confined to those with risk factors for T2DM. Both types of studies had methodological limitations and the potential for misleading results. While JUPITER was the first trial of statin which included incident T2DM as a prespecified outcome[168], the post-hoc subgroup analyses were not pre-planned[87]. It stratified participants into low and high-risk T2DM groups based on established risk factors. The study concluded that the effect of statins on new-onset T2DM was limited to those already at high risk despite largely overlapping confidence intervals for the subgroups of participants at high and low risk. The interpretation of the results was solely based

on p-values for individual subgroups rather than on a statistical test of interaction across subgroups.

The same methodological caveat applies to the genetic study of Ference et al[51]; the authors ran two separate models to examine the associations between *HMGCR* GRS and incident T2DM in the subgroups of healthy participants and participants with impaired fasting glucose levels at baseline. The marginal association in the pre-diabetic group was interpreted as evidence of a difference in effect across the two groups, followed by a controversial conclusion that the statin-associated risk of T2DM may therefore be confined to pre-diabetic patients.

The limited evidence of the claimed differential statin-associated risk of T2DM in those already at high risk of T2DM compared to those not at risk, however, has been widely cited in published statements summarising statin safety and associated adverse events[18,8]. For example, the guideline on the management of blood cholesterol by the American College of Cardiology published in 2018[169] stated that statins increase the risk of incident T2DM in patients with risk factors such as metabolic syndrome and its components.

The current study is the first attempt to systematically assess the risk of *HMGCR*-associated T2DM in patients with different risks of T2DM. The main advantage over the previous studies is that UK Biobank data allowed the application of various approaches to identify those at different risks of T2DM, both by genetic predisposition to T2DM and by clinical risk factors. These subgroup analyses were pre-planned, and participants were stratified into risk groups by their genetic predisposition to T2DM in the primary analysis and the absolute risk of incident T2DM in the secondary analysis. There was no evidence for a differential effect of *HMGCR* GRS on T2DM; in both

models, P-values for interactions in those at different levels of T2DM risk were not statistically significant. Although no evidence for any difference in the proportional effect (HR) emerged, it is nevertheless the case that in terms of the absolute excess in diabetes most of the cases will be in people at high diabetes risk.

Leveraging both individual-level genetic data from UK Biobank and GWAS summary statistics from the DIAGRAM consortium, the current study was able to estimate an individual's genetic susceptibility to T2DM for each participant and, therefore, to include all participants in the subgroup analyses based on genetic risk of T2DM. This approach outperforms the previous genetic study by analysing ten times more participants and three times more cases of T2DM.

The secondary analysis stratified participants without prevalent T2DM to the different risk groups by QDiabetes-2018 risk score, which calculates T2DM risk based on established risk factors. This approach corresponded to post hoc analyses of the JUPITER[87],TNT and IDEAL trialss[86]. Both of those studies again claimed that the risk of incident T2DM was confined to the participants with pre-existing risk factors for T2DM (such as elevated blood glucose, obesity, metabolic syndrome, and HbA1c >6%). The model used for the secondary analysis in the current study included most of these risk factors, plus additional clinical risk factors which are known to increase the risk of T2DM. Due to the necessary exclusion of the participants with prevalent T2DM at the UK Biobank baseline assessment in this subgroup analysis, the model utilised a smaller number of events (6,467 cases of incident T2DM); however, was able to discriminate between diabetes and non-diabetes effectively. There was no evidence that participants with more clinical risk factors for T2DM experienced excess proportional risk of *HMGCR* GRS-associated incident T2DM. Contrary to the previous

findings from the JUPITER subgroup analysis[168], the current study noted a trend towards decreasing proportional risk of T2DM associated with *HMGCR* GRS in the groups with a higher risk of incident T2DM (i.e. absolute risk estimated by QDiabetes-2018 risk score). Tests for trend showed a marginally significant (P-value for trend = 0.032) systematic decrease in proportional risk of incident T2DM across the subgroups with an increased absolute risk of incident T2DM.

A sensitivity analysis excluded participants with HbA1c levels meeting T2DM diagnostic criteria; the remaining participants were stratified into normoglycaemic and pre-diabetes subgroups based on the laboratory-measured HbA1c levels. There was no statistically significant difference in the effect of *HMGCR* GRS on the proportional risk of incident T2DM across the groups; however, consistent with the secondary analysis, participants with higher clinical risk of T2DM appeared to have a lower proportional risk of incident T2DM associated with the *HMGCR* GRS.

The consistency in the pattern of associations in both models, reflecting a life-long genetic risk of T2DM and predisposing clinical risk factors for T2DM, supports the rigour of the findings. Furthermore, coherent results from sensitivity analyses reinforce the conclusion that *HMGCR* GRS associated risk of T2DM is not confined to the patients with pre-existing risk factors for T2DM.

8.2.1.3 Circulating levels of LDL-C and risk of T2DM

Consistent findings regarding both *HMGCR* and *NPC1L1* GRSs (when scaled to 1 SD lower LDL-C) and T2DM suggest that circulating LDL-C levels may directly impact on T2DM risk. Evidence from the current study is in accord with two genetic studies[119,158]

showing an inverse association between genetically-instrumented lower LDL-C and T2DM. White *et al.* [119] constructed a polygenic LDL-C GRS, which explained 7.9% of its variance. The LDL-C GRS was associated with a 21% higher risk of T2DM per 1 SD lower genetically determined LDL-C, leading to the conclusion that elevated LDL-C level was causally related to reduced risk of T2DM. Similarly, Liu *et al.*[110] reported a weak inverse correlation between the effects of 113 independent variants associated with LDL-C and their effects on T2DM. However, the authors emphasised the presence of heterogeneity in the effects on LDL-C and T2DM between SNPs representing different lipid-modulating genetic mechanisms. Therefore, it has been suggested that some lipid-modulating genetic mechanisms have an effect on T2DM, whereas others may not[110].

The current study found that both *HMGCR* and *NPC1L1* GRSs, which modify the circulating LDL-C levels through different biological mechanisms of action, have similar effects on LDL-C and the risk of T2DM. *HMGCR* GRS was used to proxy pharmacological inhibition of HMG-CoA reductase, a rate-limiting enzyme, which downregulates the cholesterol biosynthesis pathway and upregulates hepatic LDL receptors. As a result, the LDL receptors catabolise circulating LDL and lower circulating levels of LDL-C. *NPC1L1* GRS was used to proxy a pharmacological inhibition of a multi-pass membrane protein, NPC1L1, which inhibits the absorption of intestinal cholesterol. Despite different biological mechanisms of action and tissue-specific functions, both GRSs were associated with lower LDL-C and ApoB-100 levels and a higher risk of T2DM. Therefore, the higher risk of T2DM associated with LDL-C lowering drug targets could be the result of the effect on circulating LDL-C rather than be attributed to a particular biological mechanism. This conclusion is supported by the recent findings that other LDL-C modulating genetic drug targets are also associated

with T2DM. For example, the genetic proxy of proprotein convertase subtilisin kexin 9 (PCSK9) inhibitors, which upregulate the level of LDL-C receptors by a different mechanism to statins and reduce LDL-C by ≈60%, is also associated with a higher risk of T2DM[51,170]. An inverse association between LDL-C and T2DM has also been reported for *LDLR* GRS[50], which plays an essential role in regulating the expression of LDL-C receptors. These genetic targets have a similar magnitude of association with T2DM per 1 SD lower genetically-predicted LDL-C, ranging between 21% to 29%, with closely corresponding results for *HMGCR* (26%) and *NPC1L1* (25%) from the current study.

While genetic studies provided evidence of the inverse associations between LDL-C and T2DM for genetic proxies of different LDL-C lowering agents, RCTs of different LDL-C lowering drugs have not provided consistent results. For example, the only large-scale RCTs of ezetimibe found no statistically significant association between lowering LDL-C with a combination therapy statin/ezetimibe compared to statin/placebo and the risk of T2DM[167]. Similarly, large RCTs of the PCSK9 inhibitors evolocumab[171] and alirocumab[172] concluded that lowering LDL-C with a PCSK9 inhibitor does not increase the risk of incident T2DM[171].

Clear evidence of the association between statin therapy and T2DM came from a meta-analysis of 13 large-scale long-terms placebo-controlled and standard-care controlled RCTs of statins[19]. In contrast, the available evidence from ezetimibe[167] and PCSK9 inhibitors[172],[171] has likely provided insufficient data to reliably assess their associations with T2DM.

The Cholesterol Treatment Trialists' (CTT) Collaboration was established in 1994 to assess the efficacy and safety of lowering LDL-C with statins in a wide range of patient

populations[173]. In 2016, CTT announced plans to collect individual participant data from all large statin trials and to retrospectively examine potential adverse events of statin therapy, including its effect on T2DM[174]. These analyses are currently underway and will emerge in early 2023, and they will include subgroup analyses based on diabetes risk of trial participants.

8.2.2 The associations of *HMGCR* and *NPC1L1* with anthropometric traits

The current study demonstrated the associations between *HMGCR* GRS and various anthropometric traits. 1 SD lower *HMGCR*-predicted LDL-C was associated with higher BMI, body weight, body fat percentage, waist, and hip circumferences (**Figure 5:9**). The association with waist-to-hip ratio, representing visceral adiposity, was not increased. These results suggest that the effect of *HMGCR* on adiposity is not limited to visceral adipose tissue but extends to general adiposity. Despite different body compositions in men and women[175], there was no difference in the effect of *HMGCR* on adiposity between the sexes (**Figure 5:9**).

These results are in agreement with previous studies reporting the association of individual SNPs in *HMGCR* with higher body weight, BMI, waist circumference[176] and with the associations reported by the GIANT consortium[108] in this thesis.

The effect of *HMGCR* GRS on adiposity could mediate its association with the risk of T2DM. Multiple studies have demonstrated strong associations between measures of general adiposity, such as body weight and BMI[150], and measures of central adiposity, such as waist circumference[150] and waist-to-hip ratio, with increased risk of T2DM. A recent genetic study[175] showed that *HMGCR* was involved in adipose tissue metabolic

pathways, including insulin resistance and insulin signalling. The association of *HMGCR* with adiposity measures demonstrated in the current project fits well in the line of evidence suggesting the causal role of HMGCR inhibition in T2DM. The associations of *NPC1L1* with adipose tissues appeared to be more complex and unexpectedly different from *HMGCR*. Despite concordant effects of *HMGCR* and *NPC1L1* on the risk of T2DM and an even larger effect on HbA1c with *NPC1L1*, there was no evidence of any association of *NPC1L1* with greater adiposity. On the contrary, there was a marginally significant association with a smaller waist and hip circumferences and body fat percentage. In the sex-specific analyses, *NPC1L1* GRS had a consistent protective effect on adiposity in women and was not associated with any changes in adiposity measures in men. Tests for interaction showed a statistically significant difference in the effect of *NPC1L1* on body weight ($P_{interaction}$=0.020), BMI ($P_{interaction}$=0.005), and body fat percentage ($P_{interaction}$=0.008) between sexes (**Figure 7:10)**.

This study is the first to report potential heterogeneity in the effects of *NPC1L1* on adiposity between the sexes; however, these results are suggestive as the validation analysis conducted in the GIANT consortium data did not support these findings (**Figure 7:12**). In the context of the overall association with a higher risk of T2DM and higher HbA1c levels, the sex-dimorphic effect of *NPC1L1* on adiposity needs further investigation.

8.2.3 Strengths and limitations of the study

There are several strengths of this work such as:

Outcome definitions - the broad scope of the individual-level data provided by the UK Biobank allowed the use of standard definitions for diabetes outcomes (total and incident cases of T2DM). This is a particular advantage over RCTs of statins and ezetimibe (in which new-onset T2DM was not pre-specified) that have used non-standard approaches. Previous genetic studies used summary-level data from the DIAGRAM consortium which consists of multiple studies using different approaches to identify T2DM.

Homogeneity in tests and measurements - Tests were carried out using a single assay method on the same types of analysers and using the same quality controls. Biological samples were kept in identical storage conditions. Similarly, a standard adiposity measurements procedure was used for all participants of the UK Biobank.

UK Biobank participants were genotyped using a specifically designed Axiom array from Affymetrix. They followed unified quality control procedures, which was not previously possible when data from the genetic consortia were used for MR studies.

Opportunity for a subgroup analysis – the availability of the genetic data and detailed health history for each participant allowed the design of a primary subgroup analysis (based on genetic risk of total T2DM) with a maximum number of cases in each subgroup and a secondary subgroup analysis (based on clinical risk of incident T2DM) with a very strong discriminative factor variable. Using a genetic approach to define subgroups based on T2DM risk allowed this study to use all 20 361 cases of T2DM and compare the diabetogenic effect of *HMGCR* across the groups with a different genetic predisposition to T2DM.

Validation datasets - the findings in the UK Biobank were validated in various independent consortia datasets. The effect on T2DM was validated in the DIAGRAM

consortium, the effect on HbA1c was validated in the MAGIC consortium, and the effects on the anthropometric traits were validated in the GIANT consortium.

However, there are several limitations inherent in the MR method. One of them is a weak instrument bias. An MR study requires an instrumental variable to be associated with the exposure. LDL-C is a complex trait with more than 8367 common variants that have been known to modulate the blood LDL-C concentrations and related traits[177]. Combined in a polygenic GRS, these variants explain approximately 21% of the variance in LDL-C levels[177]. The current study investigated genetic variants within two genes as genetic proxies for LDL-C lowering medications; the instrumental variables were limited to five SNPs for the *HMGCR* GRS and two SNPs for *NPC1L1* GRS, both explaining very small percentages of variance in LDL-C (0.36% and 0.14% respectively). Weak instrument bias may result in underpowered analyses (the probability of detecting a true effect of the GRS on T2DM was 100% for *HMGCR,* and 35% for *NPC1L1*).

The second limitation refers to extrapolating the results from a genetic study to the effect of a therapeutic agent (which will impart a much stronger LDL-C lowering effect but over a shorter period of time). The genetic analyses examine the life-long exposure to the risk factor; therefore, the results of such analysis cannot necessarily be directly applied to a short-term pharmacological modulation of the risk factor.

The third limitation is that genetic analyses may anticipate on-target effects of a drug but will not predict any off-target effects of therapy.

Not being able to use primary care data in the outcome definition algorithm is the fourth notable limitation of the study. Most cases of T2DM are diagnosed and managed in primary care. The definition of incident T2DM relied on the hospital records where

T2DM was recorded as a primary and secondary diagnosis. Primary care data linked to the UK Biobank participants were not available at the time when the analysis was conducted; therefore, some cases of incident T2DM were not included in the definition of T2DM.Finally, this study focused on the potential diabetogenic effects of both statins and ezetimibe but did not discuss the cardiovascular benefits of lowering LDL-C with these drugs. While considering the associations with the risk of T2DM, increased adiposity and weakened glycaemic controls, we did not report the associations with CVD and cardiovascular mortality, as it was outside the scope of the current study. However, lowering LDL-C with statins (or adding ezetimibe to statin) is a cornerstone of the primary and secondary prevention of CVD; it has been estimated that reducing circulating levels of LDL-C by one mmol/L is associated with an approximately 20-25% reduction in the risk of life-threatening CVD events[178]. Therefore, the health benefits of lipid-modulation therapies far outweigh the small risk of adverse metabolic effects.

8.2.4 Implications

Implication for public health

This research elucidated the effect of genetic proxies of statins and ezetimibe on the risk of T2DM; the evidence for statins triangulates with the evidence from meta-analyses of RCTs of statins and with other genetic studies. The increased risk of T2DM appeared to be an on-target effect of lowering LDL-C with statins and indicates a complex, and underappreciated, relationship between lipid metabolism and glucose homeostasis. This study showed that the diabetogenic effect of *HMGCR* is not limited to the population with pre-existing risk factors for T2DM but affects all groups treated

with a statin, though absolute risk of developing T2DM is substantially higher in those with risk factors for T2DM.

As statins are a commonly prescribed medication for the prevention of CVD, including in patients with T2DM, the prescription of statins should be supplemented with healthy lifestyle advice and regular tests of blood glucose control in those with T2DM or existing risk factors for T2DM. Lifestyle modification or initiation/escalation of glucose-lowering medication may be needed for those experiencing clinically relevant statin-associated glycaemic control deterioration.

Ezetimibe is prescribed as an adjunct to dietary modifications and statin therapy, where patients are not able to achieve a clinically desirable reduction in LDL-C levels with statins alone. The only large RCT[23] showed that ezetimibe did not increase the risk of incident T2DM nor worsen diabetes control. The current study, on the contrary, suggests that ezetimibe may have a comparable effect to statins on the risk of T2DM and an even stronger effect on HbA1c levels (when scaled for the same difference in LDL-C). The prescription of ezetimibe alone or in combination with statin therapy should also be accompanied by healthy lifestyle advice and HbA1c level monitoring in participants at risk of T2DM.

Implication for drug development

The development of new therapeutic agents is a lengthy[179] and resource-intensive process[180], complicated by the low rate of new drug discovery[181] and the low success rate of RCTs[182]. The mean duration of clinical and approval phases for new molecular entities in cardiovascular class during 2005-2009 was 6.7 years[183]. The average cost

of the new drug development in 2013 was $2.8 billion, including the clinical stage of the development process, which constituted $1.5 billion[180].

The escalating costs of research and development programs and the considerable amount of time needed to yield market approval represent notable hurdles in drug development. The risk associated with the high attrition rate of candidate drugs is one of the major challenges for pharmaceutical companies[184]. One consequence of those challenges is a decline in the number of drugs with novel mechanisms of action, which are brought from discovery into the development process by pharmaceutical companies[185]. The stagnation in new drug development during the last two decades resulted in the lowest number of novel drugs with cardiovascular indications that received marketing authorisation[186].

The use of genetic data can assist the development of new therapies by identifying new drug targets by providing additional evidence on potential adverse events associated with biological mechanisms of action of the proposed therapeutic agent. The current study demonstrated how analyses based on UK Biobank data confirm the increased risk of T2DM and increases in measures of adiposity associated with statin therapy, as initially reported in a meta-analysis of 20 large, long-term RCTs of statins[176]. In the case of ezetimibe, it is unrealistic to expect another large RCT, as it was approved by the FDA in 2002 for treatment of hypercholesterolemia in the US[187] and in 2003 by MHRA for use in the UK[188]; therefore, these genetic analyses contribute importantly to the understanding of the possible effects of ezetimibe on T2DM.

The MR method will allow further assessment of the potential diabetogenic effects of other lipid-modulating drug targets, including PCSK9 inhibitors[51], recently approved bempedoic acid (*ACLY*)[50] and emerging *LDLR* gene therapy[50] on T2DM risk. With the

increasing availability of genetic data linked to phenotypes and medical records, such methods will be increasingly used to assess the potential adverse effects of the therapeutic agents in the future.

8.2.5 Further research

While demonstrating the causal relationship between *HMGCR* GRS and T2DM, the biological mechanisms mediating this relationship remain unclear. Future work should include mediation analyses to investigate the extent to which HMGCR inhibition exerts its diabetogenic effect through increases in measures of adiposity and other pathways. For example, a multivariable MR study with a GRS for BMI can be used to estimate the direct and indirect effects of *HMGCR* GRS on T2DM. Moreover, the discrepancy in the effects on adiposity between *HMGCR* and *NPC1L1* suggests that there could be multiple mechanisms mediating the effects on T2DM.

The future release of UK Biobank imaging data will allow the MR approach to perform to its full potential. In the current project, data on adiposity was limited to commonly collected anthropometric measures, indices, and estimations by bio-impedance. However, a future release of abdominal magnetic resonance imaging (MRI) and dual-energy X-ray absorptiometry (DXA) of 100,000 participants will provide unprecedented data on body composition and on site-specific fat distribution[189]. Further insights into the effect of LDL-C modulating GRSs on adiposity and fat distribution as mediating factors of the effect on T2DM might be gained by examining the associations with organ-specific fat accumulation.

Previous studies emphasised the critical role of the liver as a pathophysiological mediator of glycaemic control and insulin sensitivity[190]. Hepatic fat accumulation (hepatic steatosis) is recognised as a risk factor for T2DM. Hepatic steatosis leads to hepatic insulin resistance, a precursor of systemic insulin resistance and progression to T2DM[191]. Although statin therapy appears to be safe for patients with hepatic steatosis[192], several studies demonstrated controversial effects of statins on histological liver outcomes[191,192]. Some l More research is needed to examine the effect of statins on the liver as a potential mediator of statin effect on T2DM.

Similarly, pancreatic steatosis is a risk factor for the development of T2DM[193]. Pancreatic fat infiltration reduces insulin secretion and increases the risk of development of both T2DM and CVD. MRI is an accurate tool to estimate intrapancreatic fat concentration. The associations of *HMGCR* and *NPC1L1* GRSs with hepatic and pancreatic fat will help to better understand the relationship between lipid metabolism and glucose homeostasis.

Appendices

Appendices Chapter 2

Appendix 2:1 The associations of genetic variants in the *HMGCR* gene with T2DM, anthropometric measures and glycaemic traits

Appendix 2:2 The associations of SNPs in the *HMGCR* gene with T2DM, LDL-C and glycaemic / adiposity traits

Appendix 2:3 The associations of genetic variants in the *NPC1L1* gene with T2DM, anthropometric measures and glycaemic traits

Appendix 2:4 The associations of SNPs in the *NPC1L1* gene with T2DM, LDL-C and glycaemic/adiposity traits

Appendix 2:5 RCTs of statins included in the meta-analyses

Appendix 2:6 RCTs of statins included in each meta-analysis

2.1 Appendix 2:1 | The associations of genetic variants in the *HMGCR* gene with T2DM, anthropometric measures and glycaemic traits

rsID	Paper	Pos	Effect allele/ other allele	MAF	Effect allele freq	Ancestry	Trait	Beta/OR	CI	N	Cases	Controls	N studies	Unit
Association of *HMGCR* SNPs and genetic score with T2DM														
rs5909	Ference B.A. et el (2016)	chr5: 74656175	G/A	0.10	0.90	European	T2DM	0.90	(0.71, 1.14)	54640	n/a	n/a	1	OR per 10 mg/dl lower LDL-C
rs5909	DIAGRAM	chr5: 74691931	A/G	n/a	n/a	European	T2DM	1.01	(0.96, 1.07)	69033	12171	56862	38	OR per allele
rs5744707	Lotta L.A et al (2016)	chr5: 74890618	A/G	0.1	0.90	European	T2DM	0.98	(0.956, 1.011)	321044	50775	270269	4	OR per allele
rs3846663	Knouff C.W. et al (2008)	chr5: 75359901	C/T	0.41	0.41	European	T2DM	1	(0.8, 1.2)	5635	n/a	n/a	1	OR per allele
rs2303152	Ference B.A. et el (2016)	chr5: 74641707	G/A	0.12	0.88	European	T2DM	0.86	(0.63, 1.16)	73794	n/a	n/a	1	OR per 10 mg/dl lower LDL-C
rs17238484	Swerdlow D.I. et al (2015)	chr5: 74684252	G/T	0.23	0.77	European	T2DM	1.020	(1.00, 1.04)	190257	26236	164021	35	OR per allele
rs17238484	Ference B.A. et el (2016)	chr5: 74648496	G/T	0.25	0.75	European	T2DM	1.16	(1.03, 1.31)	181111	n/a	n/a	17	OR per 10 mg/dl lower LDL-C

rsID	Paper	Pos	Effect allele/ other allele	MAF	Effect allele freq	Ancestry	Trait	Beta/OR	CI	N	Cases	Controls	N studies	Unit
rs16872526	Lotta L.A et al (2016)	chr5: 74675717	T/G	0.09	0.91	European	T2DM	1.02	(0.985, 1.047)	321044	50775	270269	4	OR per allele
rs12916	Lotta L.A et al (2016)	chr5: 74656539	T / C	0.42	0.58	European	T2DM	1.03	(1.01, 1.05)	376217	55271	320946	15	OR per allele
rs12916	Swerdlow D.I. et al (2015)	chr5: 74692295	T / C	0.40	0.60	European	T2DM	1.06	(1.03, 1.09)	88550	14976	73574	16	OR per allele
rs12916	Ference B.A. et el (2016)	chr5: 74656539	T / C	0.43	0.57	European	T2DM	1.09	(1.00, 1.18)	151329	n/a	n/a	17	OR per 10 mg/dl lower LDL-C
rs12916	Wurtz P. et al (2016)	chr5: 74656539	T / C	n/a	n/a	European	T2DM	1	(0.8, 1.2)	5636	372	5264	1	OR per allele
rs12916	DIAGRAM	chr5: 74692295	T/C	n/a	n/a	European	T2DM	1.02	(0.96306, 1.01893)	63390	9580	53810	38	OR per allele
rs10066707	Ference B.A. et el (2016)	chr5: 74560579	A/G	0.42	0.58	European	T2DM	1.06	(0.91, 1.24)	80650	n/a	n/a	1	OR per 10 mg/dl lower LDL-C
Association of *HMGCR* SNPs and genetic score with glycaemic traits														
rs5909	MAGIC	n/a	A/G		0.125	European	Glucose	0.0016	.0062 Stderr	n/a	n/a	n/a	n/a	n/a
rs5744707	Lotta L.A et al (2016)	chr5: 74890618	A/G	0.1	0.90	European	fasting plasma glucose	0.01	(-0.01, 0.02)	133010	n/a	n/a	1	SD per allele
rs17238484	Swerdlow D.I. et al (2015)	chr5: 74684252	G/T	0.23	0.77	European	Plasma glucose	0.230	(0.020, 0.441)	73490	n/a	n/a	23	per-allele beta, (%)

rsID	Paper	Pos	Effect allele/ other allele	MAF	Effect allele freq	Ancestry	Trait	Beta/OR	CI	N	Cases	Controls	N studies	Unit
rs16872526	Lotta L.A et al (2016)	chr5: 74675717	T/G	0.09	0.91	European	fasting glucose	0.01	(-0.00, 0.03)	133010	n/a	n/a	1	SD per allele
rs12916	Lotta L.A et al (2016)	chr5: 74656539	T / C	0.42	0.58	European	fasting glucose	0.00	(-0.01, 0.01)	133010	n/a	n/a	1	SD per allele
rs12916	Swerdlow D.I. et al (2015)	chr5: 74692295	T / C	0.40	0.60	European	Plasma glucose	0.195	(0.008, 0.382)	30313	n/a	n/a	11	per-allele beta, (%)
rs12916	Wurtz P. et al (2016)	chr5: 74656539	T / C	0.43	n/a	European	Glucose	0.01	(0.00, 0.02)	27914	n/a	n/a	8	per-allele beta, (mmol/L)
rs12916	MAGIC	n/a	T/C		0.403	European	Glucose	0.0027	.0038 Stderr	n/a	n/a	n/a	n/a	
genetic score	Lotta L.A et al (2016)	n/a	n/a		n/a	European	fasting glucose	0.02	(-0.07, 0.11)	133010	n/a	n/a	1	SD unit per 1 mmol/L reduction in LDL-C
rs5909	MAGIC	n/a	A/G		0.125	European	Insulin	0.0099	.0064 Stderr	n/a	n/a	n/a	n/a	
rs5744707	Lotta L.A et al (2016)	chr5: 74890618	A/G	0.1	0.90	European	fasting insulin	0.00	(-0.02 ,0.02)	108557	n/a	n/a	1	SD per allele
rs17238484	Swerdlow D.I. et al (2015)	chr5: 74684252	G/T	0.23	0.77	European	Plasma insulin	0.016	(0.005, 0.027)	37453	n/a	n/a	12	per-allele beta, (%)
rs16872526	Lotta L.A et al (2016)	chr5: 74675717	T/G	0.09	0.91	European	fasting insulin	0.01	(-0.01, 0.03)	108557	n/a	n/a	1	SD per allele
rs12916	Lotta L.A et al (2016)	chr5: 74656539	T / C	0.42	0.58	European	fasting insulin	0.01	(0.00, 0.02)	108557	n/a	n/a	1	SD per allele

rsID	Paper	Pos	Effect allele/ other allele	MAF	Effect allele freq	Ancestry	Trait	Beta/OR	CI	N	Cases	Controls	N studies	Unit
rs12916	Swerdlow D.I. et al (2015)	chr5: 74692295	T / C	0.40	0.60	European	Plasma insulin	0.130	(-0.140, 0.391)	30641	n/a	n/a	12	per-allele beta, (%)
rs12916	MAGIC	n/a	T/C	n/a	0.403	European	Insulin	0.0051	.0039 Stderr	n/a	n/a	n/a	n/a	
genetic score	Lotta L.A et al (2016)	n/a	n/a	n/a	n/a	European	fasting insulin	0.16	(0.05, 0.27)	108557	n/a	n/a	1	SD unit per 1 mmol/L reduction in LDL-C
Association of *HMGCR* SNPs and genetic score with anthropometric measures														
rs5909	GIANT	chr5: 74691931	A/G	0.1417	0.8583	European	BMI	0.0012	0.0065 Stderr	164182	n/a	n/a	59	beta per allele
rs5744707	Lotta L.A et al (2016)	chr5: 74890618	A/G	0.1	0.90	European	BMI	0.00	(-0.01, 0.01)	333495	n/a	n/a	1	SD per allele
rs17238484	Swerdlow D.I. et al (2015)	chr5: 74684252	G/T	0.23	0.77	European	BMI	0.105	(0.066, 0.145)	152004	n/a	n/a	32	per-allele beta, (kg/m2)
rs16872526	Lotta L.A et al (2016)	chr5: 74675717	T/G	0.09	0.91	European	BMI	0.02	(0.01, 0.03)	333495	n/a	n/a	1	SD per allele
rs12916	Lotta L.A et al (2016)	chr5: 74656539	T / C	0.42	0.58	European	BMI	0.02	(0.01, 0.02)	333495	n/a	n/a	1	SD per allele
rs12916	Swerdlow D.I. et al (2015)	chr5: 74692295	T / C	0.40	0.60	European	BMI	0.083	(0.031, 0.135)	75140	n/a	n/a	17	per-allele beta, (kg/m2)

rsID	Paper	Pos	Effect allele/ other allele	MAF	Effect allele freq	Ancestry	Trait	Beta/OR	CI	N	Cases	Controls	N studies	Unit
rs12916	GIANT	chr5: 74692295	T/C	0.5877	0.4123	European	BMI	0.0177	0.0038 Stderr	180586	n/a	n/a	59	beta per allele
genetic score	Lotta L.A et al (2016)	n/a	n/a	n/a	n/a	European	BMI	0.23	(0.15, 0.30)	333495	n/a	n/a	1	SD unit per 1 mmol/L reduction in LDL-C
rs5909	GIANT	chr5: 74691931	A/G	0.1417	0.8583	European	Waist-to-hip ratio	-0.0046	.0069 Stderr	141659	n/a	n/a	59	beta per allele
rs5744707	Lotta L.A et al (2016)	chr5: 74890618	A/G	0.1	0.90	European	Waist-to-hip ratio	-0.01	(-0.02, 0.01)	224047	n/a	n/a	1	SD per allele
rs17238484	Swerdlow D.I. et al (2015)	chr5: 74684252	G/T	0.23	0.77	European	Waist-to-hip ratio	0.001	(0.0003, 0.002)	95496	n/a	n/a	25	per-allele beta, (cm)
rs16872526	Lotta L.A et al (2016)	chr5: 74675717	T/G	0.09	0.91	European	Waist-to-hip ratio	0.01	(-0.01, 0.02)	224047	n/a	n/a	1	SD per allele
rs12916	Lotta L.A et al (2016)	chr5: 74656539	T / C	0.42	0.58	European	Waist-to-hip ratio	0.01	(0.00, 0.01)	224047	n/a	n/a	1	SD per allele
rs12916	GIANT	chr5: 74692295	T/C	0.5877	0.4123	European	Waist-to-hip ratio	0.0074	.0040 Stderr	156660	n/a	n/a	59	beta per allele
genetic score	Lotta L.A et al (2016)	n/a	n/a	n/a	n/a	European	Waist-to-hip ratio	0.09	(0.01, 0.18)	224047	n/a	n/a	1	SD unit per 1 mmol/L reduction in LDL-C
rs17238484	Swerdlow D.I. et al (2015)	chr5: 74684252	G/T	0.23	0.77	European	Weight	0.301	(0.175,0.428)	143113	n/a	n/a	30	per-allele beta, (kg)

rsID	Paper	Pos	Effect allele/ other allele	MAF	Effect allele freq	Ancestry	Trait	Beta/OR	CI	N	Cases	Controls	N studies	Unit
rs12916	Swerdlow D.I. et al (2015)	chr5: 74692295	T / C	0.40	0.60	European	Weight	0.203	(0.043, 0.362)	65871	n/a	n/a	15	per-allele beta, (kg)
rs5909	GIANT	chr5: 74691931	A/G	0.1417	0.8583	European	Waist circ	-0.0069	0.0069 Stderr	141659	n/a	n/a	59	beta per allele
rs17238484	Swerdlow D.I. et al (2015)	chr5: 74684252	G/T	0.23	0.77	European	Waist circ	0.315	(0.158, 0.472)	69163	n/a	n/a	19	per -allele beta, (cm)
rs12916	Swerdlow D.I. et al (2015)	chr5: 74692295	T / C	0.40	0.60	European	Waist circ	0.000	(-0.001, 0.001)	39771	n/a	n/a	14	per -allele beta, (cm)
rs12916	GIANT	chr5: 74692295	T/C	0.5877	0.4123	European	Waist circ	0.0074	0.0040 Stderr	156660	n/a	n/a	59	beta per allele
rs17238484	Swerdlow D.I. et al (2015)	chr5: 74684252	G/T	0.23	0.77	European	Hip circ	0.211	(0.101, 0.321)	69159	n/a	n/a	19	per-allele beta, (cm)
rs12916	Swerdlow D.I. et al (2015)	chr5: 74692295	T / C	0.40	0.60	European	Hip circ	0.303	(0.102, 0.504)	30313	n/a	n/a	11	per-allele beta, (cm)

2.2 Appendix 2:2 | The associations of SNPs in the *HMGCR* gene with T2DM, LDL-C and glycaemic/adiposity traits

RSID	paper	Effect/other allele	Effect on LDL-C	TRAITS							
				Effect on T2DM	fasting plasma glucose	fasting plasma insulin	BMI	Weight	Waist circum-ference	Hip circum-ference	Waist-to-hip ratio
rs12916	Lotta L.A et al (2016)	T / C	-0.07 (-0.08, -0.07) [n/a] (standardised units per 1 mmol/L reduction in LDL-C)	1.03 (1.01, 1.05)* (OR per LDL-C lowering allele)	0.00 (-0.01, 0.01)[n/a] (SD units per 1 mmol/L reduction in LDL-C)	0.01 (0.00, 0.02)[n/a] (standardised units per 1 mmol/L reduction in LDL-C)	0.02 (0.01, 0.02)[n/a] (SD units per 1 mmol/L reduction in LDL-C)	n/a	n/a	n/a	0.01 (0.00, 0.01)[n/a] (SD per 1 mmol/L reduction in LDL-C)
	Ference B.A. et el (2016)	T / C	-0.061 (-0.067, -0.055)*** (mg/dl per effect allele)	1.09 (1.00, 1.18)[n/a] (per 10 mg/dl lower LDL-C)	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	Wurtz P. et al (2016)	T / C	0.046 (0.054, 0.037)*** (Concentration difference per effect allele)	n/a	0.01 (0.00, 0.02) p>0.0006 (Concentration difference per effect allele)	n/a	n/a	n/a	n/a	n/a	n/a
	Tragante V. et al (2016)	T / C	lower level of trait	higher level of risk of outcome	higher level of trait	higher level of trait	n/a	n/a	n/a	n/a	n/a
	Swerdlow D.I. et al (2015)	T/C	-0.077 (0.091, -0.062)*** (mmol/l per effect allele)	1.06 (1.03, 1.09)*** (OR per effect allele)	0.123 (-0.135, 0.387)(% per effect allele)	0.659 (-0.631, 1.967)(% per effect allele)	0.083 (0.031, 0.135)* (kg/m2 per effect allele)	0.203 (0.043, 0.362)* (kg per effect allele)	0.303 (0.102, 0.504)**(cm per effect allele)	0.193 (0.050, 0.337)** (cm per effect allele)	n/a

RSID	paper	Effect/other allele	Effect on LDL-C	TRAITS							
				Effect on T2DM	fasting plasma glucose	fasting plasma insulin	BMI	Weight	Waist circum-ference	Hip circum-ference	Waist-to-hip ratio
	Ference B.A. et el (2016)	G/T	-0.052 (0.062, -0.042)*** (mg/dl per effect allele)	1.16 (1.03, 1.31) [n/a] (per 10 mg/dl lower LDL-C)	n/a	n/a	n/a	n/a	n/a	n/a	n/a
rs3846663	Knouff C.W. et al (2008)	C/T	data reported in mean +-SD per allele AA 3.29 ± 0.04; Aa 3.36± 0.04***; aa 3.42 ± 0.07***	1.00 (0.8, 1.2) (OR per effect allele)	data reported in mean +-SD per allele AA 5.47± 0.03; Aa 5.46± 0.03; aa 5.46± 0.05	n/a	n/a	n/a	n/a	n/a	n/a
	Tragante V. et al (2016)	C/T	lower level of trait	higher level of risk of outcome	higher level of trait	higher level of trait	n/a	n/a	n/a	n/a	n/a
rs5744707	Lotta L.A et al (2016)	A/G	-0.05 (-0.07, -0.04)	0.98 (0.96, 1.01) (OR per LDL-C lowering allele)	0.01 (-0.01, 0.02)[n/a] (SD units per 1 mmol/L reduction in LDL-C)	0.01 (-0.02,0.02)[n/a] (SD units per 1 mmol/L reduction in LDL-C)	0.00 (-0.01, 0.01)[n/a] (SD units per 1 mmol/L reduction in LDL-C)	n/a	n/a	n/a	-0.01 (-0.02, 0.01)[n\a] (SD units per 1 mmol/L reduction in LDL-C)
rs16872526	Lotta L.A et al (2016)	T/G	-0.04 (-0.05, -0.03)	1.016 (0.985, 1.047) (OR per LDL-C lowering allele)	0.01 (-0.00, 0.03)[n/a] (SD units per 1 mmol/L reduction in LDL-C)	0.01 (-0.01, 0.03)[n/a] (SD units per 1 mmol/L reduction in LDL-C)	0.02 (0.01, 0.03)[n/a] (SD units per 1 mmol/L reduction in LDL-C)	n/a	n/a	n/a	0.01 (-0.01, 0.02)[n/a] (SD units per 1 mmol/L reduction in LDL-C)

11RSID	paper	Effect/other allele	Effect on LDL-C	TRAITS							
				Effect on T2DM	fasting plasma glucose	fasting plasma insulin	BMI	Weight	Waist circum-ference	Hip circum-ference	Waist-to-hip ratio

rs5909	Ference B.A. et el (2016)	G/A	-0.051 (-0.065, -0.037)*** (mg/dl per effect allele)	0.90 (0.71, 1.14) [n/a] (per 10 mg/dl lower LDL-C)	n/a	n/a	n/a	n/a	n/a	n/a	n/a
rs2303152	Ference B.A. et el (2016)	G/A	-0.035 (-0.045, -0.025)	0.86 (0.63, 1.16) [n/a] (per 10 mg/dl lower LDL-C	n/a	n/a	n/a	n/a	n/a	n/a	n/a
rs10066707	Ference B.A. et el (2016)	A/G	-0.041 (0.050, -0.032)*** (mg/dl per effect allele)	1.06 (0.91, 1.24) [n/a] (per 10 mg/dl lower LDL-C)	n/a	n/a	n/a	n/a	n/a	n/a	n/a
rs2006760	Ference B.A. et el (2016)	C/G	-0.044 (-0.057, -0.032)*** (mg/dl per effect allele)	1.08 (1.02, 1.14) [n/a] (per 10 mg/dl lower LDL-C)	n/a	n/a	n/a	n/a	n/a	n/a	n/a
rs3846662	Tragante V. et al (2016)	n/a	lower level of trait	higher level of risk of outcome	higher level of trait	higher level of trait	n/a	n/a	n/a	n/a	n/a
rs7703051	Tragante V. et al (2016)	n/a	lower level of trait	higher level of risk of outcome	higher level of trait	higher level of trait	n/a	n/a	n/a	n/a	n/a
rs12654264	Tragante V. et al (2016)	n/a	lower level of trait	higher level of risk of outcome	higher level of trait	higher level of trait	n/a	n/a	n/a	n/a	n/a

Effects are shown per one copy of LDL-C lowering allele. Point estimates with P value < 0.05 marked with *, P value <0.001 marked with **, P value $<5*10^{-8}$ marked with ***.

2.3 Appendix 2:3 | The associations of genetic variants in the *NPC1L1* gene with T2DM, anthropometric measures and glycaemic traits

rsID	Paper	Pos	Effect allele/ other allele	MAF	Effect allele freq	Ancestry	Trait	Beta/OR	CI	N	Cases	Controls	N studies	Unit
Association of *NPC1L1* SNPs and genetic score with T2DM														
rs2073547	Lotta L.A. *et al* (2016)	chr7:44582331	A/G	0.19	0.81	Europe and the United States	T2DM	1.05 OR per allele	1.03, 1.08	321044	50 775	270 269	3	OR per effect allele
rs217386	Lotta L.A. *et al* (2016)	chr7:44600695	A/G	0.42	0.42	Europe and the United States	T2DM	1.03 OR per allele	1.01, 1.05	321044	50 775	270 269	3	OR per effect allele
GRS (SNP:2)	Lotta L.A. *et al* (2016)					Europe and the United States	T2DM	2.42 OR	1.70, 3.43	321044	51 775	271 269	3	OR per 1 SD higher LDL-C
GRS (SNP:5)	Ference B.A. *et al* (2016)					European	T2DM	1.073 OR	0.98, 1.18	108376	10635	97741	14	OR per 10mg/dl higher LDL-C
Association of *NPC1L1* SNPs and genetic score with glycaemic traits														
rs2073547	Lotta L.A. *et al* (2016)	chr7:44582331	A/G	0.19	0.81	Europe and the United States	FPG	0.01 SDU	0.00, 0.03	133010			3	SD per effect allele
rs217386	Lotta L.A. *et al* (2016)	chr7:44600695	A/G	0.42	0.42	Europe and the United States	FPG	0.00 SDU	-0.01 to 0.01	133010			3	SD per effect allele
GRS (SNP:2)	Lotta L.A. *et al* (2016)					Europe and the United States	FPG	0.14 SDU	-0.05, 0.33	133010			3	SD per 1 mmol/L reduction in LDL-C

rsID	Paper	Pos	Effect allele/ other allele	MAF	Effect allele freq	Ancestry	Trait	Beta/OR	CI	N	Cases	Controls	N studies	Unit
rs2073547	Lotta L.A. *et al* (2016)	chr7:44582331	A/G	0.19	0.81	Europe and the United States	Ln-FI	0.00 SDU	-0.02 to 0.01	108557			3	SD per effect allele
rs217386	Lotta L.A. *et al* (2016)	chr7:44600695	A/G	0.42	0.42	Europe and the United States	Ln-FI	-0.01 SDU	-0.02 to 0.00	108557			3	SD per effect allele
GRS (SNP:2)	Lotta L.A. *et al* (2016)					Europe and the United States	Ln-FI	-0.15 SDU	-0.37 to 0.07)	108557			3	SD per 1 mmol/L reduction in LDL-C
Association of *NPC1L1* SNPs and genetic score with anthropometric traits														
rs2073547	Lotta L.A. *et al* (2016)	chr7:44582331	A/G	0.19	0.81	Europe and the United States	BMI	0.00 SDU	-0.01 to 0.01	333495			3	SD per effect allele
rs217386	Lotta L.A. *et al* (2016)	chr7:44600695	A/G	0.42	0.42	Europe and the United States	BMI	0.00 SDU	0.00 to 0.01	333495			3	SD per effect allele
GRS (SNP:2)	Lotta L.A. *et al* (2016)					Europe and the United States	BMI	0.08 SDU	-0.04 to 0.2	333495			3	SD per 1 mmol/L reduction in LDL-C
rs2073547	Lotta L.A. *et al* (2016)	chr7:44582331	A/G	0.19	0.81	Europe and the United States	WHR	0.00 SDU	-0.01 to 0.01	224047			3	SD per effect allele
rs217386	Lotta L.A. *et al* (2016)	chr7:44600695	A/G	0.42	0.42	Europe and the United States	WHR	0.00 SDU	-0.01 to 0.01	224047			3	SD per effect allele

rsID	Paper	Pos	Effect allele/ other allele	MAF	Effect allele freq	Ancestry	Trait	Beta/OR	CI	N	Cases	Controls	N studies	Unit
GRS (SNP:2)	Lotta L.A. *et al* (2016)					Europe and the United States	WHR	-0.01 SDU	-0.14 to 0.12	224047			3	SD per 1 mmol/L reduction in LDL-C
Association of *NPC1L1* SNPs and genetic score with lipids														
rs2073547	Lotta L.A. *et al* (2016)	chr7:44582331	A/G	0.19	0.81	Europe and the United States	LDL-C	-0.05 mmol/l per allele	-0.06, 0.04	169 889			3	b mmol/l per effect allele
rs217386	Lotta L.A. *et al* (2016)	chr7:44600695	A/G	0.42	0.42	Europe and the United States	LDL-C	-0.04 mmol/l per allele	-0.04, -0.03	173 021			3	b mmol/l per effect allele
rs217386	Ference B.A. *et al* (2016)	chr7:44600695	A/G	0.41	0.41	European	LDL-C	-1.1253 mg/dl	-1.36, -0.89	173021			14	mg/dl per effect allele
rs2073547	Ference B.A. *et al* (2016)	chr7:44582331	A/G	0.19	0.81	European	LDL-C	-1.5035 mg/dl	-1.74, -1.27	169889			14	mg/dl per effect allele
rs7791240	Ference B.A. *et al* (2016)	chr7:44602589	T/C	0.09	0.91	European	LDL-C	-1.3175 mg/dl	-1.68, -0.96	161845			14	mg/dl per effect allele
rs10234070	Ference B.A. *et al* (2016)	chr7:44537696	C/T	0.1	0.90	European	LDL-C	-0.9145 md/dl	-1.27, -0.56	160051			14	mg/dl per effect allele
rs2300414	Ference B.A. *et al* (2016)	chr7:44682938	G/A	0.07	0.93	European	LDL-C	-1.0943 mg/dl	-1.58, -0.61	115286			14	mg/dl per effect allele

rs2073547	Lotta L.A. *et al* (2016)	chr7:44582331	A/G	0.19	0.81	Europe and the United States	HDL	0.00 SDU	0.00, 0.01	187087		3	SD per effect allele
rs217386	Lotta L.A. *et al* (2016)	chr7:44600695	A/G	0.42	0.42	Europe and the United States	HDL	0.00 SDU	-0.01, 0.01	187087		3	SD per effect allele

2.4 Appendix 2:4 | The associations of SNPs in the *NPC1L1* gene with T2DM, LDL-C and glycaemic/adiposity traits

RSID	paper	Effect/ other allele	Effect on LDL-C	TRAITS							
				Effect on T2DM	Fasting plasma glucose	Fasting plasma insulin	BMI	Weight	Waist circum-ference	Hip circum-ference	Waist-to-hip ratio
rs2073547	Lotta L.A *et al* (2016)	A/G	-0.05 (-0.05, -0.04) SD unit per effect allele	OR 1.05(1.03, 1.08) per effect allele	0.01 (0.00, 0.03) SD per effect allele	0.00 (-0.02, 0.01) SD per effect allele	0.00 (-0.01, 0.01) SD per effect allele	n/a	n/a	n/a	0.00 (-0.01, 0.01) SD per effect allele
	Ference B.A. *et el* (2016)	A/G	-1.504 (-1.74, -1.27) mg/dl per effect allele	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
rs217386	Lotta L.A *et al* (2016)	A/G	-0.04 (-0.03, -0.04) SD unit per effect allele	OR 1.03(1.01, 1.05) per effect allele	0.01 (-0.01, 0.02) SD per effect allele	-0.01 (-0.02, 0.00) SD per effect allele	0.00 (0.0, 0.01) SD per effect allele	n/a	n/a	n/a	0.00 (-0.01, 0.01) SD per effect allele
	Ference B.A. *et el* (2016)	A/G	-1.125 (-1.36, -0.89) mg/dl per effect allele	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
rs7791240	Ference B.A. *et el* (2016)		-1.318 (-1.68, -0.96) mg/dl per effect allele	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
rs10234070	Ference B.A. *et el* (2016)		-0.915 (-1.27, -0.56) mg/dl per effect allele	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
rs2300414	Ference B.A. *et el* (2016)		-1.094 (-1.58, -0.61) mg/dl per effect allele	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

2.5 Appendix 2:5 | Statins RCTs included in the meta-analyses

RCT	Type of study	N developing diabetes/N non-diabetic		Rahal A.J. et al (2016)	Swerdlow D., et al (2015)	Cai R., et al (2014)	Taylor F, et al (2013)	Preiss D., Sattar N. (2011)	Sattar N., et al (2010)	Rajpathak S.N et al (2009)	Coleman C. et al (2009)	Report-ed (N)
N of studies:				14	20	14	2	18	13	6	5	
N of diabetic/non-diabetic:				2393/47444 vs 2167/47481	3858/64558 vs 3481/64612	2392/47504 vs 2167/47598	342/12205 vs 290/12202	1449/16 408 vs 1300/16344	4278/91140	1088/28842 vs 994/28751	1407/39791	
Effect on T2DM (OR):				1.11 (1.00,1.2)	1.12 (1.06,1.18)	1.16 (1.06,1.28)	1.18 (1.01,1.39)	1.09 (1.02,1.17)	1.09 (1.02,1.17)	1.13 (1.03,1.23)	1.03 (0.89,1.19)	
4S	Placebo /controlled	198/2116	vs 193/2126	x	x	x		x	x			5
A to Z	int/mod dose	530/2230	vs 529/2262		x			x				2
AFCAPS/TEXCAPS	Placebo /controlled	72/3094	vs 74/3117	x	x	x	x	x	x			6
ALLHAT-LLT	Placebo /controlled	238/3017	vs 212/3070	x	x	x		x	x			5
ASCOT	Placebo /controlled	154/3910	vs 134/3863							x		1
ASCOT-LLA	Placebo /controlled	61/2189	vs 70/2256	x	x	x		x	x		x	6
CORONA	Placebo /controlled	100/1771	vs 88/1763	x	x	x		x	x	x	x	7
GISSI HF	2x2 factorial design	225/1660	vs 215/1718	x	x	x		x	x			5
GISSI PREV	2x2 factorial design	96/1743	vs 105/1717	x	x	x		x	x			5

N of diabetic/non-diabetic:			2393/47444 vs 2167/47481	3858/64558 vs 3481/64612	2392/47504 vs 2167/47598	342/12205 vs 290/12202	1449/16 408 vs 1300/16344	4278/91140	1088/28842 vs 994/28751	1407/39791	
RCT	Type of study	N developing diabetes/N non-diabetic	Rahal A.J. et al (2016)	Swerdlow D., et al (2015)	Cai R., et al (2014)	Taylor F, et al (2013)	Preiss D., Sattar N. (2011)	Sattar N., et al (2010)	Rajpathak S.N et al (2009)	Coleman C. et al (2009)	Report-ed (N)
HPS	Placebo /controlled	335/7291 vs 293/7282	x	x	x		x	x	x	x	7
IDEAL	int/mod dose	240/3907 vs 209/3912		x			x				2
JUPITER	Placebo /controlled	270/8901 vs 216/8901	x	x	x	x	x	x	x		7
LIPID	Placebo /controlled	126/3150 vs 138/3067	x	x	x		x	x	x	x	7
LIPS	Placebo /controlled	172/3970 vs 181/3967		x							1
MEGA	drug +diet/diet	172/3013 vs 164/3073	x	x	x		x	x			5
PROSPER	Placebo /controlled	165/2510 vs 127/2513	x	x	x		x	x			5
PROVE IT-TIMI 22	Int /mod dose	101/1707 vs 99/1688		x			x				2
SEARCH	Int /mod dose	625/6031 vs 587/6033		x			x				2
SPARCL	Placebo /controlled	166/1908 vs 115/1916	x	x	x						3
TNT	Int /mod dose	308/4253 vs 351/4274		x			x				2
WOSCOPS	Placebo /controlled	75/2999 vs 93/2975	x	x	x		x	x	x	x	7

2.6 Appendix 2:6 | RCTs of statins included in each meta-analysis

Reference	Drug	Trait	OR	95% CI/SE	P-value	N of studies	STATIN GROUP / Intensive dose	CONTROL GROUP / Moderate dose	N of parti-cipants	Mean follow-up (yr)	POPULATION/ RCTs
							Cases/Total, No.(%)	Cases/Total, No.(%)			
NEW-ONSET DIABETES MELLITUS											
Rahal A.J. et al (2016)	Statins	New DM	1.11	1.0, 1.2	0.007	14	2392/47444 (5.0%)	2167/47481 (4.6)	94925	4	non diabetic / ASCOT-LLA HPS JUPITER WOSCOPS LIPID CORONA PROSPER MEGA AFCAPS TexCAPS 4s LLT-ALLHAT GISSI HF GISSI PREVENZIONE SPARCL
Swerdlow D., et al (2015)	Statins	New DM	1.12	1.06, 1.18	n/a	20	3858/64558 (6.0%)	3481/64612 (5.4%)	129170	4.2	non-diabetic / 4S, WOSCOPS, AFCAPS_TexCAPS, LIPID, GISSI_PREVENZIONE, LIPS, HPS, ALLHAT-LLT, PROSPER, ASCOT-LLA, SPARCL, MEGA, CORONA, JUPITER, GISSI-HF, PROVE-IT TIMI22, AtoZ, TNT, IDEAL, SEARCH
Cai R., et al (2014)	Statins intensified LDL-C target <1.8 mmol/L	New DM	1.33	1.14, 1.56	n/a	2	436/10900 (4.0%)	331/10677 (3.1%)	21577	2	non-diabetic participants/ JUPITER, SPARCL
Cai R., et al (2014)	Statins moderate LDL-C target 1.8 to 2.59 mmol/L	New DM	1.16	1.06, 1.28	n/a	5	979/17175 (5.7%)	857/17140 (5.0%)	34315	2	non-diabetic participants/ HPS, ASCOT-LLA, CORONA, PROSPER, GISSI-HF
Cai R., et al (2014)	Statins standard LDL-C target 2.59 mmol/L	New DM	1.01	0.92, 1.10	n/a	7	977/19540 (5.0%)	979/19580 (5.0%)	39120	2	non-diabetic participants/ 4S, WOSCOPS, LIPID, AFCAPS TexCAPS, ALLHAT-LLT, MEGA, GISSI PREVENZIONE

Reference	Drug	Trait	OR	95% CI/SE	P-value	N of studies	STATIN GROUP / Intensive dose	CONTROL GROUP / Moderate dose	N of parti-cipants	Mean follow-up (yr)	POPULATION/ RCTs
							Cases/Total, No.(%)	Cases/Total, No.(%)			
Taylor F, et al (2013)	Statins	New DM	1.18	1.01 , 1.39	0.037	2	342/12,205 (2.8%)	290/12202 (2.4%)	56934	0.2	n/a / JUPITER, AFCAPS-TexCAPS
Preiss D., et al (2011)	Intensive/ moderate dose	New DM	1.12	1.04, 1.22	0.60	5	1449/16408 (8.8%)	1300/16344 (8.0%)	32 752	4.9	non-deabetic / TNT, IDEAL, A to Z, PROVE-IT-TIMI, SEARCH
Sattar N., et al (2010)	Statins	New DM	1.09	1.02, 1.17	0.61	13	2226/45521 (4.9%)	2052/45619 (4.5%)	91 140	4	non-diabetic / ASCOT-LLA, HPS, JUPITER, WOSCOPS, LIPID, CORONA, PROSPER, MEGA, AFCAPS TexCAPS, 4S ALLHAT-LLT, GISSI HF, GISSI PREVENZIONE
Rajpathak S.N et al (2009)	Statins	New DM	1.13	1.03, 1.23	0.08	5	1088/28842 (3.8.0%)	994/28751 (3.5%)	57,593	3.9	both / HPS, LIPID, ASCOT, JUPITER, CORONA
Coleman C. et al (2009)	Statins	New DM	1.03	0.89, 1.19	n/a	5	n/a	n/a	39,791	2.7–6.0	N/A
ADIPOSITY TRAIT											
Swerdlow D., et al (2015)	Statins	Bodyweight	0.24 kg	0.10, 0.38	n/a	15	/45724	/45669	91393	3.9	non-diabetic / 4S WOSCOPS AFCAPS_TexCAPS LIPID GISSI_PREVENZIONE PROSPER ASCOT-LLA SPARCL MEGA CORONA JUPITER GISSI-HF PROVE-IT TNT IDEAL

Appendices Chapter 3

Appendix 3:1	The list of oral hypoglycaemic agents and insulin by class with UK Biobank codes
Appendix 3:2	UK Biobank non-cancer illnesses codes for diabetes
Appendix 3:3	ICD-10 codes for diabetes
Appendix 3:4	The algorithm for defining prevalent diabetes cases
Appendix 3:5	The algorithms for discrimination between type 1 and type 2 diabetes
Appendix 3:6	The algorithm computing incident cases of T2DM
Appendix 3:7	SNPs within 100 kb from the *HMGCR* gene which are associated with LDL-C at GWAS level of significance
Appendix 3:8	The results of the *HMGCR* SNPs clumping procedure
Appendix 3:9	The comparison of genetic data for *HMGCR* index SNPs in UK Biobank, GLGC, and Ensemble databases
Appendix 3:10	The distribution of the *HMGCR* GRS
Appendix 3:11	SNPs within 100 kb from the *NPC1L1* gene which are associated with LDL-C at GWAS level of significance
Appendix 3:12	The results of the *NPC1L1* SNPs clumping procedure
Appendix 3:13	The comparison of genetic data for *NPC1L1* index SNPs in UK Biobank, GLGC, and Ensemble databases
Appendix 3:14	The distribution of the *NPC1L1* GRS

3.1 Appendix 3:1 | The list of oral hypoglycaemic agents and insulin by class with UK Biobank codes

Medication class and trade name	UK Biobank medication code
Insulin	
insulin product	1140883066
Alpha-glucosidase inhibitors	
acarbose	1140868902
glucobay 50mg tablet	1140874646
Glitazone	
actos	1141171652
pioglitazone	1141171646
rosiglitazone	1141177600
rosiglitazone 1mg / metformin 500mg tablet	1141189090
Sulfonylurea	
diamicron 80mg tablet	1140874746
diabetamide 2.5mg tablet	1140874736
glibenclamide	1140874718
gliclazide	1140874744
glimepiride	1141152590
amaryl 1mg tablet	1141156984
minodiab 2.5mg tablet	1140874652
glipizide product	1141157284
semi-daonil 2.5mg tablet	1140874726
daonil 5mg tablet	1140874724
glibenese 5mg tablet	1140874650
glyclizide	1140910566
glipizide	1140874646

tolbutamide	1140874674
tolanase 100mg	1140874666
Metformin	
avandamet 1mg / 500mg tablet	1141189094
metformin	1140884600
glucophage 500mg tablet	1140874686
Meglitinide	
starlix 60mg tablet	1141173786
repaglinide	1141168660
nateglinide	1141173882
Novonorm	1141168668

3.2 Appendix 3:2 | UK Biobank non-cancer illnesses codes for diabetes

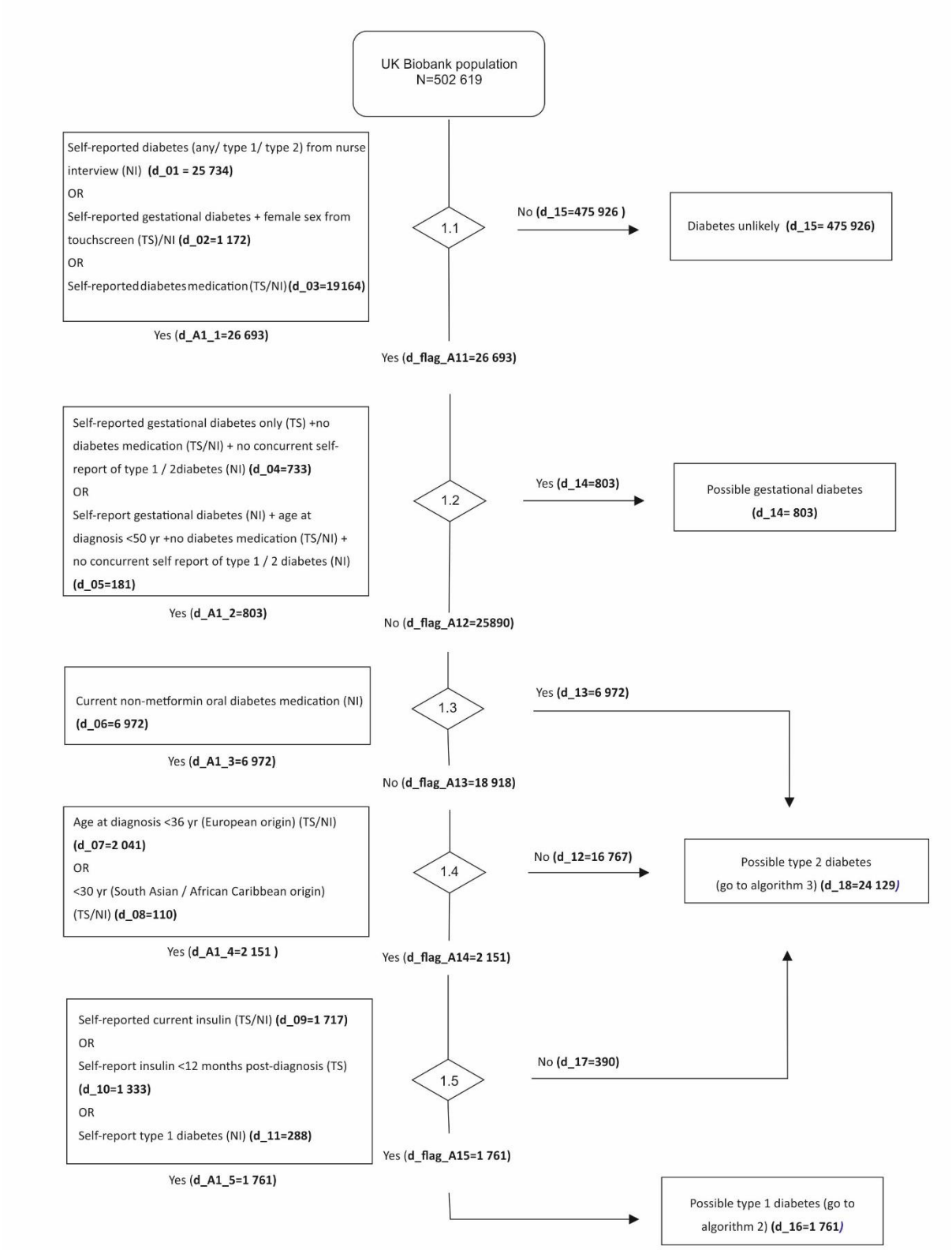
Diagnosis	UKB specific code
Unspecified Diabetes Mellitus	1220
Type 1 Diabetes Mellitus	1222
Type 2 Diabetes Mellitus	1223

3.3 Appendix 3:3 | List of ICD-10 codes for diabetes

Type of Diabetes	Code	Description
Type 2 Diabetes Mellitus	E110	Type 2 diabetes mellitus with hyperosmolarity
Type 2 Diabetes Mellitus	E112	Type 2 diabetes mellitus with kidney complications
Type 2 Diabetes Mellitus	E113	Type 2 diabetes mellitus with ophthalmic complications
Type 2 Diabetes Mellitus	E114	Type 2 diabetes mellitus with neurological complications
Type 2 Diabetes Mellitus	E115	Type 2 diabetes mellitus with circulatory complications
Type 2 Diabetes Mellitus	E116	Type 2 diabetes mellitus with other specified complications
Type 2 Diabetes Mellitus	E118	Type 2 diabetes mellitus with unspecified complications
Type 2 Diabetes Mellitus	E119	Type 2 diabetes mellitus without complications
Unspecified Diabetes Mellitus	E140	Unspecified diabetes mellitus
Unspecified Diabetes Mellitus	E142	Unspecified diabetes mellitus with renal complications
Unspecified Diabetes Mellitus	E143	Unspecified diabetes mellitus with ophthalmic complications
Unspecified Diabetes Mellitus	E144	Unspecified diabetes mellitus with neurological complications
Unspecified Diabetes Mellitus	E145	Unspecified diabetes mellitus with peripheral circulatory complications
Unspecified Diabetes Mellitus	E146	Unspecified diabetes mellitus with other specified complications
Unspecified Diabetes Mellitus	E147	Unspecified diabetes mellitus with multiple complications
Unspecified Diabetes Mellitus	E148	Unspecified diabetes mellitus with unspecified complications
Unspecified Diabetes Mellitus	E149	Unspecified diabetes mellitus without complications
Unspecified Diabetes Mellitus	H360	Diabetic retinopathy
Unspecified Diabetes Mellitus	G590	Diabetic mononeuropathy
Unspecified Diabetes Mellitus	G632	Diabetic polyneuropathy
Unspecified Diabetes Mellitus	H280	Diabetic cataract
Unspecified Diabetes Mellitus	M142	Diabetic arthropathy
Unspecified Diabetes Mellitus	N083	Glomerular disorders in diabetes mellitus

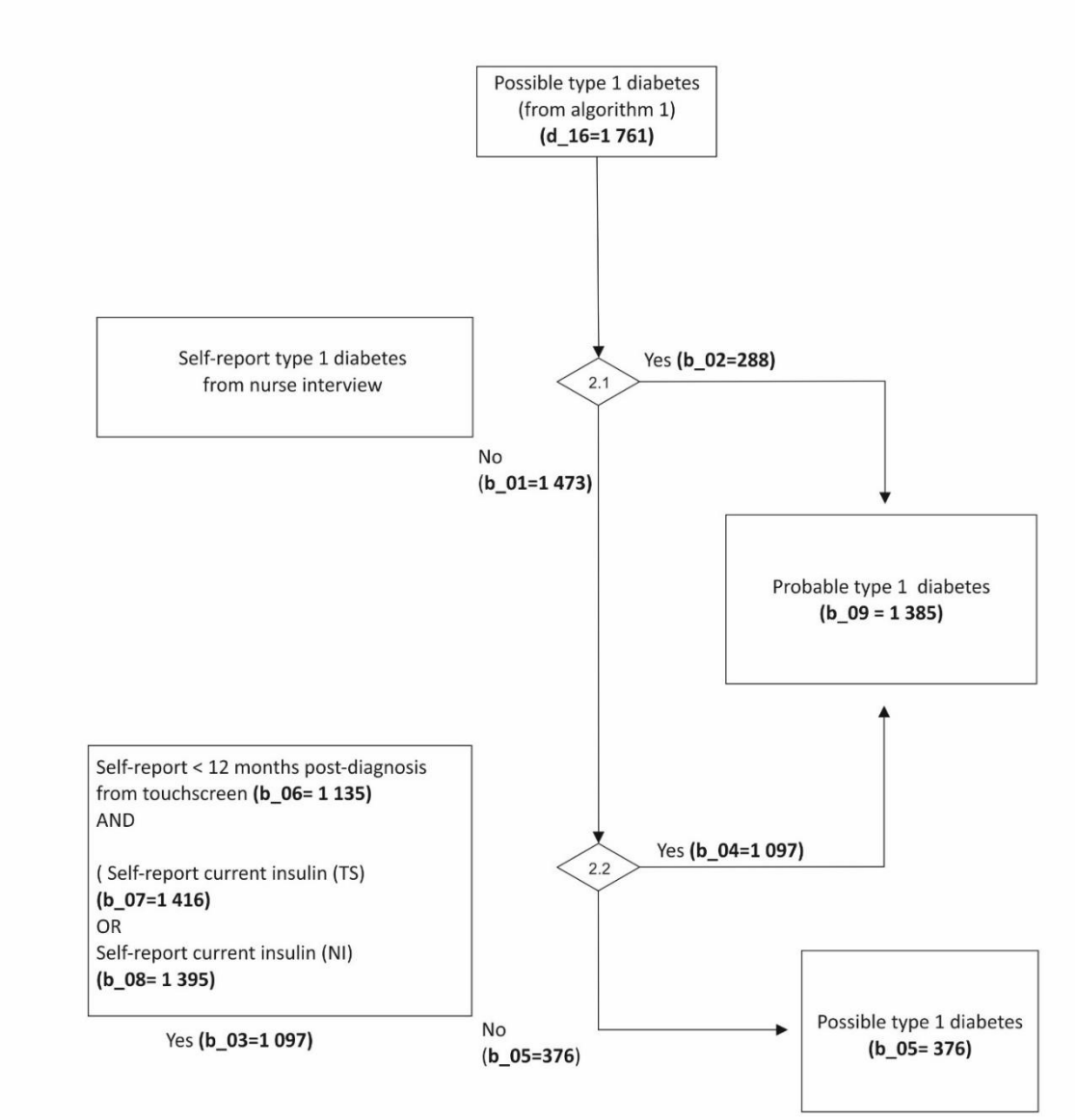
Type 1 Diabetes Mellitus	E100	Type 1 diabetes mellitus with coma
Type 1 Diabetes Mellitus	E101	Type 1 diabetes mellitus with ketoacidosis
Type 1 Diabetes Mellitus	E102	Type 1 diabetes mellitus with renal complications
Type 1 Diabetes Mellitus	E103	Type 1 diabetes mellitus with ophthalmic complications
Type 1 Diabetes Mellitus	E104	Type 1 diabetes mellitus with neurological complications
Type 1 Diabetes Mellitus	E105	Type 1 diabetes mellitus with peripheral circulatory complications
Type 1 Diabetes Mellitus	E106	Type 1 diabetes mellitus with other specified complications
Type 1 Diabetes Mellitus	E107	Type 1 diabetes mellitus with multiple complications
Type 1 Diabetes Mellitus	E108	Type 1 diabetes mellitus with unspecified complications
Type 1 Diabetes Mellitus	E109	Type 1 diabetes mellitus without complications

3.4 Appendix 3:4 | The algorithm for defining prevalent diabetes cases

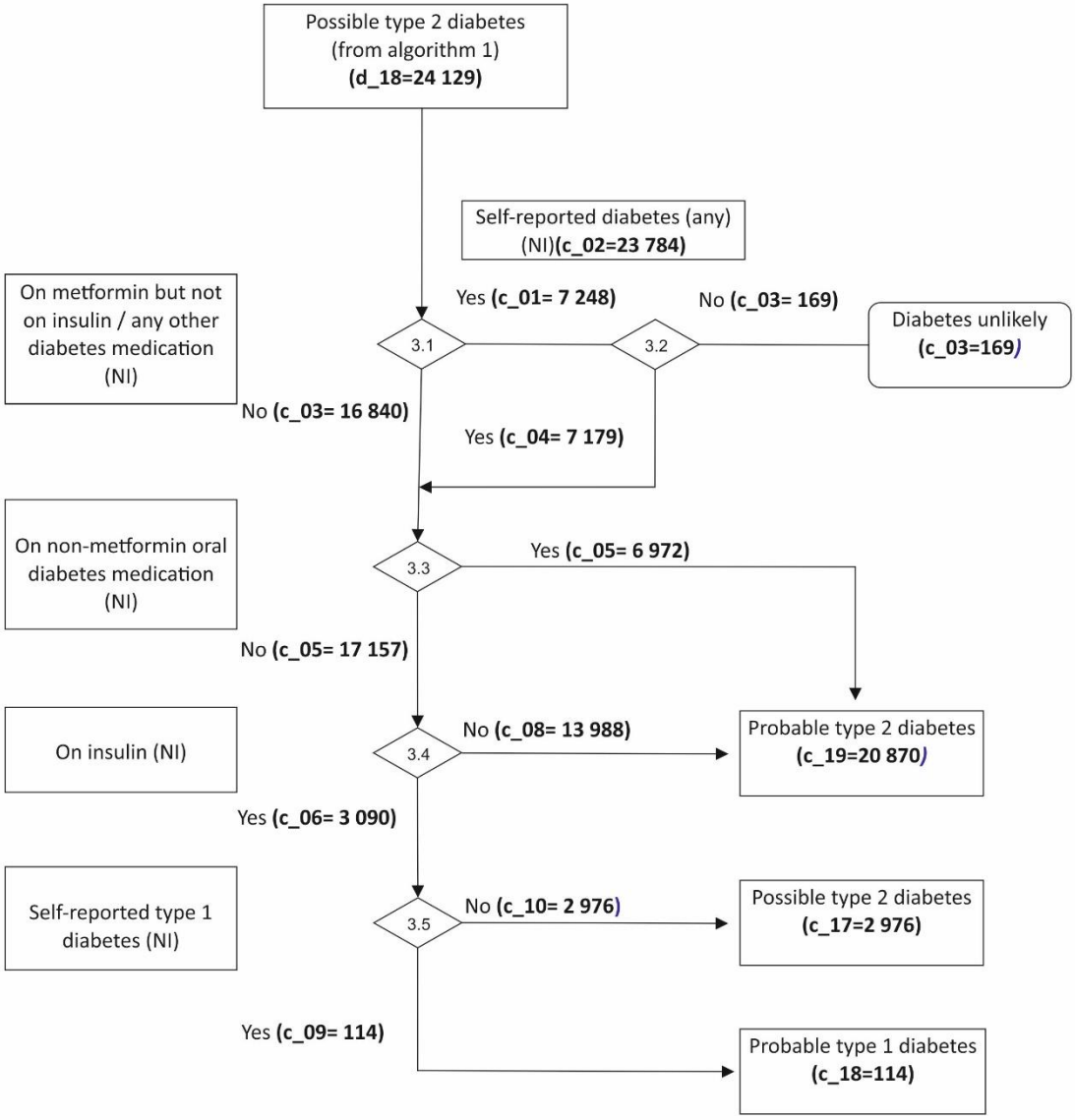


Prevalence algorithm 1: Distinction between diabetes presence or absence, and initial sorting of diabetes type using baseline UK Biobank assessment data. Results of the current analysis are shown in comparison with Eastwood's paper (blue numbers in italic) .

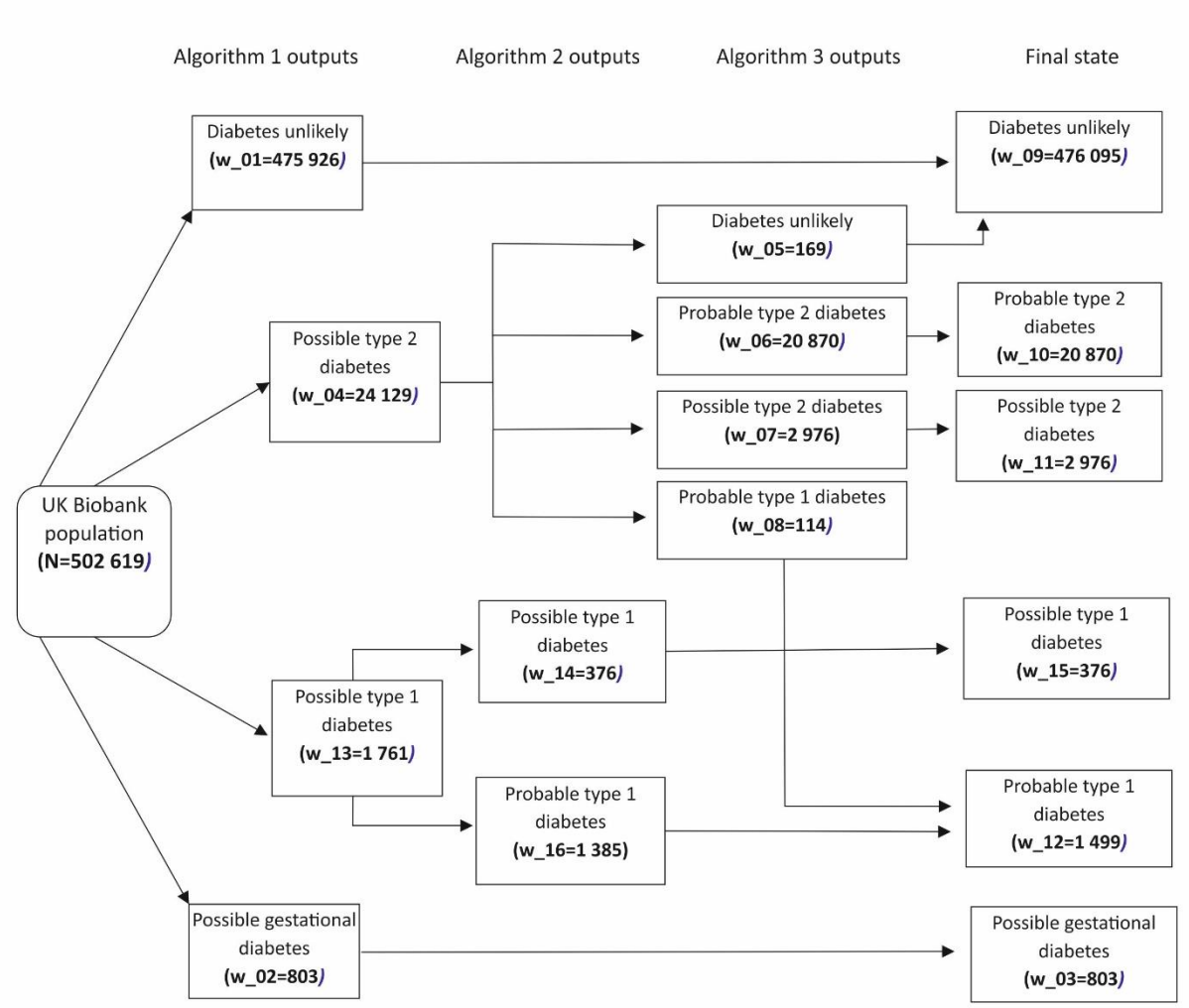
3.5 Appendix 3:5 | The algorithms for discrimination between type 1 and type 2 diabetes



Prevalence algorithm 2: Finalising type 1 diabetes diagnosis and classification into probable and possible categories.

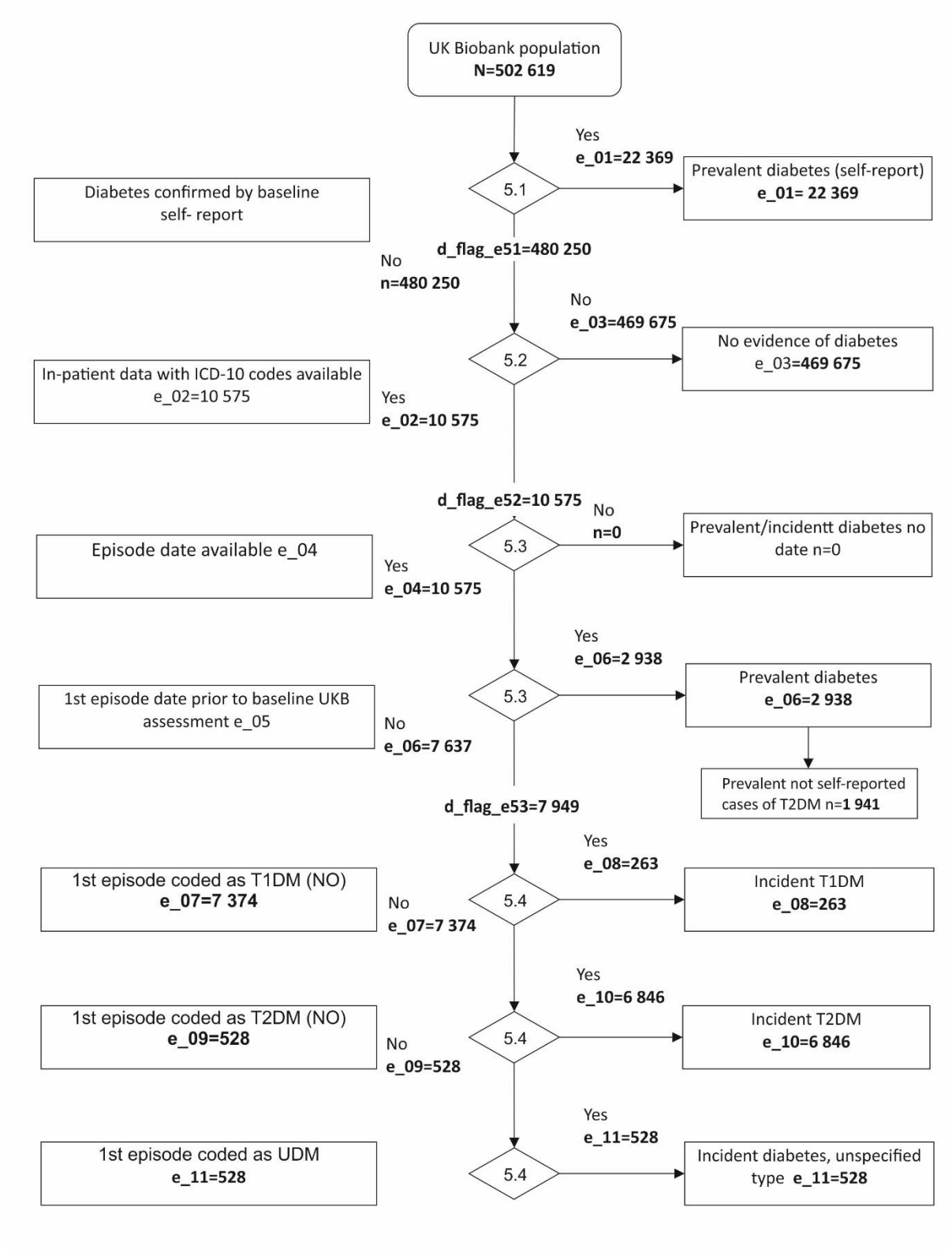


Prevalence algorithm 3: Finalising type 2 diabetes diagnosis and classification into probable and possible categories.



Final diabetic diagnostic status in UK Biobank.

3.6 Appendix 3:6 | The algorithm computing incident cases of T2DM



3.7 Appendix 3:7 | SNPs within 100 kb from the *HMGCR* gene which are associated with LDL-C at GWAS level of significance

Chromosome and bp position	RSID	Minor allele	Major allele	Effect size	SE	P value	Minor allele frequency
chr5:74534372	rs6866661	g	a	0.0465	0.0053	2.64E-18	0.4182
chr5:74536723	rs10064936	a	c	0.0348	0.0062	1.78E-09	0.2731
chr5:74537751	rs4704197	t	c	0.0466	0.0053	2.23E-18	0.4195
chr5:74540187	rs9942407	a	g	0.0489	0.0053	3.13E-19	0.4182
chr5:74540397	rs7711235	g	a	0.0375	0.0063	5.00E-10	0.2678
chr5:74560579	rs10066707	a	g	0.0497	0.0054	2.97E-19	0.4169
chr5:74562564	rs2112653	c	t	0.0583	0.0052	1.25E-26	0.4749
chr5:74565153	rs10062361	t	c	0.0695	0.0043	5.41E-52	0.2348
chr5:74569028	rs72633962	c	t	0.0600	0.0072	3.33E-15	0.1412
chr5:74569856	rs4704200	t	g	0.0644	0.0037	1.30E-61	0.4459
chr5:74570531	rs1551894	a	g	0.0694	0.0045	4.17E-48	0.2348
chr5:74571914	rs2052515	a	g	0.0691	0.0043	1.00E-51	0.2348
chr5:74574984	rs2126736	g	a	0.0642	0.0037	2.62E-61	0.4354
chr5:74576285	rs7722186	a	g	0.0613	0.0051	1.63E-29	0.4459
chr5:74582162	rs80037857	a	g	0.0594	0.0088	2.12E-10	0.0950
chr5:74596682	rs918623	g	t	0.0651	0.0043	5.82E-49	0.2467
chr5:74602699	rs1423527	a	c	0.0697	0.0037	1.33E-72	0.4380
chr5:74603479	rs2335418	a	g	0.0654	0.0037	8.58E-65	0.4578
chr5:74604742	rs6871667	a	g	0.0675	0.0037	1.28E-68	0.4327
chr5:74605220	rs10056811	a	g	0.0719	0.0039	3.58E-69	0.3747
chr5:74608732	rs11957260	t	c	0.0657	0.0037	1.06E-65	0.4538
chr5:74611574	rs6892023	t	g	0.0652	0.0062	5.91E-23	0.4538
chr5:74615021	rs17671591	t	c	0.0713	0.0038	6.77E-70	0.3813
chr5:74616843	rs10474433	c	t	0.0737	0.0055	5.95E-38	0.3799
chr5:74617262	rs2878417	a	g	0.0704	0.0037	5.85E-74	0.4354
chr5:74618186	rs13358429	a	g	0.0713	0.0039	5.80E-70	0.3799
chr5:74619132	rs13354746	t	c	0.0696	0.0043	2.19E-53	0.2467
chr5:74620225	rs3895886	t	c	0.0420	0.0061	1.02E-10	0.1280
chr5:74620377	rs3857388	c	t	0.0421	0.0059	2.20E-11	0.1280
chr5:74624234	rs6895057	a	g	0.0478	0.0060	4.70E-14	0.2665
chr5:74624482	rs3843480	c	t	0.0754	0.0044	4.40E-60	0.4103
chr5:74625487	rs7703051	a	c	0.0727	0.0037	1.40E-77	0.4103

Chromosome and bp position	RSID	Minor allele	Major allele	Effect size	SE	P value	Minor allele frequency
chr5:74625512	rs7701925	a	g	0.0623	0.0058	7.69E-24	0.2533
chr5:74626082	rs11749783	c	t	0.0724	0.0037	4.42E-76	0.4103
chr5:74628014	rs3931914	c	g	0.0613	0.0044	1.70E-39	0.2546
chr5:74628375	rs3843481	t	a	0.0682	0.0038	7.52E-65	0.4103
chr5:74628497	rs2878418	a	c	0.0626	0.0061	2.94E-22	0.2533
chr5:74631139	rs3761738	t	c	0.0621	0.0087	5.02E-13	0.1029
chr5:74631262	rs17238330	g	a	0.0464	0.0080	2.54E-08	0.1029
chr5:74631352	rs13436857	c	t	0.0531	0.0056	3.27E-19	0.2691
chr5:74631501	rs3761739	t	c	0.0458	0.0051	7.16E-20	0.1570
chr5:74632133	rs3761740	a	c	0.0593	0.0088	2.57E-12	0.1029
chr5:74636484	rs10045497	a	c	0.0766	0.0053	2.32E-44	0.4077
chr5:74637711	rs10038095	t	a	0.0688	0.0055	4.23E-32	0.4103
chr5:74641560	rs10515198	a	g	0.0599	0.0061	5.99E-22	0.1029
chr5:74641707	rs2303152	a	g	0.0423	0.0064	1.04E-09	0.1201
chr5:74643670	rs17244848	c	t	0.0544	0.0084	1.07E-09	0.1069
chr5:74644681	rs10474434	t	g	0.0625	0.0061	4.11E-22	0.2520
chr5:74644706	rs6453131	g	t	0.0682	0.0052	4.46E-35	0.4090
chr5:74646878	rs11742194	t	c	0.0594	0.0060	8.13E-22	0.1029
chr5:74648496	rs17238484	t	g	0.0627	0.0062	1.35E-21	0.2533
chr5:74648603	rs12654264	t	a	0.0687	0.0037	3.89E-70	0.4103
chr5:74651084	rs3846662	g	a	0.0692	0.0038	2.24E-69	0.4525
chr5:74655726	rs3846663	t	c	0.0722	0.0037	1.13E-75	0.4103
chr5:74656175	rs5909	a	g	0.0617	0.0088	4.93E-13	0.1016
chr5:74656539	rs12916	c	t	0.0733	0.0038	7.79E-78	0.4314
chr5:74659099	rs78109779	g	a	0.0561	0.0085	5.56E-10	0.1042
chr5:74659587	rs11743717	t	c	0.0583	0.0085	1.37E-10	0.1042
chr5:74660356	rs4703670	c	t	0.0615	0.0044	4.54E-39	0.2375
chr5:74665860	rs3733907	c	g	0.0456	0.0051	7.72E-20	0.1557
chr5:74666492	rs7733436	t	c	0.0601	0.0055	6.16E-25	0.4090
chr5:74668589	rs17562686	t	c	0.0585	0.0088	1.09E-11	0.1003
chr5:74670454	rs10515197	t	c	0.0421	0.0054	6.84E-16	0.1544
chr5:74673707	rs13356670	g	a	0.0441	0.0044	8.34E-22	0.2678
chr5:74675076	rs7700468	c	t	0.0534	0.0042	2.62E-31	0.2573
chr5:74675717	rs16872526	g	t	0.0407	0.0068	2.36E-08	0.1029
chr5:74678485	rs16872536	t	g	0.0422	0.0068	9.54E-09	0.1029

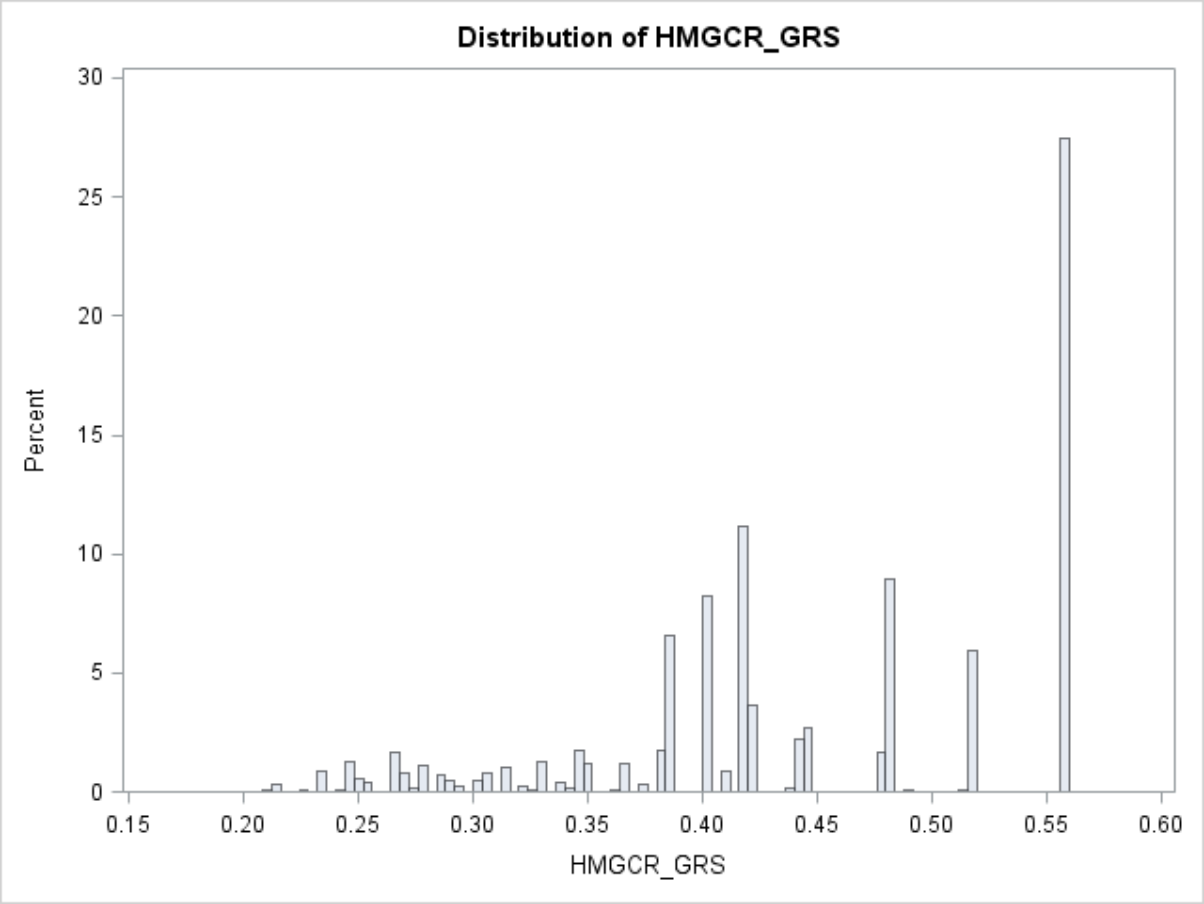
Chromosome and bp position	RSID	Minor allele	Major allele	Effect size	SE	P value	Minor allele frequency
chr5:74678937	rs3761742	a	g	0.0624	0.0044	1.36E-39	0.2375
chr5:74681500	rs3846665	a	c	0.0626	0.0044	1.01E-39	0.2361
chr5:74681773	rs698912	c	t	0.0623	0.0044	1.38E-39	0.2361
chr5:74683646	rs7356637	c	t	0.0532	0.0059	3.67E-16	0.2573
chr5:74685520	rs3761743	g	a	0.0536	0.0043	8.11E-32	0.2573
chr5:74687252	rs12234032	t	c	0.0595	0.0064	2.53E-18	0.2361
chr5:74689249	rs6878990	c	t	0.0645	0.0037	5.83E-63	0.4129
chr5:74692776	rs6453133	g	a	0.0497	0.0040	1.08E-32	0.3087
chr5:74696638	rs17648288	g	c	0.0569	0.0061	4.60E-20	0.1016
chr5:74696779	rs3804231	a	g	0.0642	0.0053	1.88E-29	0.1319
chr5:74697050	rs12173076	g	t	0.0654	0.0056	2.33E-27	0.1227
chr5:74701463	rs72633965	c	t	0.0648	0.0074	8.13E-17	0.1319
chr5:74704701	rs10515195	c	g	0.0446	0.0072	3.81E-09	0.1042
chr5:74704829	rs9293652	c	t	0.0642	0.0053	3.14E-29	0.1319
chr5:74706661	rs10067900	c	t	0.0616	0.0052	4.99E-28	0.1319
chr5:74710550	rs16872556	a	g	0.0414	0.0069	1.29E-08	0.1042
chr5:74717653	rs10038723	t	c	0.0622	0.0044	1.70E-39	0.2375
chr5:74717767	rs6453134	a	g	0.0504	0.0051	3.54E-20	0.2586
chr5:74717811	rs77385263	a	t	0.0564	0.0075	7.03E-13	0.1319
chr5:74724227	rs78979878	t	g	0.0657	0.0074	4.70E-17	0.1319
chr5:74724832	rs72633967	c	a	0.0660	0.0075	4.46E-17	0.1346
chr5:74725609	rs10075139	c	t	0.0617	0.0044	5.28E-39	0.2375
chr5:74731732	rs10434644	g	a	0.0541	0.0061	4.69E-17	0.2586
chr5:74732445	rs76478815	t	c	0.0651	0.0074	6.25E-17	0.1319
chr5:74732937	rs79058517	a	t	0.0531	0.0086	6.96E-09	0.0989
chr5:74734362	rs4385188	g	a	0.0533	0.0044	1.10E-29	0.2586
chr5:74735038	rs10474436	g	a	0.0539	0.0071	5.58E-13	0.2586
chr5:74736065	rs6872314	c	a	0.0529	0.0044	2.02E-29	0.2586
chr5:74740756	rs75673664	t	c	0.0651	0.0074	7.25E-17	0.1319
chr5:74745228	rs10070207	t	g	0.0493	0.0045	1.10E-24	0.2296
chr5:74748505	rs4704219	t	c	0.0647	0.0037	2.74E-63	0.4142
chr5:74749639	rs10041470	c	g	0.0488	0.0079	6.16E-09	0.1319
chr5:74753634	rs72633968	g	a	0.0643	0.0078	5.14E-15	0.1227
chr5:74754355	rs6874781	t	a	0.0501	0.0070	1.47E-11	0.2427

chr5:74756118	rs10076819	c	t	0.0648	0.0074	9.88E-17	0.1319
chr5:74757286	rs72633969	a	g	0.0650	0.0074	7.37E-17	0.1319
chr5:74757556	rs4704220	a	g	0.0639	0.0037	5.13E-62	0.4156
chr5:74757758	rs12659791	c	t	0.0433	0.0050	1.42E-18	0.1557
chr5:74658176	rs373596487	t	g	0.0555	0.0085	8.78E-10	NA
chr5:74659988	rs372811307	a	g	0.0602	0.0080	1.16E-12	NA
chr5:74641091	rs183357401	c	t	0.0625	0.0063	8.02E-21	1.0000
chr5:74700126	rs185707939	c	a	0.0639	0.0076	1.60E-15	1.0000
chr5:74700127	rs190616816	c	a	0.0638	0.0084	8.89E-13	1.0000
chr5:74734757	rs74703093	c	t	0.0633	0.0076	2.54E-15	1.0000
chr5:74562029	rs2006760	g	c	0.0533	0.0076	1.67E-13	1.0000
chr5:74644793	rs181748302	a	t	0.0685	0.0055	9.34E-32	NA

3.8 Appendix 3:8 | The results of the *HMGCR* SNPs clumping procedure

SNP	P-value (association with LDL-C in GLGC)	Number of other variants in clump	Other SNPs IDs in the clump
rs12916	7.79E-78	80	rs6866661 ,rs4704197 ,rs9942407 ,rs10066707 , rs2112653 ,rs10062361 ,rs72633962 ,rs4704200 , rs1551894 ,rs2052515 ,rs2126736 ,rs7722186 , rs918623 ,rs1423527 ,rs2335418 ,rs6871667 , rs10056811 ,rs11957260, rs6892023, rs17671591 , rs10474433 ,rs2878417 ,rs13358429 ,rs13354746 , rs6895057 ,rs3843480 ,rs7703051 ,rs7701925 , rs11749783 ,rs3931914 ,rs3843481 ,rs2878418 , rs13436857 ,rs3761739 ,rs10045497 ,rs10038095 , rs10474434 ,rs6453131 ,rs17238484 ,rs12654264 , rs3846662 ,rs3846663 ,rs4703670 ,rs3733907 , rs7733436 ,rs10515197 ,rs13356670 ,rs7700468 , rs3761742 ,rs3846665 ,rs698912 ,rs7356637 , rs3761743 ,rs12234032 ,rs6878990 ,rs6453133 , rs3804231 ,rs72633965 ,rs9293652 ,rs10067900 , rs10038723 ,rs6453134 ,rs77385263 ,rs78979878 , rs72633967 ,rs10075139 ,rs10434644 ,rs76478815, rs4385188 ,rs10474436 ,rs6872314 ,rs75673664 , rs10070207 ,rs4704219 ,rs10041470 ,rs6874781 , rs10076819 ,rs72633969 ,rs4704220 ,rs12659791
rs12173076	2.33E-27	1	rs72633968
rs10515198	5.99E-22	13	rs2006760 ,rs80037857 ,rs3761738 ,rs17238330 , rs3761740 ,rs17244848 ,rs11742194 ,rs5909 , rs78109779 ,rs11743717 ,rs17562686 ,rs17648288, rs79058517
rs185707939	1.60E-15	0	
rs74703093	2.54E-15	0	
rs3857388	2.20E-11	8	rs10064936 ,rs7711235 ,rs3895886 ,rs2303152, rs16872526 ,rs16872536 ,rs10515195 ,rs16872556

3.9 Appendix 3:9 | The distribution of *HMGCR* GRS



3.10 Appendix 3:10 | SNPs within 100 kb from the *NPC1L1* gene which are associated with LDL-C at GWAS level of significance

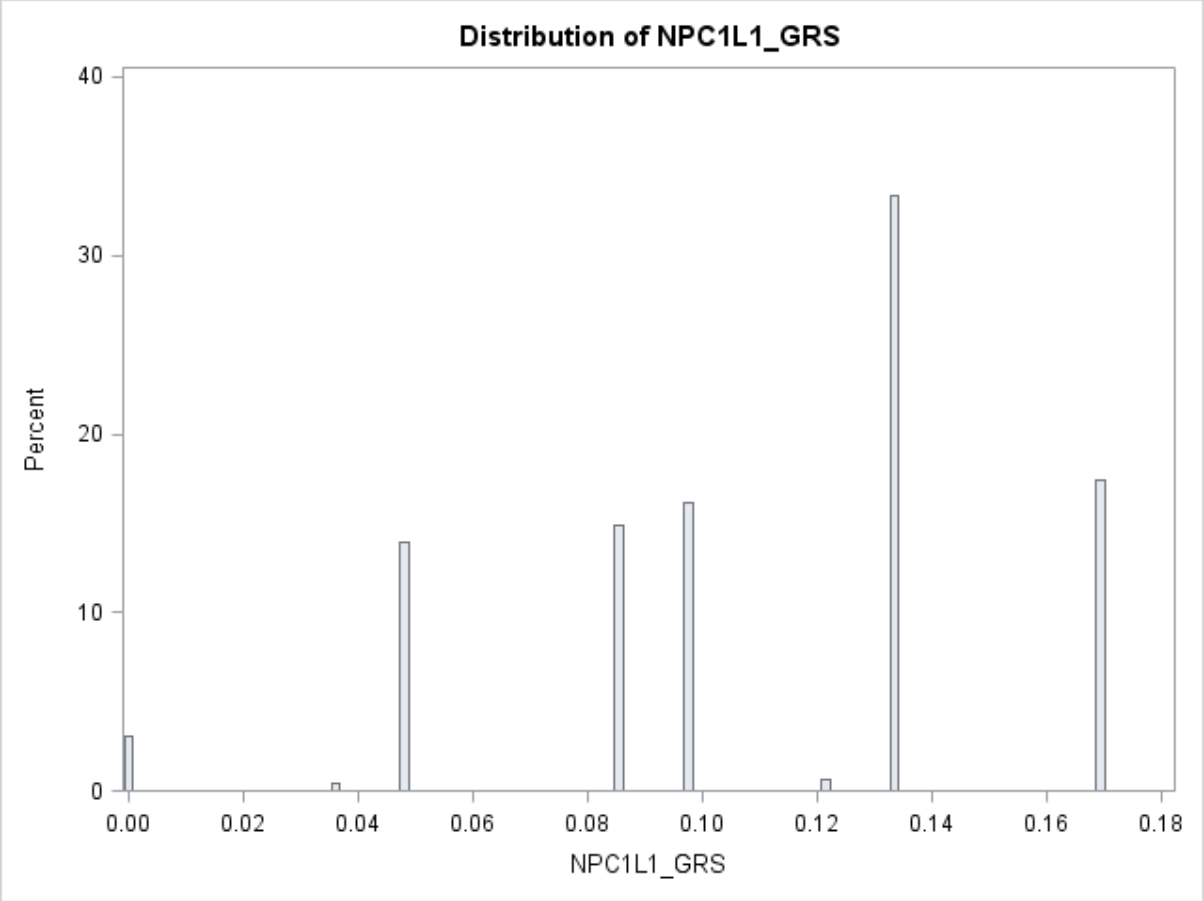
Chromosome and bp position	RSID	Effect allele*	Other allele	Effect size	SE	P value	Effect allele frequency
chr7:44570717	rs7798185	c	a	0.0409	0.0066	1.88E-09	0.84
chr7:44573761	rs217406	c	g	0.0388	0.0052	4.09E-14	0.84
chr7:44577622	rs73107473	g	a	0.0446	0.0068	2.69E-10	0.86
chr7:44579180	rs2072183	g	c	0.0386	0.0047	7.12E-16	0.76
chr7:44580876	rs41279633	g	t	0.0515	0.0069	2.27E-12	0.87
chr7:44580991	rs17655652	c	t	0.0280	0.0044	2.18E-10	0.29
chr7:44581986	rs17725246	t	c	0.0472	0.0049	1.49E-20	0.80
chr7:44582331	rs2073547	a	g	0.0485	0.0049	1.92E-21	0.81
chr7:44584551	rs10260606	g	c	0.0427	0.0049	3.33E-17	0.81
chr7:44600354	rs217387	a	c	0.0310	0.0045	4.45E-11	0.43
chr7:44600695	rs217386	a	g	0.0363	0.0038	1.20E-19	0.41
chr7:44602187	rs217385	t	g	0.0356	0.0038	4.40E-19	0.41
chr7:44602589	rs7791240	t	c	0.0425	0.0065	1.84E-10	0.91
chr7:44606217	rs217381	g	c	0.0342	0.0037	3.05E-18	0.41
chr7:44607172	rs217379	t	c	0.0330	0.0037	3.00E-17	0.43
chr7:44610490	rs217376	t	c	0.0330	0.0037	2.86E-17	0.43
chr7:44611077	rs217375	c	t	0.0331	0.0037	2.17E-17	0.43
chr7:44612454	rs217373	a	c	0.0329	0.0037	1.63E-17	0.43
chr7:44613103	rs217371	g	a	0.0329	0.0037	3.51E-17	0.43
chr7:44615186	rs217369	c	t	0.0325	0.0038	1.08E-15	0.43
chr7:44617793	rs7723	g	a	0.0380	0.0064	8.20E-09	0.91
chr7:44622286	rs217358	a	g	0.0329	0.0037	4.11E-17	0.42
chr7:44626377	rs217355	c	t	0.0294	0.0037	4.13E-14	0.43
chr7:44632212	rs10240746	a	g	0.0320	0.0052	2.76E-09	0.44
chr7:44646763	rs732797	g	a	0.0332	0.0052	1.09E-09	0.43
chr7:44657475	rs1434961	c	t	0.0263	0.0037	8.39E-12	0.48

*Effect allele is an allele with is associated with a decrease in LDL-C

3.11 Appendix 3:11| The results of the *NPC1L1* SNPs clumping procedure

SNP	P-value (association with LDL-C in GLGC)	Number of other variants in the clump	Other SNPs IDs in the clump
rs2073547	1.92E-21	9	rs7798185, rs217406, rs73107473, rs2072183, rs41279633, rs17725246, rs10260606, rs7791240, rs7723
rs217386	1.20E-19	15	rs17655652, rs217387, rs217385, rs217381, rs217379, rs217376, rs217375, rs217373, rs217371, rs217369, rs217358, rs217355, rs10240746, rs732797, rs1434961

3.12 Appendix 3:12 | The distribution of the *NPC1L1* GRS



Appendices Chapter 4

Appendix 4:1 The list of LDL-C lowering medications and the UK Biobank medication codes

Appendix 4:2 LDL-C and Apo B-100 levels in participants on-statin and not on-statin

4.1 Appendix 4:1 | The list of LDL-C lowering medications and the UK Biobank medication codes

UK Biobank medication code	UK Biobank medication name
1140861958	Simvastatin
1140881748	Zocor 10mg tablet
1141200040	Zocor heart-pro 10mg tablet
1141188146	Simvador 10mg tablet
1140888594	Fluvastatin
1140864592	Lescol 20mg capsule
1140888648	Pravastatin
1140861970	Lipostat 10mg tablet
1140910632	Eptastatin
1140910654	Velastatin
1141146234	Atorvastatin
1141146138	Lipitor 10mg tablet
1141192410	Rosuvastatin
1141192414	Crestor 10mg tablet
1140910652	Synvinolin

4.2 Appendix 4:2 | LDL-C and Apo B-100 levels in participants on-statin and not on-statin

Levels of lipoproteins in groups of UK Biobank participants by use of statin

Lipoprotein	N	Mean (95% CI) mmol/L	SD mmol/L	SE mmol/L
LDL-C				
On-statin uncorrected	55 534	2.83 (2.83, 84)	0.717	0.003
On-statin corrected	55 534	4.04 (4.04, 4.05)	1.024	0.004
Not on statin	261 197	3.74 (3.73, 3.74)	0.802	0.002
Overall uncorrected	316 731	3.58 (3.57, 3.58)	0.859	0.002
Overall corrected	316 731	3.79 (3.79, 379)	0.853	0.002
Apo B-100				
On-statin uncorrected	55 632	0.88 (0.87,0.88)	0.205	0.001
On-statin corrected	55 632	1.07 (0.07, 1.08)	0.251	0.001
Not on statin	261 727	1.07 (1.07, 1.08)	0.230	0.000
Overall uncorrected	317 359	1.04 (1.04, 1.04)	0.238	0.000
Overall corrected	317 359	1.07 (1.07, 1.07)	0.234	0.000

Appendices Chapter 5

Appendix 5:1	List of UKB-specific hypoglycaemic agents by classes
Appendix 5:2	Fixed-effect meta-analyses of the associations of *HMGCR* genetic variants with risk of T2DM in UK Biobank and the DIAGRAM consortium
Appendix 5:3	Fixed-effect meta-analyses of the associations of *HMGCR* genetic variants with HbA1c in UK Biobank and the MAGIC consortium
Appendix 5:4	The effects of the *HMGCR* SNPs on anthropometric measures in UK Biobank and the GIANT consortium
Appendix 5:5	Fixed-effect meta-analyses of the associations of *HMGCR* genetic variants with BMI in UK Biobank and the GIANT consortium
Appendix 5:6	Fixed-effect meta-analyses of the associations of *HMGCR* genetic variants with waist circumference in UK Biobank and the GIANT consortium
Appendix 5:7	Fixed-effect meta-analyses of the associations of *HMGCR* genetic variants with hip circumference in UK Biobank and the GIANT consortium
Appendix 5:8	Fixed-effect meta-analyses of the associations of *HMGCR* genetic variants with waist-to-hip ratio in UK Biobank and the GIANT consortium
Appendix 5:9	Fixed-effect meta-analyses of the associations of *HMGCR* genetic variants height in UK Biobank and the GIANT consortium

5.1 Appendix 5:1 | List of UKB-specific hypoglycaemic agents by classes

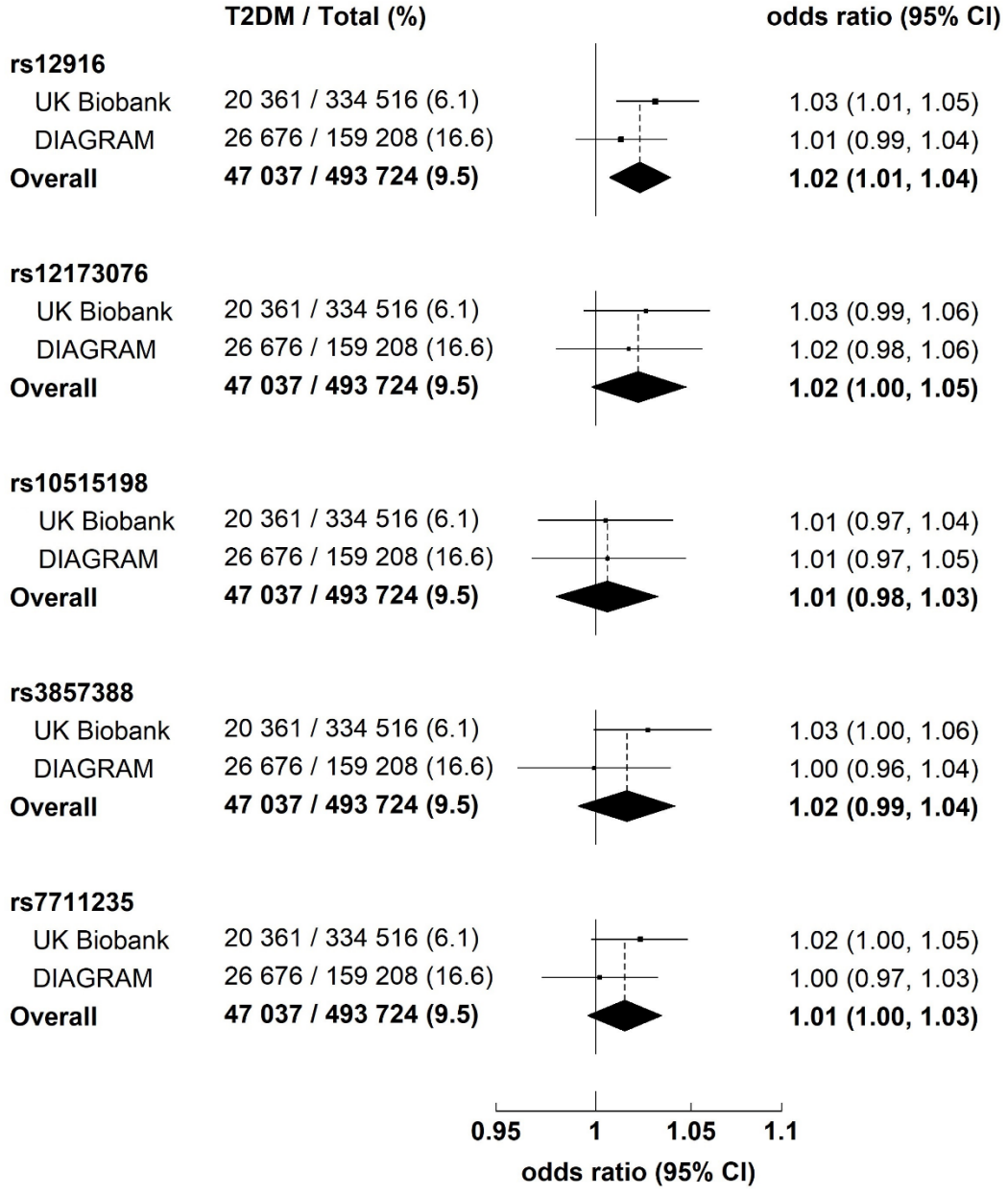
PHARMACOLOGICAL CLASS NAME	UKB MEDICATION CODE
Insulin	
insulin product	1140883066
Alpha-glucosidase inhibitors	
acarbose	1140868902
glucobay 50mg tablet	1140874646
Glitazone	
actos	1141171652
pioglitazone	1141171646
rosiglitazone	1141177600
rosiglitazone 1mg / metformin 500mg tablet	1141189090
Sulfonylurea	
diamicron 80mg tablet	1140874746
diabetamide 2.5mg tablet	1140874736
glibenclamide	1140874718
gliclazide	1140874744
glimepiride	1141152590
amaryl 1mg tablet	1141156984
minodiab 2.5mg tablet	1140874652
glipizide product	1141157284
semi-daonil 2.5mg tablet	1140874726
daonil 5mg tablet	1140874724
glibenese 5mg tablet	1140874650
glyclizide	1140910566
glipizide	1140874646
tolbutamide	1140874674
tolanase 100mg	1140874666

Metformin	
avandamet 1mg / 500mg tablet	1141189094
metformin	1140884600
glucophage 500mg tablet	1140874686

Meglitinide	
starlix 60mg tablet	1141173786
repaglinide	1141168660
nateglinide	1141173882
novonorm	1141168668

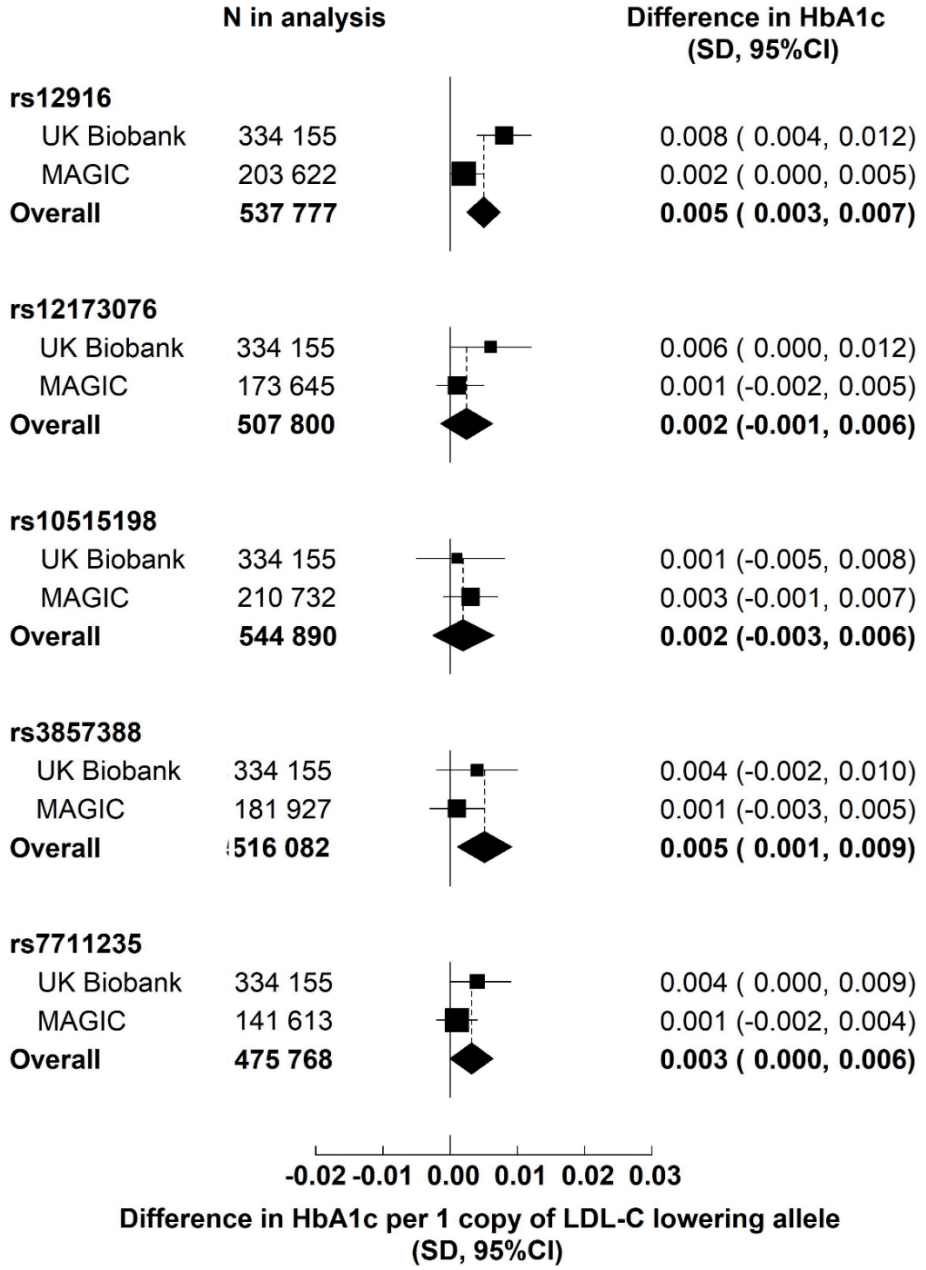
5.2 Appendix 5:2 | Fixed-effect meta-analyses of the associations of *HMGCR* genetic variants with risk of T2DM in UK Biobank and the DIAGRAM consortium

Associations of the genetic variants in *HMGCR* with T2DM in UK Biobank and the DIAGRAM consortium



The effect is shown in SD units per 1 copy of LDL-C lowering allele

5.3 Appendix 5:3 | Fixed-effect meta-analyses of the associations of *HMGCR* genetic variants with HbA1c in UK Biobank and the MAGIC consortium



The effect is shown in SD units per 1 copy of LDL-C lowering allele

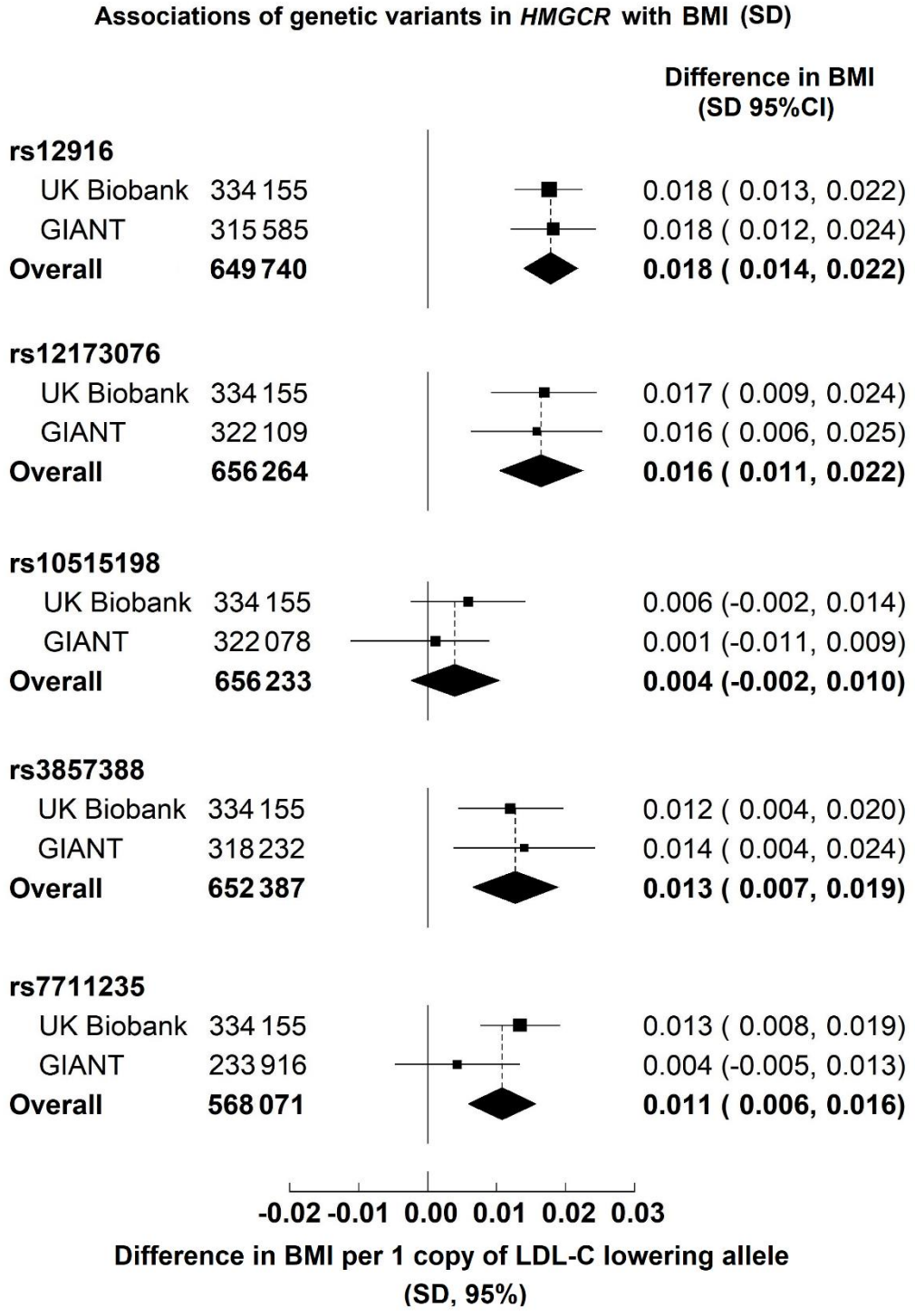
5.4 Appendix 5:4 | The effects of the *HMGCR* SNPs on anthropometric measures in UK Biobank and the GIANT consortium

	Effect allele/ other allele	GIANT difference in trait, SD (95%CI)	UK Biobank difference in trait, SD (95% CI)
rs12916	T/C	0.018 (0.012, 0.024)	0.018 (0.013, 0.022)
rs12173076	T/G	0.016 (0.006, 0.025)	0.017 (0.009, 0.024)
rs10515198	G/A	-0.001 (-0.011, 0.009)	0.006 (-0.002, 0.014)
rs3857388	T/C	0.014 (0.004, 0.024)	0.012 (0.005, 0.020)
rs7711235	A/G	0.004 (-0.005, 0.013)	0.013 (0.008, 0.019)
Waist circumference			
rs12916	T/C	0.017 (0.010, 0.024)	0.015 (0.010, 0.020)
rs12173076	T/G	0.017 (0.007, 0.027)	0.019 (0.011, 0.027)
rs10515198	G/A	-0.001 (-0.012, 0.01)	0.001 (-0.008, 0.009)
rs3857388	T/C	0.013 (0.002, 0.024)	0.013 (0.005, 0.021)
rs7711235	A/G	0.010 (0.000, 0.020)	0.013 (0.007, 0.018)
Hip circumference			
rs12916	T/C	0.015 (0.008, 0.022)	0.017 (0.012, 0.022)
rs12173076	T/G	0.017 (0.006, 0.028)	0.020 (0.012, 0.027)
rs10515198	G/A	0.002 (-0.009, 0.014)	0.007 (-0.002, 0.015)
rs3857388	T/C	0.007 (-0.005, 0.019)	0.009 (0.002, 0.017)
rs7711235	A/G	0.008 (-0.002, 0.019)	0.012 (0.007, 0.018)
Waist-to-hip ratio			
rs12916	T/C	0.009 (0.002, 0.016)	0.004 (-0.001, 0.009)
rs12173076	T/G	0.008 (-0.002, 0.018)	0.007 (0.000, 0.015)
rs10515198	G/A	-0.005 (-0.016, 0.006)	-0.005 (-0.013, 0.004)
rs3857388	T/C	0.011 (0.000, 0.022)	0.006 (-0.002, 0.013)
rs7711235	A/G	0.010 (0.000, 0.019)	0.007 (0.001, 0.012)
Height			

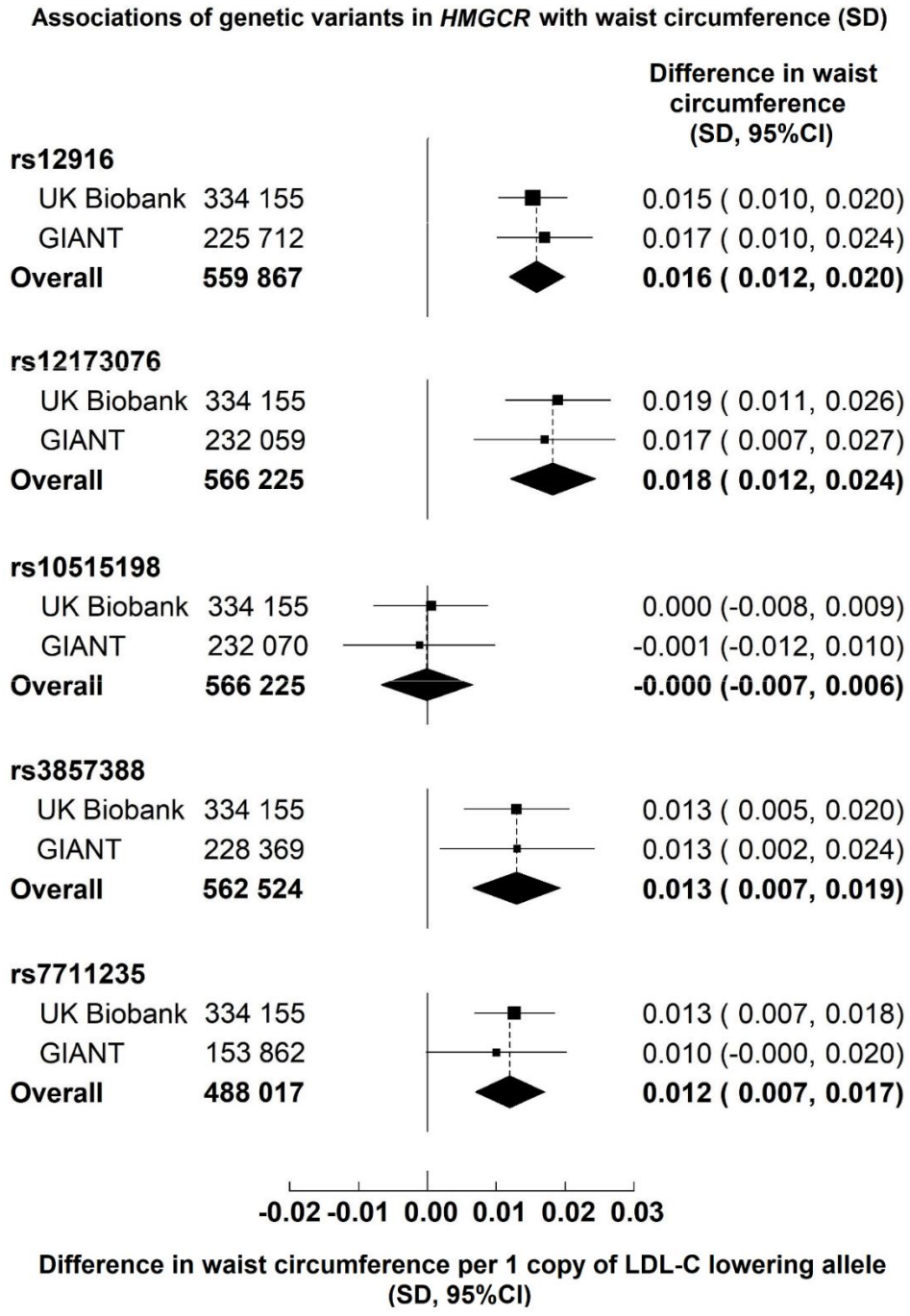
rs12916	T/C	-0.005 (-0.010, 0.001)	0.017 (0.012, 0.022)
rs12173076	T/G	0.006 (-0.003, 0.015)	0.024 (0.016, 0.032)
rs10515198	G/A	-0.008 (-0.017, 0.002)	-0.005 (-0.013, 0.004)
rs3857388	T/C	-0.001 (-0.010, 0.009)	0.009 (0.001, 0.016)
rs7711235	A/G	-0.005 (-0.012, 0.002)	0.006 (0.001, 0.012)

The effect is shown in SD (95% CI) units per 1 copy of effect allele.

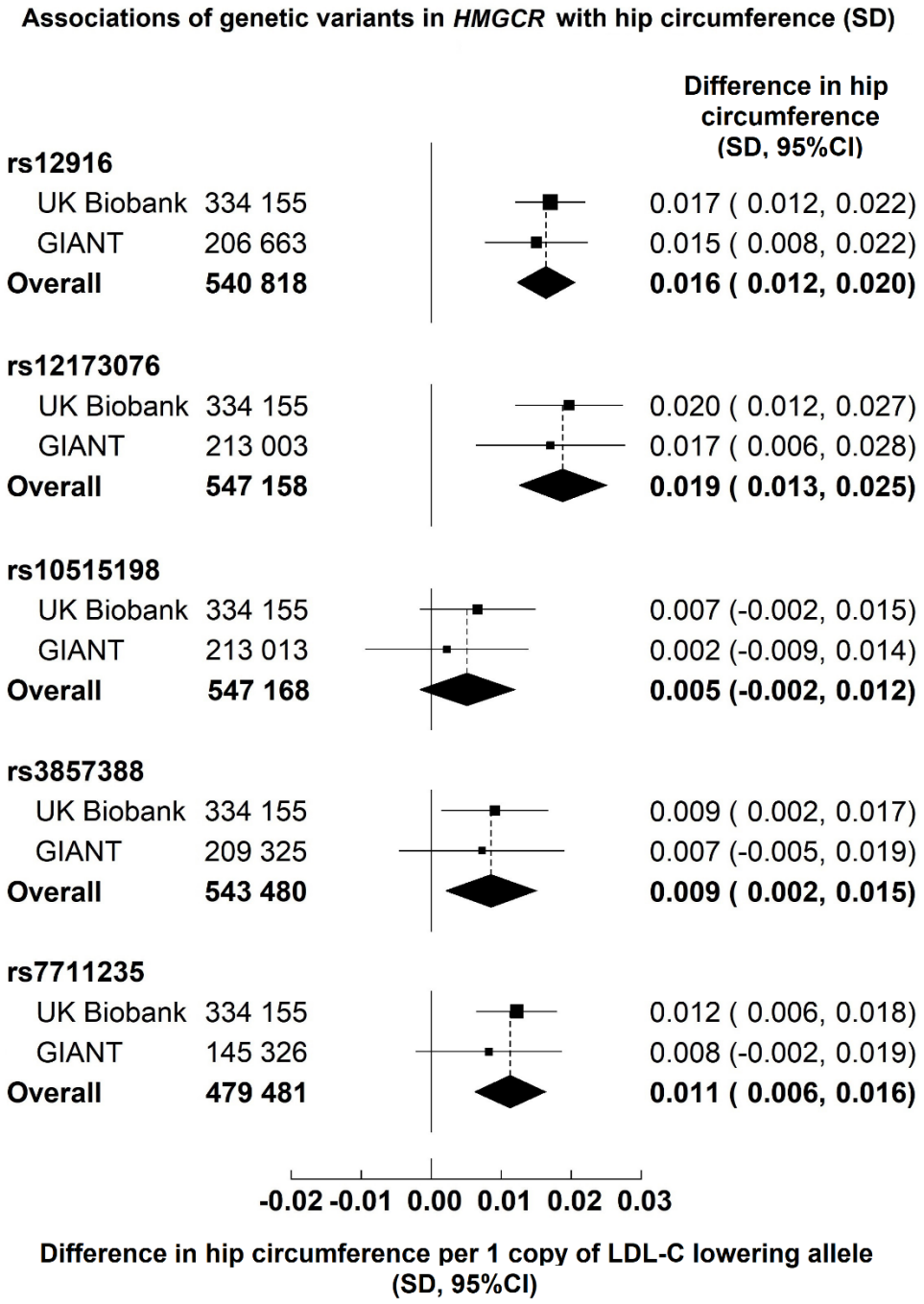
5.5 Appendix 5:5 | Fixed-effect meta-analyses of the associations of *HMGCR* genetic variants with BMI in UK Biobank and the GIANT consortium



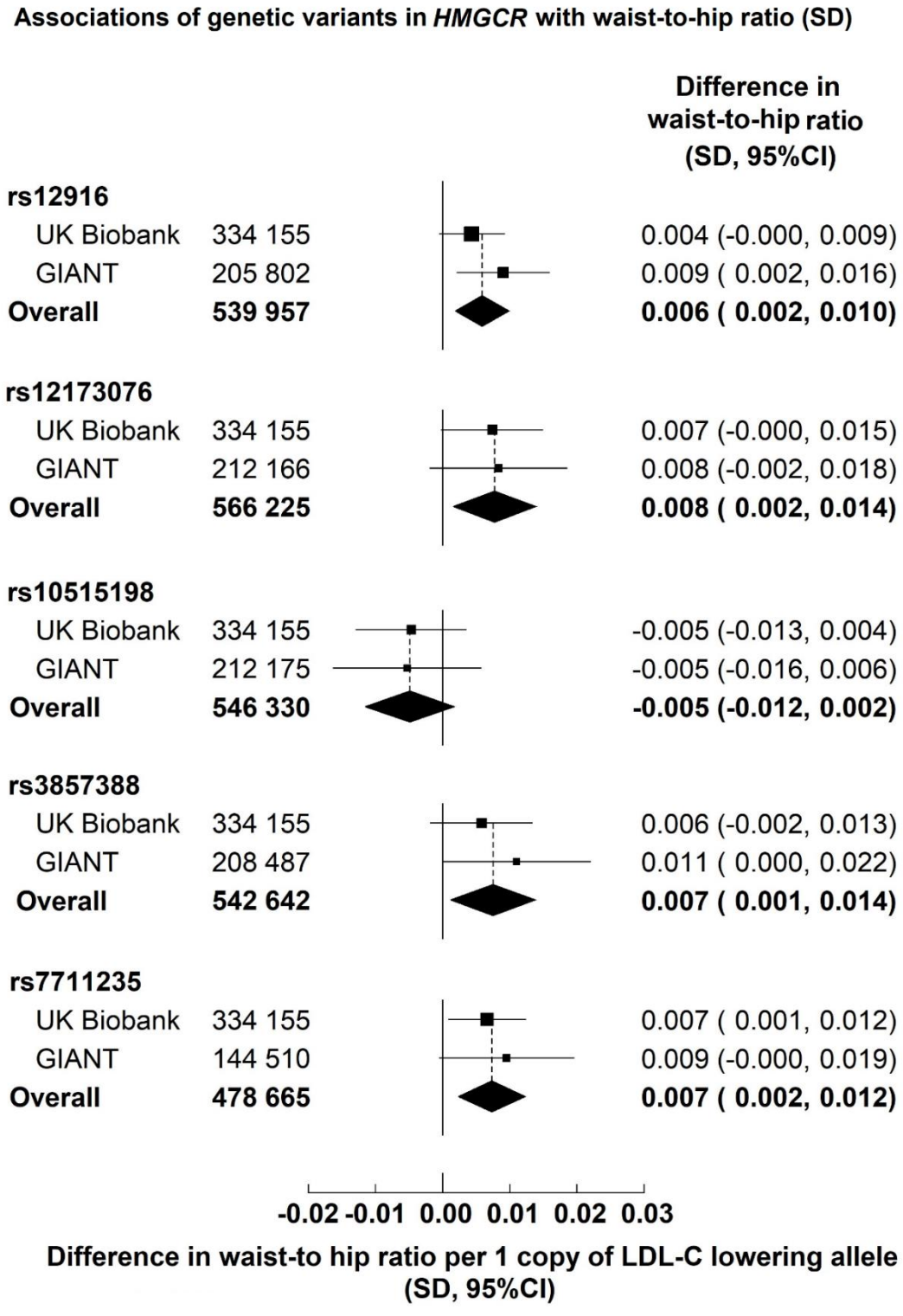
5.6 Appendix 5:6 | Fixed-effect meta-analyses of the associations of *HMGCR* genetic variants with waist circumference in UK Biobank and the GIANT consortium



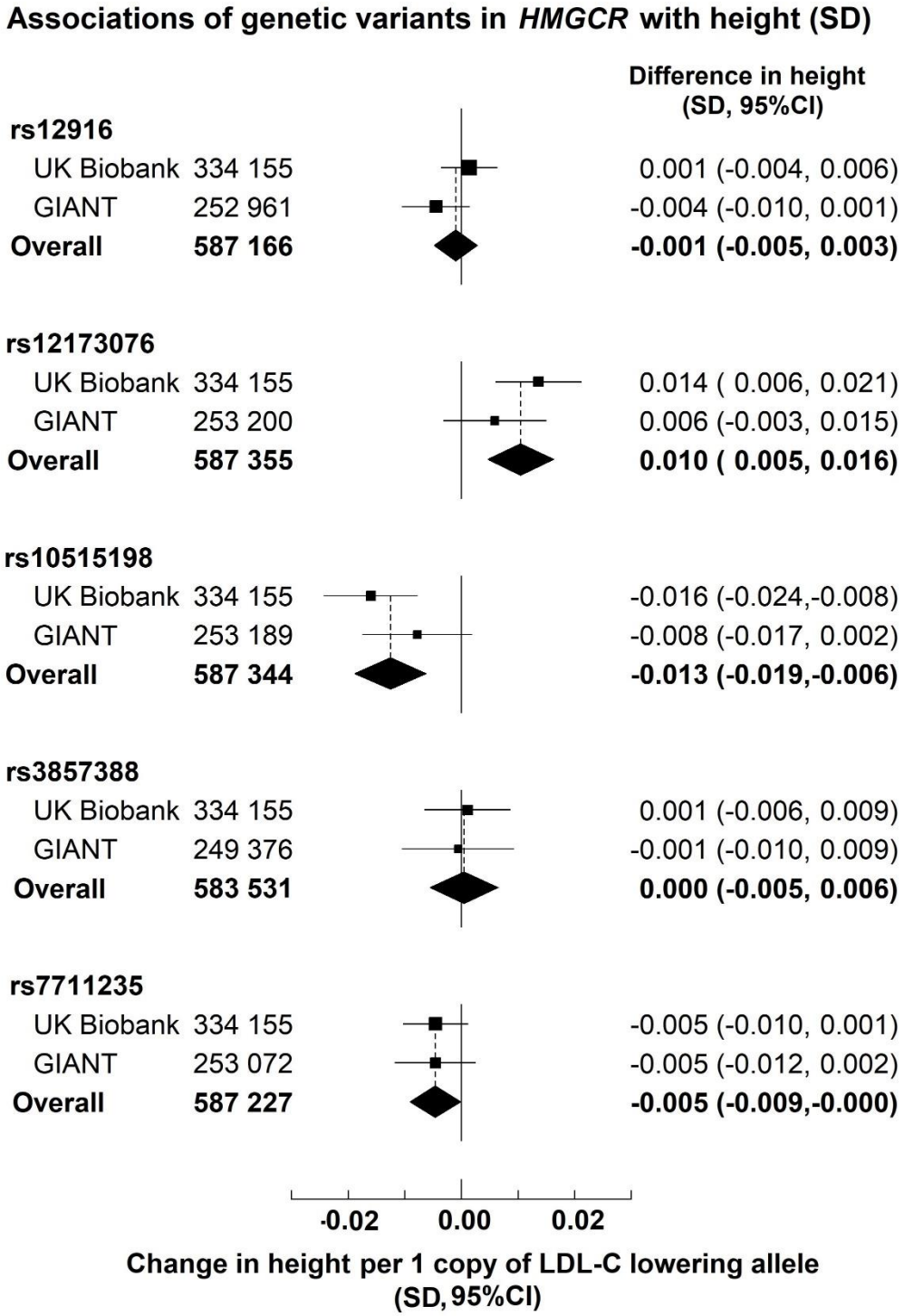
5.7 Appendix 5:7 | Fixed-effect meta-analyses of the associations of *HMGCR* genetic variants with hip circumference in UK Biobank and the GIANT consortium



5.8 Appendix 5:8 | Fixed-effect meta-analyses of the associations of *HMGCR* genetic variants with waist-to-hip ratio in UK Biobank and the GIANT consortium



5.9 Appendix 5:9 | Fixed-effect meta-analyses of the associations of *HMGCR* genetic variants with height in UK Biobank and the GIANT consortium



Appendices Chapter 6

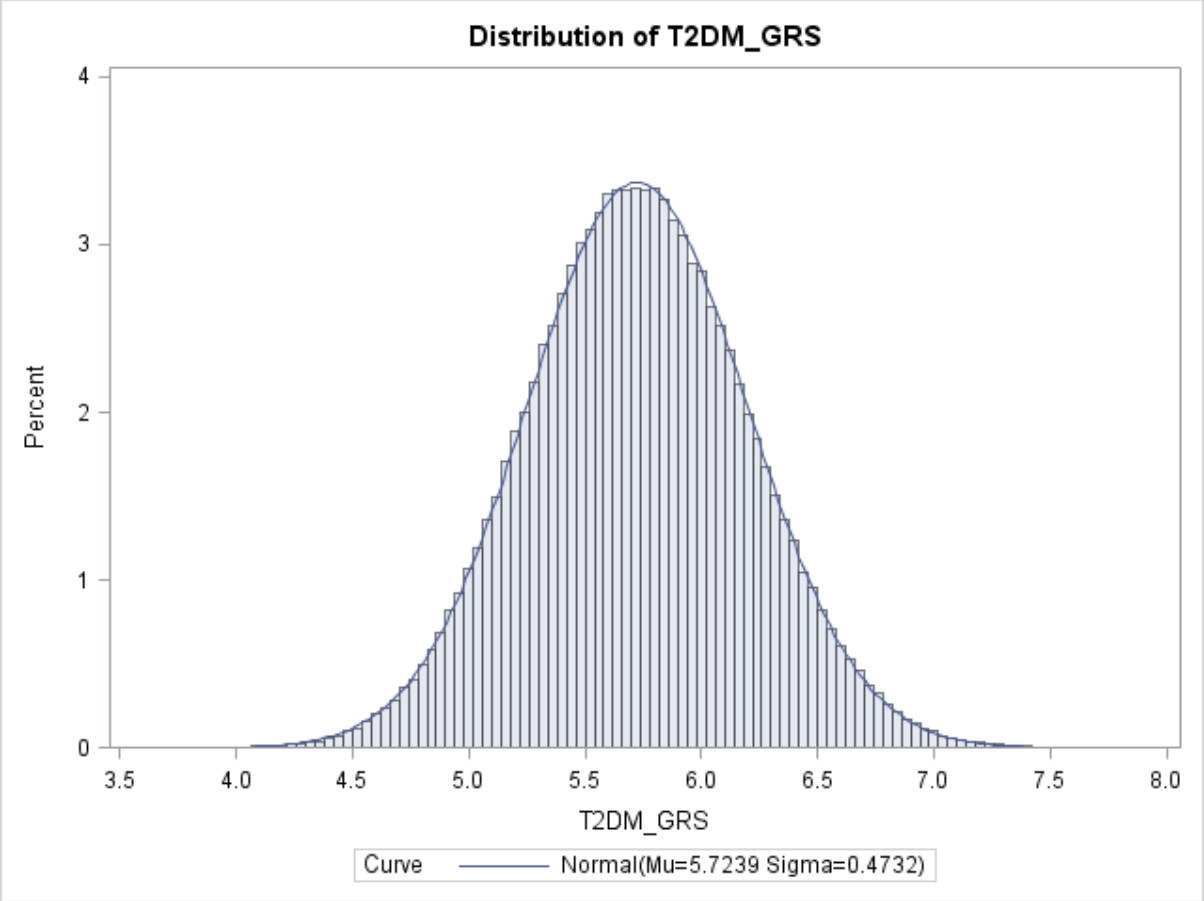
Appendix 6:1	57 SNPs associated with higher risk of T2DM, Morrison meta-analysis and UK Biobank genetic data
Appendix 6:2	The distribution of the polygenic T2DM GRS
Appendix 6:3	The distribution of QDiabetes-2018 RS
Appendix 6:4	Density Plot and Histogram of HbA1c distribution: a comparison of natural units with adjusted for *HMGCR* GRS residuals
Appendix 6:5	The correlation between HbA1c in mmol/mol and HbA1c residuals
Appendix 6:6	The associations of anthropometric measures (in SD units) with the *HMGCR* GRS in the subgroups stratified by sex and T2DM GRS

6.1 Appendix 6:1 | 57 SNPs associated with higher risk of T2DM, Morrison meta-analysis and UK Biobank genetic data

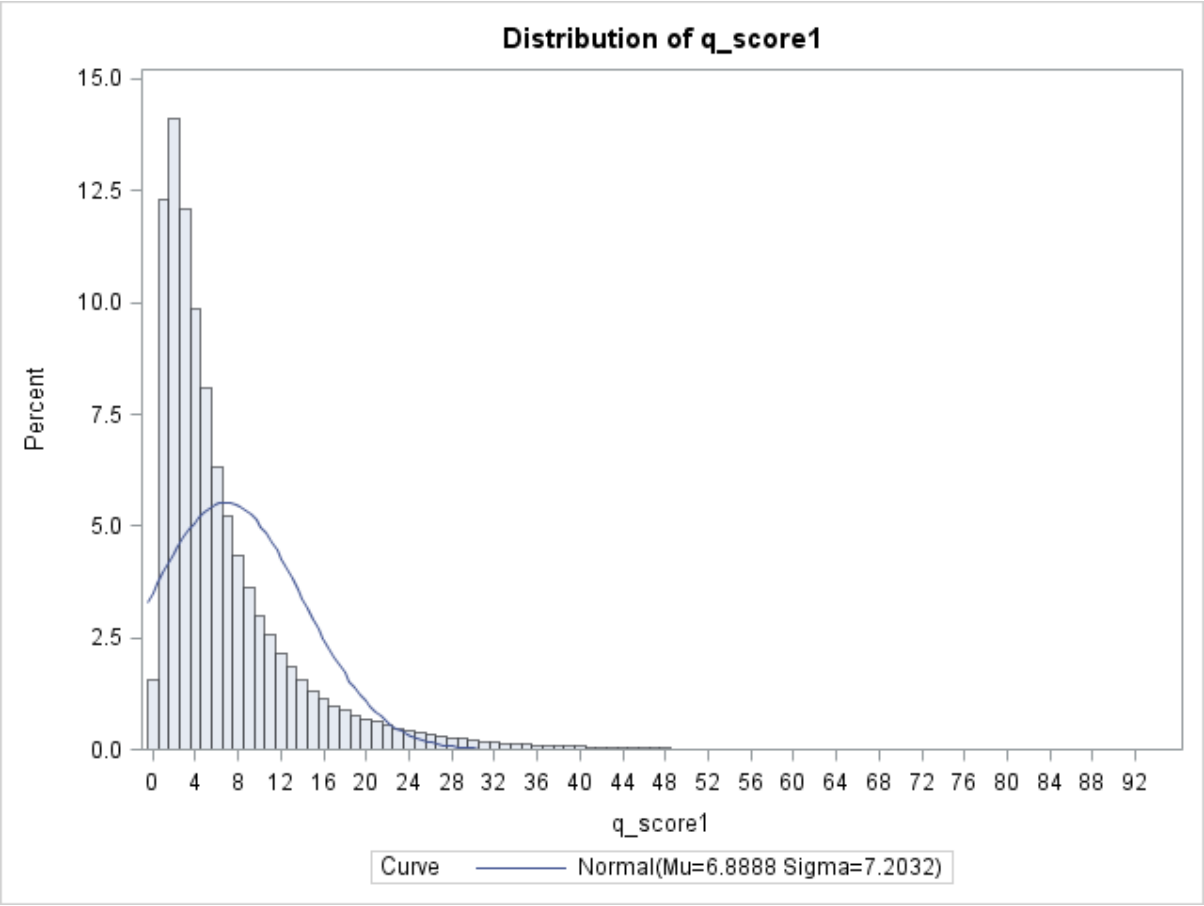
SNP	Chr	Locus	Morrison		UKB		SNPs in UKB for inversion	Effect allele freq	Effect on T2DM OR (95% CI)	P-value
			Effect allele	Other allele	Allele A	Allele B				
rs10203174	2	THADA	C	T	C	T	1	0.880975	1.14 (1.10-1.19)	9.50E-12
rs10278336	7	GCK	A	G	A	G	1	0.579009	1.07 (1.04-1.10)	6.38E-06
rs10401969	19	CILP2	C	T	T	C	0	0.0787202	1.13 (1.09-1.18)	7.04E-09
rs10758593	9	GLIS3	A	G	G	A	0	0.403204	1.06 (1.04-1.09)	2.58E-07
rs10811661	9	CDKN2A/B	T	C	T	C	1	0.828448	1.18 (1.15-1.22)	3.72E-27
rs10830963	11	MTNR1B	G	C	C	G	0	0.273823	1.10 (1.07-1.13)	5.32E-13
rs10842994	12	KLHDC5	C	T	C	T	1	0.808442	1.10 (1.06-1.13)	6.08E-10
rs10923931	1	NOTCH2	T	G	G	T	0	0.112705	1.08 (1.04-1.12)	0.000013
rs11063069	12	CCND2	G	A	A	G	0	0.205247	1.08 (1.05-1.11)	3.25E-07
rs1111875	10	HHEX/IDE	C	T	C	T	1	0.590354	1.11 (1.09-1.14)	1.98E-19
rs11257655	10	CDC123/CAMK1	T	C	C	T	0	0.211317	1.07 (1.04-1.10)	2.13E-06
rs11651052	17	HNF1B (TCF2)	A	G	A	G	1	0.476583	1.10 (1.07-1.14)	1.97E-11
rs11717195	3	ADCY5	T	C	T	C	1	0.758546	1.11 (1.08-1.14)	6.47E-14
rs11873305	18	MC4R	A	C	A	C	1	0.956324	1.18 (1.11-1.26)	3.77E-07
rs12427353	12	HNF1A (TCF1)	G	C	G	C	1	0.815228	1.08 (1.05-1.12)	6.52E-08
rs12571751	10	ZMIZ1	A	G	A	G	1	0.539947	1.08 (1.05-1.10)	1.02E-10
rs12899811	15	PRC1	G	A	A	G	0	0.319922	1.08 (1.05-1.10)	6.34E-09
rs12970134	18	MC4R	A	G	G	A	0	0.264332	1.08 (1.05-1.11)	1.19E-08
rs13389219	2	GRB14	C	T	C	T	1	0.605123	1.07 (1.05-1.10)	1.00E-08
rs1359790	13	SPRY2	G	A	G	A	1	0.718994	1.08 (1.05-1.10)	1.39E-08
rs1496653	3	UBE2E2	A	G	A	G	1	0.792267	1.09 (1.06-1.12)	3.56E-09
rs1552224	11	ARAP1 (CENTD)	A	C	A	C	1	0.849036	1.11 (1.07-1.14)	1.79E-10
rs163184	11	KCNQ1	G	T	T	G	0	0.470907	1.09 (1.06-1.11)	1.18E-11
rs17168486	7	DGKB	T	C	C	T	0	0.17641	1.11 (1.07-1.14)	5.94E-11
rs17791513	9	TLE4	A	G	A	G	1	0.937523	1.12 (1.07-1.17)	2.83E-07
rs1801282	3	PPARG	C	G	C	G	1	0.883008	1.13 (1.09-1.17)	1.05E-12
rs2075423	1	PROX1	G	T	G	T	1	0.665376	1.07 (1.05-1.10)	8.10E-09
rs2261181	12	HMGA2	T	C	C	T	0	0.101841	1.13 (1.08-1.17)	1.16E-09

SNP	Chr	Locus	Morrison		UKB		SNPs in UKB for inversion	Effect allele freq	Effect on T2DM OR (95% CI)	P-value
			Effect allele	Other allele	Allele A	Allele B				
rs231361	11	KCNQ1	A	G	G	A	0	0.259863	1.09 (1.06-1.12)	1.21E-09
rs243088	2	BCL11A	T	A	A	T	0	0.458342	1.07 (1.04-1.09)	1.81E-08
rs2796441	9	TLE1	G	A	G	A	1	0.58536	1.07 (1.05-1.10)	5.39E-09
rs2943640	2	IRS1	C	A	A	C	0	0.662478	1.10 (1.07-1.12)	2.69E-14
rs3802177	8	SLC30A8	G	A	G	A	1	0.695472	1.14 (1.11-1.17)	1.26E-21
rs4402960	3	IGF2BP2	T	G	G	T	0	0.320573	1.13 (1.10-1.16)	2.39E-23
rs4458523	4	WFS1	G	T	T	G	0	0.599837	1.10 (1.07-1.12)	2.02E-15
rs4502156	15	C2CD4A	T	C	T	C	1	0.555369	1.06 (1.03-1.08)	2.25E-06
rs459193	5	ANKRD55	G	A	A	G	0	0.739069	1.08 (1.05-1.11)	5.99E-09
rs4607517	7	GCK	A	G	G	A	0	0.174913	1.08 (1.04-1.11)	0.0000103
rs4760790	12	TSPAN8/LGR5	A	G	A	G	1	0.281058	1.10 (1.05-1.14)	7.99E-06
rs516946	8	ANK1	C	T	T	C	0	0.763448	1.09 (1.06-1.12)	2.49E-10
rs5215	11	KCNJ11	C	T	C	T	1	0.352101	1.07 (1.05-1.10)	8.50E-10
rs6795735	3	ADAMTS9	C	T	C	T	1	0.571294	1.08 (1.06-1.11)	7.39E-11
rs6878122	5	ZBED3	G	A	G	A	1	0.312165	1.10 (1.07-1.13)	5.04E-11
rs6960043	7	DGKB	C	T	T	C	0	0.517533	1.06 (1.04-1.09)	3.41E-07
rs7177055	15	HMG20A	A	G	G	A	0	0.703465	1.08 (1.05-1.10)	4.60E-09
rs7202877	16	BCAR1	T	G	T	G	1	0.904412	1.12 (1.07-1.16)	3.50E-08
rs7569522	2	RBMS1	A	G	A	G	1	0.457356	1.05 (1.03-1.07)	0.0000406
rs7593730	2	RBMS1	C	T	T	C	0	0.775134	1.11 (1.06-1.15)	1.50E-06
rs7756992	6	CDKAL1	G	A	A	G	0	0.273503	1.17 (1.14-1.20)	6.95E-35
rs780094	2	GCKR	C	T	T	C	0	0.621752	1.06 (1.04-1.09)	5.37E-07
rs7845219	8	TP53INP1	T	C	T	C	1	0.504465	1.06 (1.03-1.08)	4.57E-06
rs7903146	10	TCF7L2	T	C	C	T	0	0.289628	1.39 (1.35-1.42)	1.20E-139
rs7955901	12	TSPAN8/LGR5	C	T	C	T	1	0.440634	1.07 (1.05-1.10)	6.51E-09
rs8108269	19	GIPR	G	T	T	G	0	0.285574	1.07 (1.04-1.10)	4.36E-07
rs849135	7	JAZF1	G	A	G	A	1	0.512067	1.11 (1.08-1.13)	3.06E-17
rs944801	9	CDKN2A/B	C	G	G	C	0	0.588027	1.08 (1.05-1.10)	2.42E-09
rs9936385	16	FTO	C	T	T	C	0	0.393633	1.13 (1.10-1.16)	2.61E-23

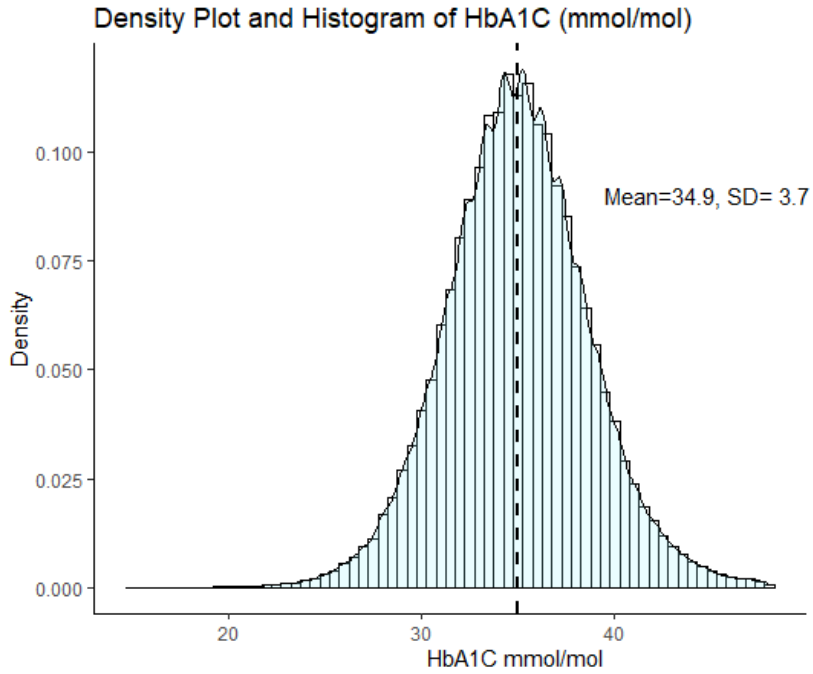
6.2 Appendix 6:2 | The distribution of the polygenic T2DM GRS

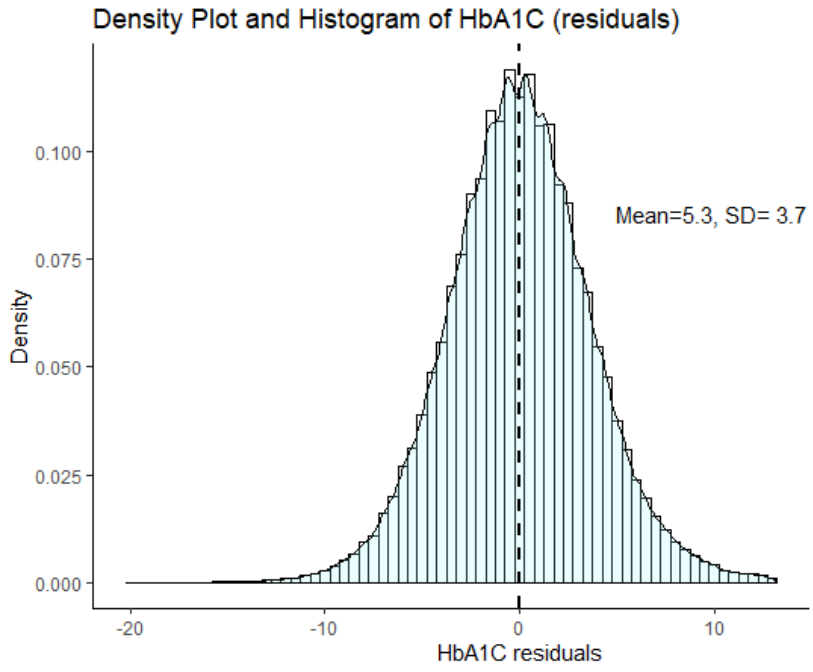


6.3 Appendix 6:3 | The distribution of QDiabetes-2018 RS

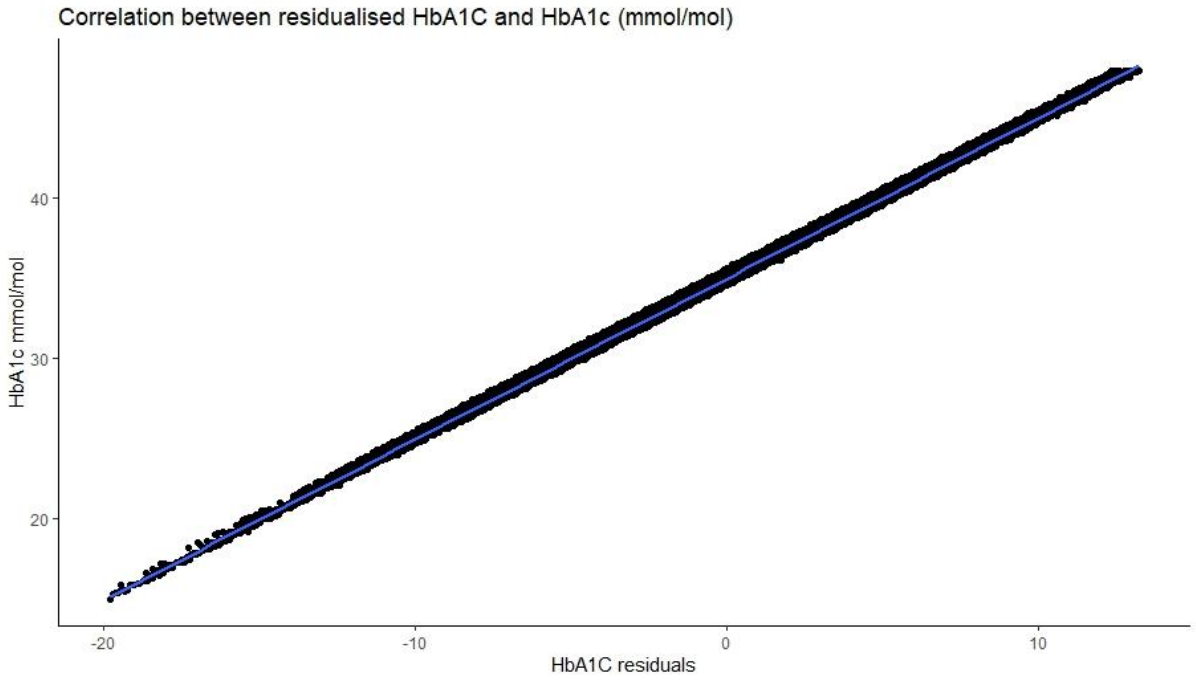


6.4 Appendix 6:4 | Density Plot and Histogram of HbA1c distribution: a comparison of natural units with adjusted for *HMGCR* GRS residuals





6.5 Appendix 6:5 | The correlation between HbA1c in mmol/mol and HbA1c residuals



6.6 Appendix 6:6 | The associations of anthropometric measures (in SD units) with the *HMGCR* GRS in the subgroups stratified by sex and T2DM GRS

	Men (N=154 796)	Women (N=179 720)	Overall (Average of stratified by sex values)	Overall (One model without sex stratification)	P-value for interaction with sex *	P-value for interaction with a risk group**
Body Weight (kg)						
Low risk	0.19 (0.11, 0.27)	0.20 (0.12, 0.27)	0.19 (0.12, 0.27)	0.19 (0.14, 0.25)		
Medium risk	0.04 (-0.04, 0.12)	0.09 (0.01, 0.16)	0.06 (-0.01, 0.14)	0.07 (0.01, 0.12)		
High risk	0.09 (0.01, 0.17)	0.14 (0.07, 0.21)	0.11 (0.04, 0.19)	0.12 (0.06, 0.17)		
Overall	0.11 (0.06, 0.15)	0.14 (0.10, 0.18)	0.12 (0.08, 0.17)	0.13 (0.09, 0.16)	0.2888	0.0105
BMI (kg/m2)						
Low risk	0.17 (0.09, 0.25)	0.25 (0.16, 0.33)	0.21 (0.12, 0.29)	0.21 (0.15, 0.27)		
Medium risk	0.06 (-0.02, 0.14)	0.10 (0.01, 0.19)	0.08 (0.00, 0.16)	0.08 (0.02, 0.14)		
High risk	0.10 (0.02, 0.18)	0.18 (0.09, 0.27)	0.14 (0.06, 0.22)	0.14 (0.08, 0.20)		
Overall	0.11 (0.06, 0.15)	0.18 (0.13, 0.23)	0.14 (0.09, 0.19)	0.15 (0.11, 0.18)	0.062	0.0077
Body Fat Mass (kg)						
Low risk	0.16 (0.09, 0.24)	0.21 (0.13, 0.30)	0.19 (0.11, 0.27)	0.19 (0.13, 0.25)		
Medium risk	0.03 (-0.05, 0.11)	0.09 (0.01, 0.18)	0.06 (-0.02, 0.14)	0.07 (0.01, 0.12)		
High risk	0.10 (0.02, 0.18)	0.15 (0.07, 0.24)	0.13 (0.04, 0.21)	0.13 (0.07, 0.19)		
Overall	0.10 (0.05, 0.14)	0.15 (0.10, 0.20)	0.13 (0.08, 0.17)	0.13 (0.10, 0.16)	0.1317	0.0128
Waist circumference (cm)						
Low risk	0.16 (0.09, 0.24)	0.16 (0.08, 0.23)	0.16 (0.08, 0.24)	0.16 (0.11, 0.21)		
Medium risk	0.05 (0.09, 0.24)	0.08 (0.01, 0.16)	0.07 (-0.01, 0.14)	0.07 (0.01, 0.12)		

High risk	0.07 (0.00, 0.15)	0.13 (0.05, 0.21)	0.10 (0.03, 0.18)	0.10 (0.05,0.16)		
Overall	0.10 (0.05, 0.14)	0.12 (0.08, 0.17)	0.11 (0.07, 0.15)	0.11 (0.08, 0.14)	0.3502	0.0706
Hip circumference (cm)						
Low risk	0.15 (0.08, 0.22)	0.22 (0.13, 0.31)	0.18 (0.11, 0.26)	0.19 (0.13, 0.24)		
Medium risk	0.09 (0.02, 0.16)	0.12 (0.03, 0.21)	0.11 (0.03, 0.19)	0.11 (0.05, 0.16)		
High risk	0.08 (0.01, 0.15)	0.14 (0.05, 0.23)	0.11 (0.03, 0.19)	0.11 (0.05, 0.17)		
Overall	0.11 (0.07, 0.15)	0.16 (0.11, 0.21)	0.13 (0.09, 0.18)	0.14 (0.10, 0.17)	0.1207	0.0896

* The model with fitted interaction term (TRAIT ~ HMGCR_GRS SEX **HMGCR_GRS*SEX** PC0-PC40 ARRAY)

** The model with fitted interaction term (TRAIT ~ HMGCR_GRS T2DM_RISK_GROUP **HMGCR_GRS*T2DM_RISK_GROUP** SEX PC0-PC40 ARRAY

T2DM_RISK_GROUP is a categorical variable with 3 levels derived from the polygenic T2DM risk score

Appendices Chapter 7

Appendix 7:1 Fixed-effect meta-analyses of the associations of *NPC1L1* genetic variants with risk of T2DM in UK Biobank and the DIAGRAM consortium

Appendix 7:2 Fixed-effect meta-analyses of the associations of *NPC1L1* genetic variants with HbA1c in UK Biobank and the MAGIC consortium

Appendix 7:3 The effects of the *NPC1L1* SNPs on anthropometric measures in UK Biobank and the GIANT consortium.

Appendix 7:4 Fixed-effect meta-analyses of the associations of *NPC1L1* genetic variants with BMI in UK Biobank and the GIANT consortium

Appendix 7:5 Fixed-effect meta-analyses of the associations of *NPC1L1* genetic variants with waist circumference in UK Biobank and the GIANT consortium

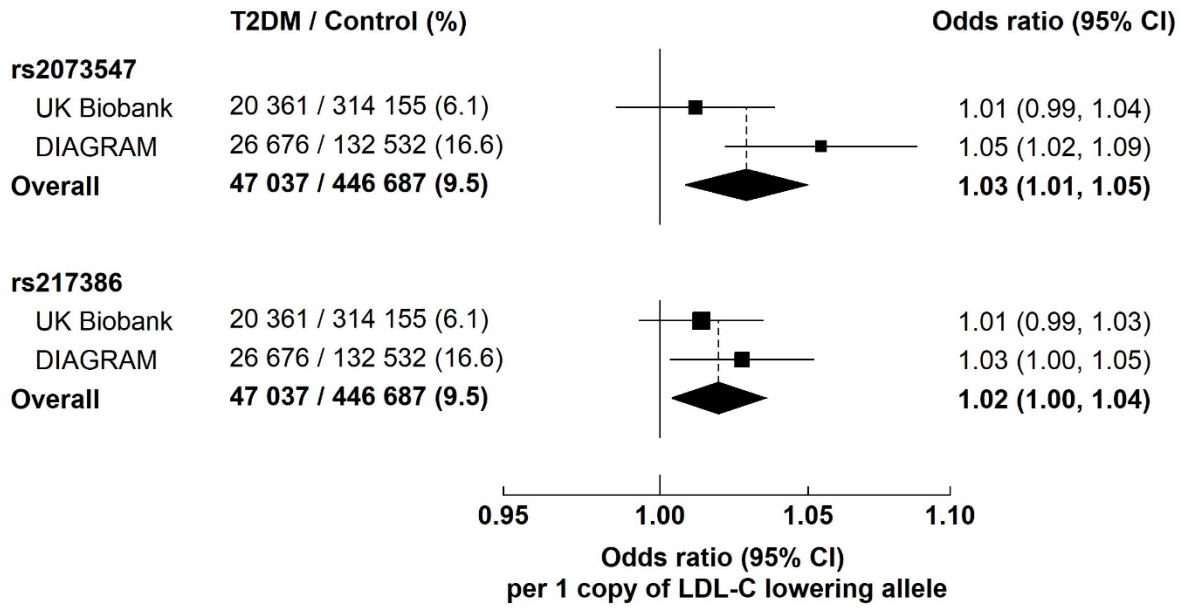
Appendix 7:6 Fixed-effect meta-analyses of the associations of *NPC1L1* genetic variants with hip circumference in UK Biobank and the GIANT consortium

Appendix 7:7 Fixed-effect meta-analyses of the associations of *NPC1L1* genetic variants with waist-to-hip ratio in UK Biobank and the GIANT consortium

Appendix 7:8 Fixed-effect meta-analyses of the associations of *NPC1L1* genetic variants with height in UK Biobank and the GIANT consortium

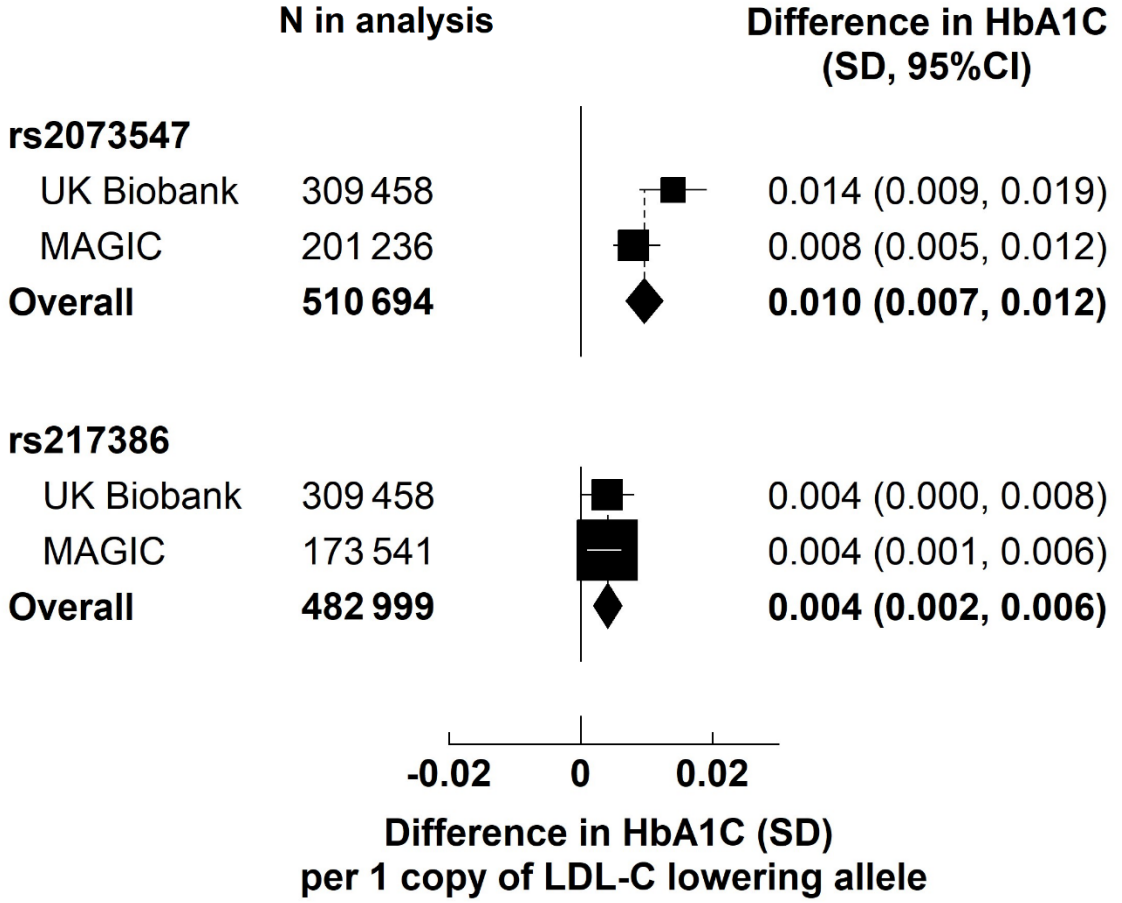
7.1 Appendix 7:1 | Fixed-effect meta-analyses of the associations of *NPC1L1* genetic variants with risk of T2DM in UK Biobank and the DIAGRAM consortium

Associations of the genetic variants in *NPC1L1* with T2DM in UK Biobank and the DIAGRAM consortium



7.2 Appendix 7:2 | Fixed-effect meta-analyses of the associations of *NPC1L1* genetic variants with HbA1c in UK Biobank and the MAGIC consortium

Associations of the genetic variants in *NPC1L1* with HbA1c in UK Biobank and the DIAGRAM consortium

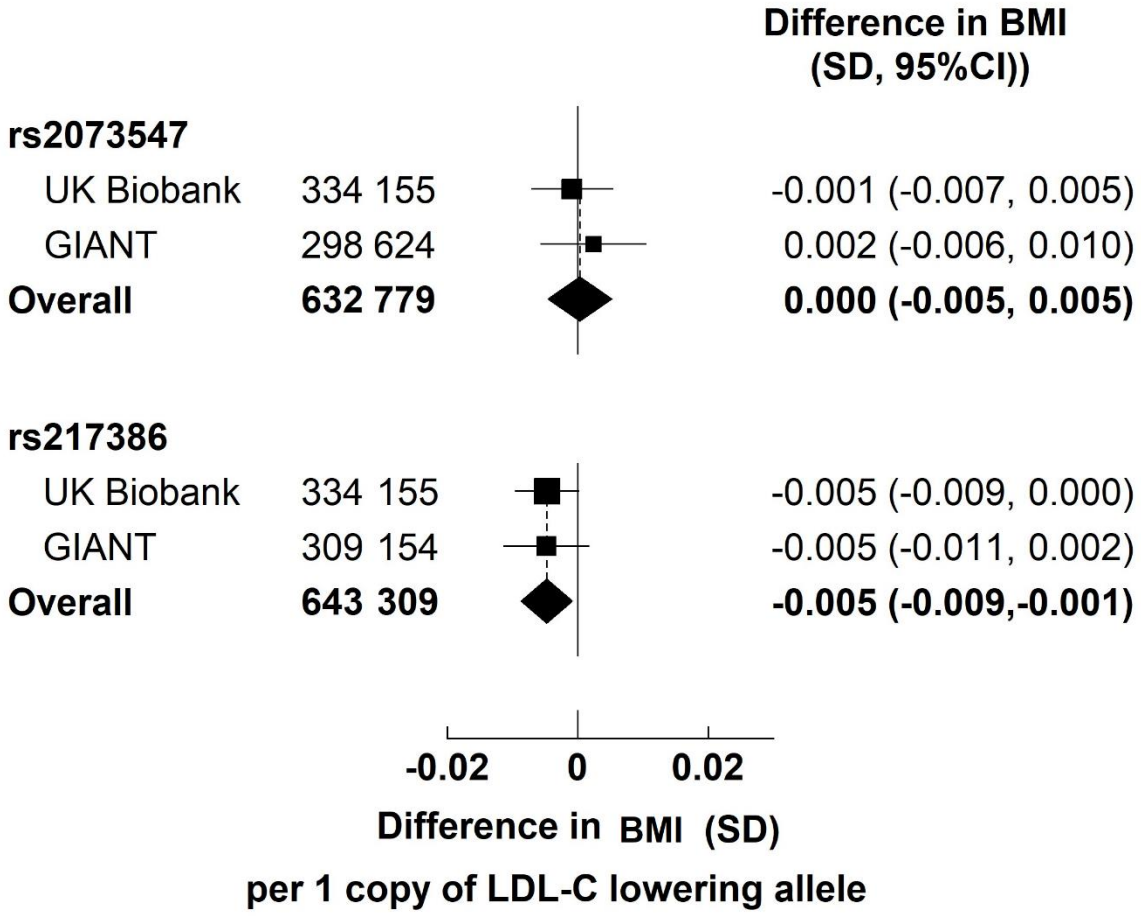


7.3 Appendix 7:3 | The effects of the *NPC1L1* SNPs on anthropometric measures in UK Biobank and the GIANT consortium.

RSID	Effect allele / other allele	UK Biobank difference in trait, SD (95% CI)	GIANT difference in trait, SD (95% CI)
BMI			
rs2073547	A/G	-0.001 (-0.007, 0.005)	0.009 (0.004, 0.014)
rs217386	A/G	-0.005 (-0.010, 0.000)	0.003 (-0.001, 0.007)
Waist circumference			
rs2073547	A/G	-0.001 (-0.007, 0.006)	0.002 (-0.007, 0.011)
rs217386	A/G	-0.008 (-0.013, -0.003)	0.003 (-0.004, 0.010)
Hip circumference			
rs2073547	A/G	-0.003 (-0/009, 0.003)	0.003 (-0.006, 0.013)
rs217386	A/G	-0.007 (-0.012, -0.002)	0.003 (-0.004, 0.011)
Waist-to-hip ratio			
rs2073547	A/G	0.002 (-0.004, 0.008)	0.001 (-0.008, 0.010)
rs217386	A/G	-0.004 (-0.008, 0.001)	0.002 (-0.005, 0.009)
Height			
rs2073547	A/G	0.004 (-0.002, 0.010)	-0.000 (-0.009, 0.008)
rs217386	A/G	-0.003 (-0.008, 0.002)	0.003 (-0.004, 0.009)

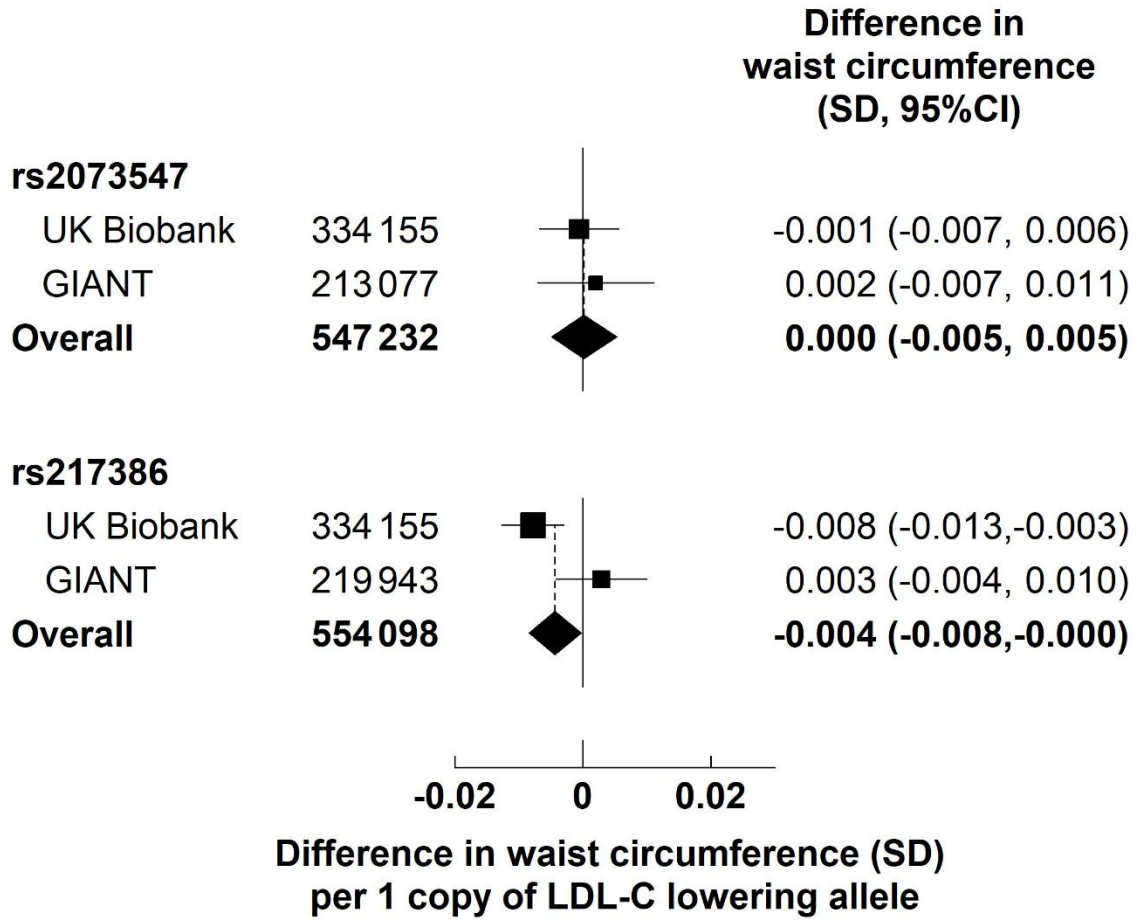
7.4 Appendix 7:4 | Fixed-effect meta-analyses of the associations of *NPC1L1* genetic variants with BMI in UK Biobank and the GIANT consortium

Associations of the genetic variants in *NPC1L1* with BMI (SD) in UK Biobank and the GIANT consortium



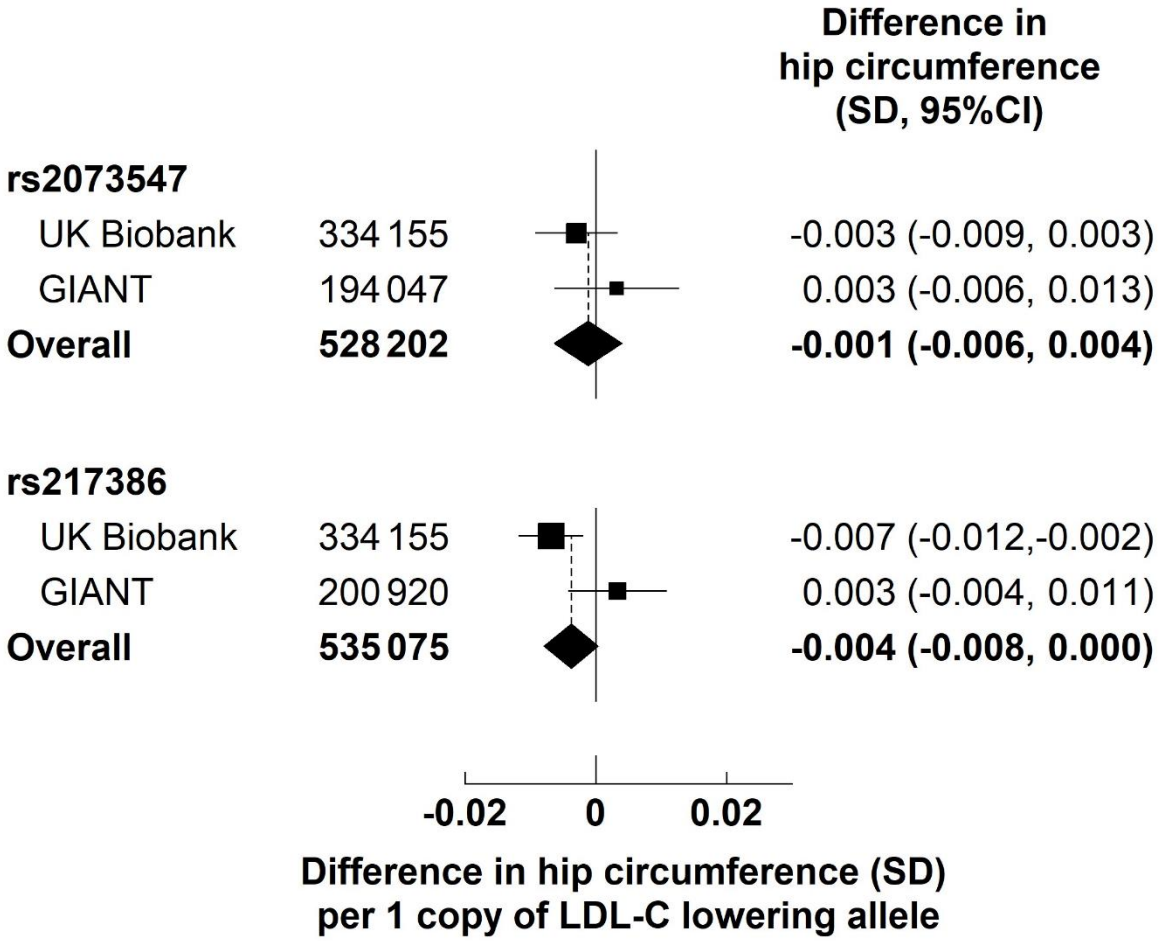
7.5 Appendix 7:5 | Fixed-effect meta-analyses of the associations of *NPC1L1* genetic variants with waist circumference in UK Biobank and the GIANT consortium

Associations of the genetic variants in *NPC1L1* with waist circumference (SD) in UK Biobank and the GIANT consortium



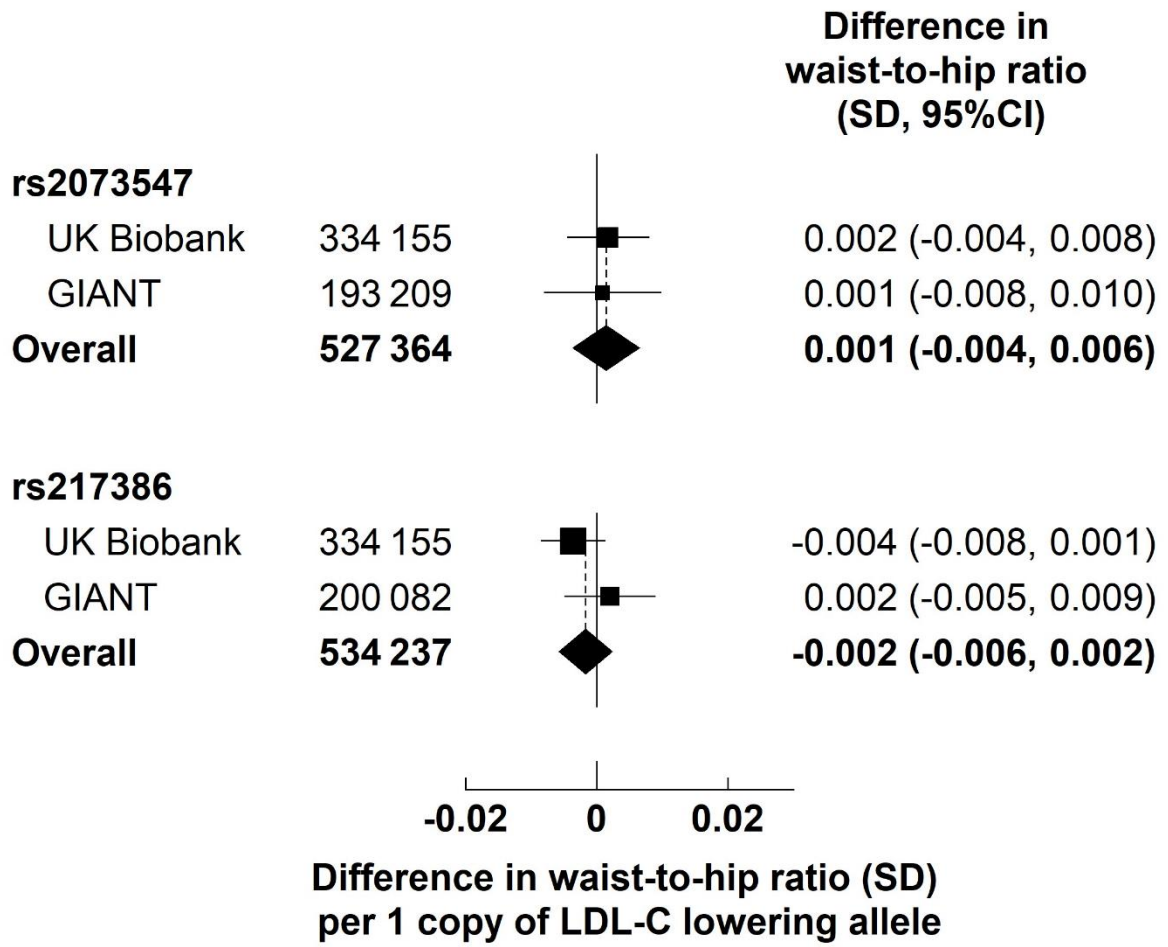
7.6 . Appendix 7:6 | Fixed-effect meta-analyses of the associations of *NPC1L1* genetic variants with hip circumference in UK Biobank and the GIANT consortium

Associations of the genetic variants in *NPC1L1* with hip circumference (SD) in UK Biobank and the GIANT consortium



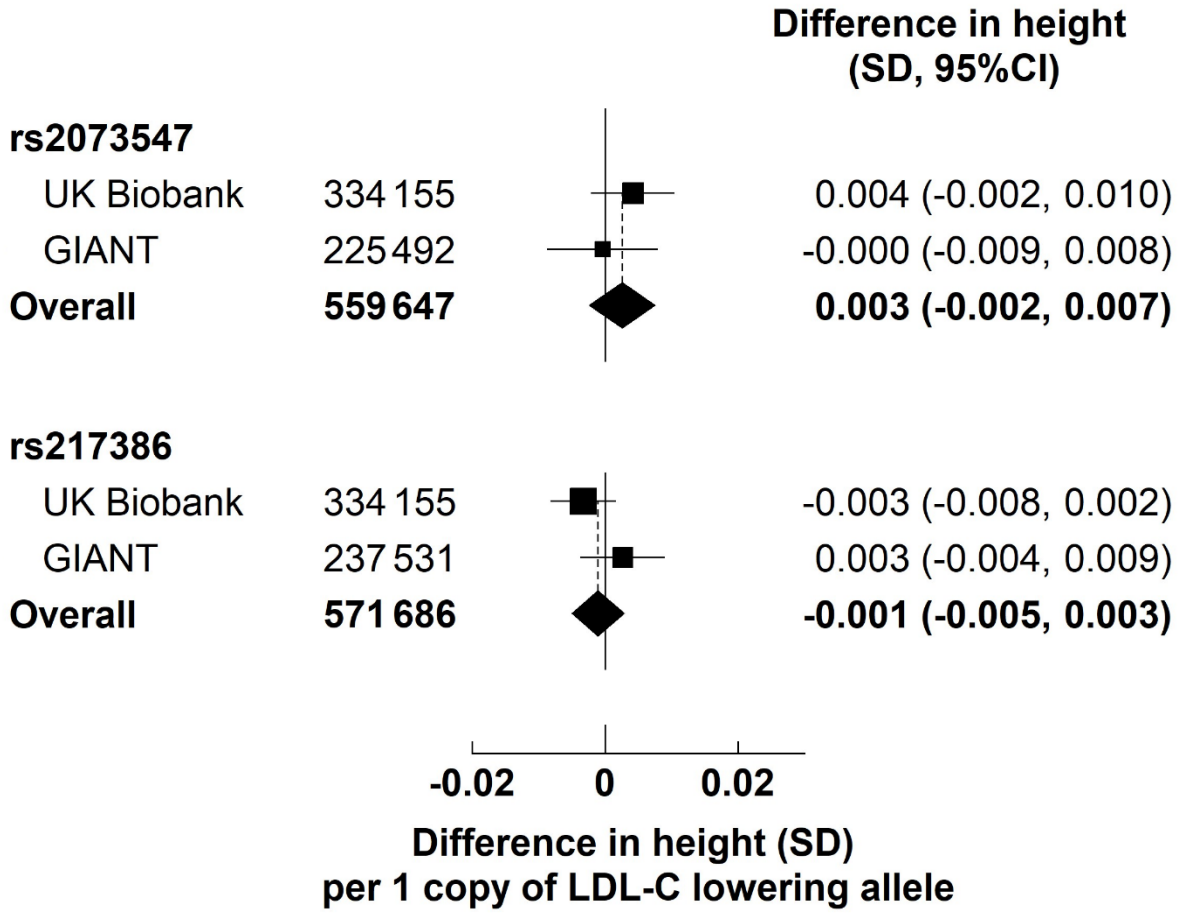
7.7 Appendix 7:7 | Fixed-effect meta-analyses of the associations of *NPC1L1* genetic variants with waist-to-hip ratio in UK Biobank and the GIANT consortium

Associations of the genetic variants in *NPC1L1* with waist-to-hip ratio (SD) in UK Biobank and the GIANT consortium



7.8 Appendix 7:8 | Fixed-effect meta-analyses of the associations of *NPC1L1* genetic variants with height in UK Biobank and the GIANT consortium

Associations of the genetic variants in *NPC1L1* with height (SD) in UK Biobank and the GIANT consortium



References:

1. GBD 2015 Mortality and Causes of Death Collaborators. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet.* 2016;388(10053):1459-1544.
2. Roth GA, Mensah GA, Johnson CO, et al. Global Burden of Cardiovascular Diseases and Risk Factors, 1990-2019: Update From the GBD 2019 Study. *J Am Coll Cardiol.* 2020;76(25):2982-3021.
3. Leong DP, Joseph PG, McKee M, et al. Reducing the Global Burden of Cardiovascular Disease, Part 2: Prevention and Treatment of Cardiovascular Disease. *Circ Res.* 2017;121(6):695-710.
4. Mensah GA, Wei GS, Sorlie PD, et al. Decline in Cardiovascular Mortality: Possible Causes and Implications. *Circ Res.* 2017;120(2):366-380.
5. Andersson C, Johnson AD, Benjamin EJ, Levy D, Vasan RS. 70-year legacy of the Framingham Heart Study. *Nat Rev Cardiol.* 2019;16(11):687-698.
6. Ward NC, Watts GF, Eckel RH. Statin Toxicity. *Circ Res.* 2019;124(2):328-350.
7. Wang H, Naghavi M, Allen C, et al. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet.* 2016;388(10053):1459-1544.
8. Collins R, Reith C, Emberson J, et al. Interpretation of the evidence for the efficacy and safety of statin therapy. *Lancet.* 2016;388(10059):2532-2561.
9. Cholesterol Treatment Trialists C, Baigent C, Blackwell L, et al. Efficacy and safety of more intensive lowering of LDL cholesterol: a meta-analysis of data from 170,000 participants in 26 randomised trials. *Lancet.* 2010;376(9753):1670-1681.
10. Curtis HJ, Walker AJ, MacKenna B, Croker R, Goldacre B. Prescription of suboptimal statin treatment regimens: a retrospective cohort study of trends and variation in English primary care. *Br J Gen Pract.* 2020;70(697):e525-e533.
11. British Heart Foundation. UK Factsheet. https://www.bhf.org.uk/-/media/files/research/heart-statistics/bhf-cvd-statistics-uk-factsheet.pdf?la=en. Published 2020. Accessed.
12. OpenPrescribing. The DataLab. University of Oxford. OpenPrescribing.net. Accessed 27/03/22, 2022.
13. National Institute for Health and Care Excellence. NICE clinical guideline 181. Lipid modification: cardiovascular risk assessment and the modification of blood lipids for the primary and secondary prevention of cardiovascular disease. https://www.nice.org.uk/guidance/cg181. Published 2016. Accessed.
14. SIGN. *Risk estimation and the prevention of cardiovascular disease.* 2017.
15. Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. *BMJ.* 2008;336(7659):1475-1482.
16. Tunstall-Pedoe H, Woodward M. ASSIGN score. Estimate the risk. ASSIGN score Web site. www.assign-score.com/estimate-the-risk/. Published 2008. Accessed.
17. Woodward M, Brindle P, Tunstall-Pedoe H, estimation Sgor. Adding social deprivation and family history to cardiovascular risk assessment: the ASSIGN score from the Scottish Heart Health Extended Cohort (SHHEC). *Heart.* 2007;93(2):172-176.
18. Newman CB, Preiss D, Tobert JA, et al. Statin Safety and Associated Adverse Events: A Scientific Statement From the American Heart Association. *Arterioscler Thromb Vasc Biol.* 2019;39(2):e38-e81.

19. Sattar N, Preiss D, Murray HM, et al. Statins and risk of incident diabetes: a collaborative meta-analysis of randomised statin trials. *The Lancet.* 2010;375(9716):735-742.
20. Preiss D, Seshasai SR, Welsh P, et al. Risk of Incident Diabetes With Intensive-Dose Compared With Moderate-Dose Statin Therapy A Meta-analysis. *JAMA.* 2011;305(24):2556–2564.
21. Carmena R, Betteridge DJ. Diabetogenic Action of Statins: Mechanisms. *Curr Atheroscler Rep.* 2019;21(6):23.
22. Davis HR, Zhu LJ, Hoos LM, et al. Niemann-Pick C1 like 1 (NPC1L1) is the intestinal phytosterol and cholesterol transporter and a key modulator of whole-body cholesterol homeostasis. *J Biol Chem.* 2004;279(32):33586-33592.
23. Cannon CP, Blazing MA, Giugliano RP, et al. Ezetimibe Added to Statin Therapy after Acute Coronary Syndromes. *N Engl J Med.* 2015;372(25):2387-2397.
24. Agnelli G, Avezum A, Brieger D, et al. Rationale and design of the GRACE (Global Registry of Acute Coronary Events) Project: A multinational registry of patients hospitalized with acute coronary syndromes. *Am Heart J.* 2001;141(2):190-199.
25. Qamar A, Alexander K, M. Incidence of New-Onset Diabetes in the IMPROVE-IT Trial: Does Adding Ezetimibe to Simvastatin Increase Risk Compared to Simvastatin Alone? . Latest in cardiology Web site. https://www.acc.org/Latest-in-Cardiology/Articles/2015/11/20/14/20/Incidence-of-New-Onset-Diabetes-in-the-IMPROVE-IT-Trial. Published 2015. Accessed 05.09.2022, 2022.
26. Blazing MA, Giugliano RP, de Lemos JA, et al. On-treatment analysis of the Improved Reduction of Outcomes: Vytorin Efficacy International Trial (IMPROVE-IT). *Am Heart J.* 2016;182:89-96.
27. Reed J, Bain S, Kanamarlapudi V. A Review of Current Trends with Type 2 Diabetes Epidemiology, Aetiology, Pathogenesis, Treatments and Future Perspectives. *Diabetes Metab Syndr Obes.* 2021;14:3567-3602.
28. Forouhi NG, Wareham NJ. Epidemiology of diabetes. *Medicine (Abingdon).* 2014;42(12):698-702.
29. Malone JI, Hansen BC. Does obesity cause type 2 diabetes mellitus (T2DM)? Or is it the opposite? *Pediatric Diabetes.* 2019;20(1):5-9.
30. Lotta LA, Sharp SJ, Burgess S, et al. Association Between Low-Density Lipoprotein Cholesterol–Lowering Genetic Variants and Risk of Type 2 Diabetes. *Jama.* 2016;316(13).
31. Chawla A, Chawla R, Jaggi S. Microvasular and macrovascular complications in diabetes mellitus: Distinct or continuum? *Indian J Endocrinol Metab.* 2016;20(4):546-551.
32. Julian GS, Accetturi Pititto LM, Miresashvili N, Broe Honore J, Lopes Assis Coelho RC, Campos D. Cost of microvascular complications in people with diabetes from a public healthcare perspective: a retrospective database study in Brazil. *J Med Econ.* 2021;24(1):1002-1010.
33. Hex N, Bartlett C, Wright D, Taylor M, Varley D. Estimating the current and future costs of Type 1 and Type 2 diabetes in the UK, including direct health costs and indirect societal and productivity costs. *Diabet Med.* 2012;29(7):855-862.
34. Laakso M. Epidemiology of Type 2 Diabetes. In: Goldstein JB, Muller-Wieland D, eds. *Type 2 Diabetes. Principles and Practice.* Sscond ed.: CRC Press; 2007.
35. Deshpande AD, Harris-Hayes M, Schootman M. Epidemiology of diabetes and diabetes-related complications. *Phys Ther.* 2008;88(11):1254-1264.
36. Khan MAB, Hashim MJ, King JK, Govender RD, Mustafa H, Al Kaabi J. Epidemiology of Type 2 Diabetes - Global Burden of Disease and Forecasted Trends. *J Epidemiol Glob Health.* 2020;10(1):107-111.
37. Ling W, Huang Y, Huang YM, Fan RR, Sui Y, Zhao HL. Global trend of diabetes mortality attributed to vascular complications, 2000-2016. *Cardiovasc Diabetol.* 2020;19(1):182.
38. Rao Kondapally Seshasai S, Kaptoge S, Thompson A, et al. Diabetes mellitus, fasting glucose, and risk of cause-specific death. *N Engl J Med.* 2011;364(9):829-841.

39. Chávez-Reyes J, Escárcega-González CE, Chavira-Suárez E, et al. Susceptibility for Some Infectious Diseases in Patients With Diabetes: The Key Role of Glycemia. *Front Public Health.* 2021;9:559595.
40. International Diabetes Federation. IDF Diabetes Atlas. www.diabetesatlas,org. Published 2021. Accessed 04.11.2021, 2021.
41. Reay WR, Cairns MJ. Advancing the use of genome-wide association studies for drug repurposing. *Nature Reviews Genetics.* 2021;22(10):658-671.
42. The Global Diabetes Community. Cost of Diabetes. www.diabetes.co.uk/cost-of-diabetes.html. Published 2021. Accessed 04.11.2021, 2019.
43. Tinajero MG, Malik VS. An Update on the Epidemiology of Type 2 Diabetes: A Global Perspective. *Endocrinol Metab Clin North Am.* 2021;50(3):337-355.
44. Shu L, Blencowe M, Yang X. Translating GWAS Findings to Novel Therapeutic Targets for Coronary Artery Disease. *Front Cardiovasc Med.* 2018;5:56.
45. Willer CJ, Schmidt EM, Sengupta S, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet.* 2013;45(11):1274-1283.
46. Nikpay M, Goel A, Won H-H, et al. A comprehensive 1000 Genomes–based genome-wide association meta-analysis of coronary artery disease. *Nature Genetics.* 2015;47(10):1121.
47. Morris AP, Voight BF, Teslovich TM, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet.* 2012;44(9):981-990.
48. Teslovich TM, Musunuru K, Smith AV, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature.* 2010;466(7307):707-713.
49. Sonehara K, Okada Y. Genomics-driven drug discovery based on disease-susceptibility genes. *Inflammation and Regeneration.* 2021;41(1):8.
50. Ference BA, Ray KK, Catapano AL, et al. Mendelian Randomization Study of ACLY and Cardiovascular Disease. *N Engl J Med.* 2019;380(11):1033-1042.
51. Ference BA, Robinson JG, Brook RD, et al. Variation in PCSK9 and HMGCR and Risk of Cardiovascular Disease and Diabetes. *N Engl J Med.* 2016;375(22):2144-2153.
52. Nabirotchkin S, Peluffo AE, Rinaudo P, Yu J, Hajj R, Cohen D. Next-generation drug repurposing using human genetics and network biology. *Curr Opin Pharmacol.* 2020;51:78-92.
53. Nguyen PA, Born DA, Deaton AM, Nioi P, Ward LD. Phenotypes associated with genes encoding drug targets are predictive of clinical trial side effects. *Nat Commun.* 2019;10(1):1579.
54. Jerome RN, Joly MM, Kennedy N, et al. Leveraging Human Genetics to Identify Safety Signals Prior to Drug Marketing Approval and Clinical Use. *Drug Saf.* 2020;43(6):567-582.
55. Gill D, Georgakis MK, Walker VM, et al. Mendelian randomization for studying the effects of perturbing drug targets. *Wellcome Open Res.* 2021;6:16.
56. Hingorani AD, Kuan V, Finan C, et al. Improving the odds of drug development success through human genomics: modelling study. *Sci Rep.* 2019;9(1):18911.
57. Schmidt AF, Finan C, Gordillo-Marañón M, et al. Genetic drug target validation using Mendelian randomization. *bioRxiv.* 2019:781039.
58. Nitsch D, Molokhia M, Smeeth L, DeStavola BL, Whittaker JC, Leon DA. Limits to causal inference based on Mendelian randomization: a comparison with randomized controlled trials. *Am J Epidemiol.* 2006;163(5):397-403.
59. Jones DS, Podolsky SH. The history and fate of the gold standard. *Lancet.* 2015;385(9977):1502-1503.
60. Gill D, Burgess S. The evolution of mendelian randomization for investigating drug effects. *PLoS Med.* 2022;19(2):e1003898.
61. Bennett DA, Holmes MV. Mendelian randomisation in cardiovascular research: an introduction for clinicians. *Heart.* 2017;103(18):1400-1407.

62. Heart Protection Study Collaborative G. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20,536 high-risk individuals: a randomised placebo-controlled trial. *Lancet.* 2002;360(9326):7-22.
63. Ridker PM. The JUPITER trial: results, controversies, and implications for prevention. *Circ Cardiovasc Qual Outcomes.* 2009;2(3):279-285.
64. Maher V, Sinfuego J, Chao P, Parekh J. Primary prevention of coronary heart disease. What has WOSCOPS told us and what questions remain? West Of Scotland Coronary Prevention Study. *Drugs.* 1997;54(1):1-8.
65. Armitage J, Bowman L, Wallendszus K, et al. Intensive lowering of LDL cholesterol with 80 mg versus 20 mg simvastatin daily in 12,064 survivors of myocardial infarction: a double-blind randomised trial. *Lancet.* 2010;376(9753):1658-1669.
66. Ference BA, Majeed F, Penumetcha R, Flack JM, Brook RD. Effect of naturally random allocation to lower low-density lipoprotein cholesterol on the risk of coronary heart disease mediated by polymorphisms in NPC1L1, HMGCR, or both: a 2 x 2 factorial Mendelian randomization study. *J Am Coll Cardiol.* 2015;65(15):1552-1561.
67. Swerdlow DI, Preiss D, Kuchenbaecker KB, et al. HMG-coenzyme A reductase inhibition, type 2 diabetes, and bodyweight: evidence from genetic analysis and randomised trials. *The Lancet.* 2015;385(9965):351-361.
68. Cannon CP, Giugliano RP, Blazing MA, et al. Rationale and design of IMPROVE-IT (IMProved Reduction of Outcomes: Vytorin Efficacy International Trial): Comparison of ezetimbe/simvastatin versus simvastatin monotherapy on cardiovascular outcomes in patients with acute coronary syndromes. *Am Heart J.* 2008;156(5):826-832.
69. Blazing MA, Giugliano RP, DeLemos J, et al. On-Treatment Analysis of the IMProved Reduction of Outcomes: Vytorin Efficacy International Trial (IMPROVE-IT). *Circulation.* 2014;130(23):2112-2112.
70. Burgess S, Thompson S, G. *Mendelian Randomization: Methods for Using Genetic Variants in Causal Estimation* CRC Press; 2015.
71. Burgess S. Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. *Int J Epidemiol.* 2014;43(3):922-929.
72. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature.* 2018;562(7726):203-209.
73. Scott RA, Scott LJ, Maegi R, et al. An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. *Diabetes.* 2017;66(11):2888-2902.
74. Locke AE, Kahali B, Berndt SI, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature.* 2015;518:197.
75. Knouff CW, Lim N, Song K, et al. Pharmacological effects of lipid-lowering drugs recapitulate with a larger amplitude the phenotypic effects of common variants within their target genes. *Pharmacogenet Genomics.* 2008;18(12):1051-1057.
76. Wurtz P, Wang Q, Soininen P, et al. Metabolomic Profiling of Statin Use and Genetic Inhibition of HMG-CoA Reductase. *J Am Coll Cardiol.* 2016;67(10):1200-1210.
77. Tragante V, Asselbergs FW, Swerdlow DI, et al. Harnessing publicly available genetic data to prioritize lipid modifying therapeutic targets for prevention of coronary heart disease based on dysglycemic risk. *Hum Genet.* 2016;135(5):453-467.
78. Coleman CI, Reinhart K, Kluger J, White CM. The effect of statins on the development of new-onset type 2 diabetes: a meta-analysis of randomized controlled trials. *Curr Med Res Opin.* 2008;24(5):1359-1362.
79. Rajpathak SN, Kumbhani DJ, Crandall J, Barzilai N, Alderman M, Ridker PM. Statin therapy and risk of developing type 2 diabetes: a meta-analysis. *Diabetes Care.* 2009;32(10):1924-1929.
80. Cai R, Yuan Y, Zhou Y, et al. Lower intensified target LDL-c level of statin therapy results in a higher risk of incident diabetes: a meta-analysis. *PLoS One.* 2014;9(8):e104922.

81. Baker WL, Talati R, White CM, Coleman CI. Differing effect of statins on insulin sensitivity in non-diabetics: a systematic review and meta-analysis. *Diabetes Res Clin Pract.* 2010;87(1):98-107.
82. Erqou S, Lee CC, Adler AI. Statins and glycaemic control in individuals with diabetes: a systematic review and meta-analysis. *Diabetologia.* 2014;57(12):2444-2452.
83. Cannon C. IMProved Reduction of Outcomes: Vytorin Efficacy International Trial - IMPROVE-IT. Paper presented at: American Heart Association Scientific Sessions; 17 November, 2014, 2014; American Heart Association Scientific Sessions.
84. Swerdlow DI, Preiss D. Genetic insights into statin-associated diabetes risk. *Current Opinion in Lipidology.* 2016;27(2):125-130.
85. Al-Goblan AS, Al-Alfi MA, Khan MZ. Mechanism linking diabetes mellitus and obesity. *Diabetes Metab Syndr Obes.* 2014;7:587-591.
86. Waters DD, Ho JE, Boekholdt SM, et al. Cardiovascular event reduction versus new-onset diabetes during atorvastatin therapy: effect of baseline risk factors for diabetes. *J Am Coll Cardiol.* 2013;61(2):148-152.
87. Ridker PM, Pradhan A, MacFadyen JG, Libby P, Glynn RJ. Cardiovascular benefits and diabetes risks of statin therapy in primary prevention: an analysis from the JUPITER trial. *The Lancet.* 2012;380(9841):565-571.
88. Allen N, Sudlow C, Downey P, et al. UK Biobank: Current status and what it means for epidemiology. *Health Policy Techn.* 2012;1(3):123-126.
89. Collins R. What makes UK Biobank special? *The Lancet.* 2012;379(9822):1173-1174.
90. Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *Am J Epidemiol.* 2017;186(9):1026-1034.
91. Health D. Uk Biobank begins recruiting using NHS data. https://www.digitalhealth.net/2007/09/uk-biobank-begins-recruiting-using-nhs-data/. Published 2007. Accessed 10.11.21, 2021.
92. Batty GD, Gale CR, Kivimäki M, Deary IJ, Bell S. Comparison of risk factor associations in UK Biobank against representative, general population based studies with conventional response rates: prospective cohort study and individual participant meta-analysis. *BMJ.* 2020;368:m131.
93. UK Biobank. *UK Biobank: Protocol for a large-scale prospective epidemiological resource. Protocol No: UKBB-PROT-09-06 (Main Phase)* 2007.
94. UK Biobank. *Verbal Interview stage Version 1.1.* 2012.
95. UK Biobank. *Body Composition Measurement.* 2011.
96. UK Biobank. *Hospital Episode Statistics data in Showcase. Version 1.0.* 2013.
97. UK Biobank. Uk Biobank ethics and governance framework. In: Biobank U, ed2007.
98. Fry D, Almond R, Moffat S, Gordon M, Singh P. *UK Biobank Biomarker Project Companion Document to Accompany Serum Biomarker Data.* 2019.
99. Elliott P, Peakman TC, Biobank UK. The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. *Int J Epidemiol.* 2008;37(2):234-244.
100. van Dijk CE, Verheij RA, Hansen J, et al. Primary care nurses: effects on secondary care referrals for diabetes. *BMC Health Serv Res.* 2010;10:230.
101. UK Biobank. *Mortality data: linkage to death registries* 2020.
102. Welsh S. *Genotyping of 500,000 UK Biobank participants. Description of sample processing workflow and preparation of DNA for genotyping.* 2017.
103. Bycroft C, Freeman C, Petkova D, et al. Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv.* 2017:166298.
104. Eastwood SV, Mathur R, Atkinson M, et al. Algorithms for the Capture and Adjudication of Prevalent and Incident Diabetes in UK Biobank. *PLoS One.* 2016;11(9):e0162388.

105. Pitts-Tucker T. Asian and Afro-Caribbean Britons have double the risk of type 2 diabetes. *BMJ.* 2012;345:e6135.
106. Chen J, Spracklen CN, Marenne G, et al. The trans-ancestral genomic architecture of glycemic traits. *Nat Genet.* 2021;53(6):840-860.
107. Wood AR, Esko T, Yang J, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet.* 2014;46(11):1173-1186.
108. Winkler TW, Justice AE, Graff M, et al. The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study. *PLoS Genet.* 2015;11(10):e1005378.
109. Brion MJA, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. *Int J Epidemiol.* 2013;42(5):1497-1501.
110. Liu DJ, Peloso GM, Yu H, et al. Exome-wide association study of plasma lipids in > 300,000 individuals. *Nature Genetics.* 2017;49(12):1758-+.
111. Iliadou A, Snieder H, Wang X, Treiber FA, Davis CL. Heritabilities of lipids in young European American and African American twins. *Twin Res Hum Genet.* 2005;8(5):492-498.
112. Liu H, Wang W, Zhang C, et al. Heritability and Genome-Wide Association Study of Plasma Cholesterol in Chinese Adult Twins. *Front Endocrinol (Lausanne).* 2018;9:677.
113. Ference BA, Kastelein JJP, Ginsberg HN, et al. Association of Genetic Variants Related to CETP Inhibitors and Statins With Lipoprotein Levels and Cardiovascular Risk. *JAMA.* 2017;318(10):947-956.
114. Hoffmann TJ, Theusch E, Haldar T, et al. A large electronic-health-record-based genome-wide study of serum lipids. *Nat Genet.* 2018;50(3):401-413.
115. Sniderman AD. Differential response of cholesterol and particle measures of atherogenic lipoproteins to LDL-lowering therapy: implications for clinical practice. *J Clin Lipidol.* 2008;2(1):36-42.
116. Sniderman AD, Thanassoulis G, Glavinovic T, et al. Apolipoprotein B Particles and Cardiovascular Disease: A Narrative Review. *JAMA Cardiol.* 2019;4(12):1287-1295.
117. Langlois MR, Nordestgaard BG, Langsted A, et al. Quantifying atherogenic lipoproteins for lipid-lowering strategies: consensus-based recommendations from EAS and EFLM. *Clin Chem Lab Med.* 2020;58(4):496-517.
118. Sniderman A, Langlois M, Cobbaert C. Update on apolipoprotein B. *Curr Opin Lipidol.* 2021;32(4):226-230.
119. White J, Swerdlow DI, Preiss D, et al. Lipid Fractions and Contrasting Risks of Coronary Artery Disease and Diabetes. *JAMA Cardiology.* 2016;1(6):692–699.
120. Valdes-Marquez E, Parish S, Clarke R, et al. Relative effects of LDL-C on ischemic stroke and coronary disease: A Mendelian randomization study. *Neurology.* 2019;92(11):e1176-e1187.
121. Baigent C, Keech A, Kearney PM, et al. Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins. *Lancet.* 2005;366(9493):1267-1278.
122. Peloso GM, Auer PL, Bis JC, et al. Association of Low-Frequency and Rare Coding-Sequence Variants with Blood Lipids and Coronary Heart Disease in 56,000 Whites and Blacks. *Am J Hum Genet.* 2014;94(2):223-232.
123. Hopewell JC, Parish S, Offer A, et al. Impact of common genetic variation on response to simvastatin therapy among 18 705 participants in the Heart Protection Study. *Eur Heart J.* 2013;34(13):982-992.
124. McCaw ZR, Lane JM, Saxena R, Redline S, Lin X. Operating characteristics of the rank-based inverse normal transformation for quantitative trait analysis in genome-wide association studies. *Biometrics.* 2020;76(4):1262-1272.
125. Sofer T, Zheng X, Gogarten SM, et al. A fully adjusted two-stage procedure for rank-normalization in genetic association studies. *Genet Epidemiol.* 2019;43(3):263-275.

126. Swerdlow DI, Preiss D, Kuchenbaecker KB, et al. HMG-coenzyme A reductase inhibition, type 2 diabetes, and bodyweight: evidence from genetic analysis and randomised trials. *Lancet.* 2015;385:351–361.
127. Lotta LA, Sharp SJ, Burgess S, et al. Association Between Low-Density Lipoprotein Cholesterol-Lowering Genetic Variants and Risk of Type 2 Diabetes: A Meta-analysis. *JAMA.* 2016;316(13):1383-1391.
128. Regitz-Zagrosek V, Kararigas G. Mechanistic Pathways of Sex Differences in Cardiovascular Disease. *Physiol Rev.* 2017;97(1):1-37.
129. Logue J, Walker JJ, Colhoun HM, et al. Do men develop type 2 diabetes at lower body mass indices than women? *Diabetologia.* 2011;54(12):3003-3006.
130. Power ML, Schulkin J. Sex differences in fat storage, fat metabolism, and the health risks from obesity: possible evolutionary origins. *Br J Nutr.* 2008;99(5):931-940.
131. Kautzky-Willer A, Harreiter J, Pacini G. Sex and Gender Differences in Risk, Pathophysiology and Complications of Type 2 Diabetes Mellitus. *Endocr Rev.* 2016;37(3):278-316.
132. Peters SAE, Woodward M. Sex Differences in the Burden and Complications of Diabetes. *Curr Diab Rep.* 2018;18(6):33.
133. Moebus S, Göres L, Lösch C, Jöckel K-H. Impact of time since last caloric intake on blood glucose levels. *Eur J Epidemiol.* 2011;26(9):719-728.
134. Leong A, Daya N, Porneala B, et al. Prediction of Type 2 Diabetes by Hemoglobin A(1c) in Two Community-Based Cohorts. *Diabetes care.* 2018;41(1):60-68.
135. Emdin CA, Khera AV, Natarajan P, et al. Genetic Association of Waist-to-Hip Ratio With Cardiometabolic Traits, Type 2 Diabetes, and Coronary Heart Disease. *JAMA.* 2017;317(6):626-634.
136. Lotta LA, Wittemans LBL, Zuber V, et al. Association of Genetic Variants Related to Gluteofemoral vs Abdominal Fat Distribution With Type 2 Diabetes, Coronary Disease, and Cardiovascular Risk Factors. *JAMA.* 2018;320(24):2553-2563.
137. CTTC. Effect of statin therapy on muscle symptoms: an individual participant data meta-analysis of large-scale, randomised, double-blind trials. *Lancet.* 2022;400(10355):832-845.
138. Haines MS, Leong A, Porneala BC, Meigs JB, Miller KK. Association between muscle mass and diabetes prevalence independent of body fat distribution in adults under 50 years old. *Nutrition & Diabetes.* 2022;12(1):29.
139. Bower JK, Meadows RJ, Foster MC, Foraker RE, Shoben AB. The Association of Percent Body Fat and Lean Mass With HbA(1c) in US Adults. *J Endocr Soc.* 2017;1(6):600-608.
140. Preiss D, Seshasai SRK, Welsh P, et al. Risk of Incident Diabetes With Intensive-Dose Compared With Moderate-Dose Statin Therapy A Meta-analysis. *JAMA.* 2011;305(24).
141. Kohli P, Waters D, Nemr R, Arsenault B, Messing M. Risk of New-Onset Diabetes and Cardiovascular Risk Reduction From High-Dose Statin Therapy in Pre-Diabetics and Non–Pre-Diabetics: An Analysis From TNT and IDEAL. *Journal of the American College of Cardiology.* 2015;65(4):402-404.
142. Burke JF, Sussman JB, Kent DM, Hayward RA. Three simple rules to ensure reasonably credible subgroup analyses. *BMJ.* 2015;351:h5651.
143. Hippisley-Cox J, Coupland C. Development and validation of QDiabetes-2018 risk prediction algorithm to estimate future risk of type 2 diabetes: cohort study. *BMJ (Clinical research ed).* 2017(359).
144. Langenberg C, Lotta LA. Genomic insights into the causes of type 2 diabetes. *Lancet.* 2018;391(10138):2463-2474.
145. Morris AP. Progress in defining the genetic contribution to type 2 diabetes susceptibility. *Curr Opin Genet Dev.* 2018;50:41-51.
146. Hippisley-Cox J, Coupland C, John R, Aziz S, Peter B. Predicting risk of type 2 diabetes in England and Wales: prospective derivation and validation of QDScore. *BMJ (Clinical research ed).* 2009(338).

147. Coscia C, Gill D, Benítez R, Pérez T, Malats N, Burgess S. Avoiding collider bias in Mendelian randomization when performing stratified analyses. *Eur J Epidemiol.* 2022;37(7):671-682.
148. Holmberg MJ, Andersen LW. Collider Bias. *JAMA.* 2022;327(13):1282-1283.
149. Lai LC. Global standardisation of HbA1c. *Malays J Pathol.* 2008;30(2):67-71.
150. Feller S, Boeing H, Pischon T. Body mass index, waist circumference, and the risk of type 2 diabetes mellitus: implications for routine clinical practice. *Dtsch Arztebl Int.* 2010;107(26):470-476.
151. Katzmann JL, Sorio-Vilela F, Dornstauder E, et al. Non-statin lipid-lowering therapy over time in very-high-risk patients: effectiveness of fixed-dose statin/ezetimibe compared to separate pill combination on LDL-C. *Clin Res Cardiol.* 2022;111(3):243-252.
152. Ambegaonkar BM, Tipping D, Polis AB, Tomassini JE, Tershakovec AM. Achieving goal lipid levels with ezetimibe plus statin add-on or switch therapy compared with doubling the statin dose. A pooled analysis. *Atherosclerosis.* 2014;237(2):829-837.
153. Lee J, Egolum U, Parihar H, Cooley M, Ling H. Effect of Ezetimibe Added to High-Intensity Statin Therapy on Low-Density Lipoprotein Cholesterol Levels: A Meta-Analysis. *Cardiol Res.* 2021;12(2):98-108.
154. Myocardial Infarction Genetics Consortium I, Stitziel NO, Won HH, et al. Inactivating mutations in NPC1L1 and protection from coronary heart disease. *N Engl J Med.* 2014;371(22):2072-2082.
155. Wu H, Shang H, Wu J. Effect of ezetimibe on glycemic control: a systematic review and meta-analysis of randomized controlled trials. *Endocrine.* 2018;60(2):229-239.
156. Lee ES, Seo HJ, Back SS, et al. Transcriptional regulation of Niemann-Pick C1-like 1 gene by liver receptor homolog-1. *BMB Rep.* 2015;48(9):513-518.
157. Scott RA, Scott LJ, Magi R, et al. An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. *Diabetes.* 2017;66(11):2888-2902.
158. Klimentidis YC, Arora A, Newell M, et al. Phenotypic and Genetic Characterization of Lower LDL Cholesterol and Increased Type 2 Diabetes Risk in the UK Biobank. *Diabetes.* 2020;69(10):2194-2205.
159. Burgess S, Thompson SG. Bias in causal estimates from Mendelian randomization studies with weak instruments. *Stat Med.* 2011;30(11):1312-1323.
160. Casula M, Mozzanica F, Scotti L, et al. Statin use and risk of new-onset diabetes: A meta-analysis of observational studies. *Nutr Metab Cardiovasc Dis.* 2017;27(5):396-406.
161. Takeshita Y, Takamura T, Honda M, et al. The effects of ezetimibe on non-alcoholic fatty liver disease and glucose metabolism: a randomised controlled trial. *Diabetologia.* 2014;57(5):878-890.
162. Park H, Shima T, Yamaguchi K, et al. Efficacy of long-term ezetimibe therapy in patients with nonalcoholic fatty liver disease. *J Gastroenterol.* 2011;46(1):101-107.
163. Takase H, Dohi Y, Okado T, Hashimoto T, Goto Y, Kimura G. Effects of ezetimibe on visceral fat in the metabolic syndrome: a randomised controlled study. *Eur J Clin Invest.* 2012;42(12):1287-1294.
164. Cho Y, Kim R-H, Park H, Wang HJ, Lee H, Kang ES. Effect of Ezetimibe on Glucose Metabolism and Inflammatory Markers in Adipose Tissue. *Biomedicines.* 2020;8(11):512.
165. Young KG, McDonald TJ, Shields BM. Glycated haemoglobin measurements from UK Biobank are different to those in linked primary care records: implications for combining biochemistry data from research studies and routine clinical care. *medRxiv.* 2022:2021.2011.2020.21266511.
166. Wheeler E, Leong A, Liu CT, et al. Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. *PLoS Med.* 2017;14(9):e1002383.
167. Arman Q, Alexander KM. Incidence of New-Onset Diabetes in the IMPROVE-IT Trial: Does Adding Ezetimibe to Simvastatin Increase Risk Compared to Simvastatin Alone? Americal

College of Cardiology. Latest in cardiology Web site. https://www.acc.org/latest-in-cardiology/articles/2015/11/20/14/20/incidence-of-new-onset-diabetes-in-the-improve-it-trial. Published 2015. Accessed 03.01.2022, 2015.
168. Ridker PM, Danielson E, Fonseca FA, et al. Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. *N Engl J Med.* 2008;359(21):2195-2207.
169. Grundy SM, Stone NJ, Bailey AL, et al. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA Guideline on the Management of Blood Cholesterol: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation.* 2019;139(25):e1082-e1143.
170. Schmidt AF, Swerdlow DI, Holmes MV, et al. PCSK9 genetic variants and risk of type 2 diabetes: a mendelian randomisation study. *Lancet Diabetes Endocrinol.* 2017;5(2):97-105.
171. Sabatine MS, Leiter LA, Wiviott SD, et al. Cardiovascular safety and efficacy of the PCSK9 inhibitor evolocumab in patients with and without diabetes and the effect of evolocumab on glycaemia and risk of new-onset diabetes: a prespecified analysis of the FOURIER randomised controlled trial. *Lancet Diabetes Endocrinol.* 2017;5(12):941-950.
172. Schwartz GG, Steg PG, Szarek M, et al. Alirocumab and Cardiovascular Outcomes after Acute Coronary Syndrome. *New England Journal of Medicine.* 2018;379(22):2097-2107.
173. Cholesterol Treatment Trialists' (CTT) Collaboration. Protocol for a prospective collaborative overview of all current and planned randomized trials of cholesterol treatment regimens. *Am J Cardiol.* 1995;75(16):1130-1134.
174. Cholesterol Treatment Trialists C. Protocol for analyses of adverse event data from randomized controlled trials of statin therapy. *Am Heart J.* 2016;176:63-69.
175. Sarsenbayeva A, Jui BN, Fanni G, et al. Impaired HMG-CoA Reductase Activity Caused by Genetic Variants or Statin Exposure: Impact on Human Adipose Tissue, beta-Cells and Metabolome. *Metabolites.* 2021;11(9).
176. Swerdlow DI, Preiss D. Genetic insights into statin-associated diabetes risk. *Curr Opin Lipidol.* 2016;27(2):125-130.
177. Wu H, Forgetta V, Zhou S, Bhatnagar SR, Pare G, Richards JB. Polygenic Risk Score for Low-Density Lipoprotein Cholesterol Is Associated With Risk of Ischemic Heart Disease and Enriches for Individuals With Familial Hypercholesterolemia. *Circ Genom Precis Med.* 2021;14(1):e003106.
178. Silverman MG, Ference BA, Im K, et al. Association Between Lowering LDL-C and Cardiovascular Risk Reduction Among Different Therapeutic Interventions: A Systematic Review and Meta-analysis. *JAMA.* 2016;316(12):1289-1297.
179. Ushijima S, Matsumaru N, Tsukamoto K. Evaluation of Drug Lags in Development Initiation, New Drug Application and Approval Between Japan and the USA and the Impact of Local Versus Multi-regional Clinical Trials. *Pharmaceut Med.* 2021;35(4):253-260.
180. DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: New estimates of R&D costs. *J Health Econ.* 2016;47:20-33.
181. Dowden H, Munro J. Trends in clinical success rates and therapeutic focus. *Nat Rev Drug Discov.* 2019;18(7):495-496.
182. Arrowsmith J, Miller P. Trial watch: phase II and phase III attrition rates 2011-2012. *Nat Rev Drug Discov.* 2013;12(8):569.
183. Kaitin KI, DiMasi JA. Pharmaceutical innovation in the 21st century: new drug approvals in the first decade, 2000-2009. *Clin Pharmacol Ther.* 2011;89(2):183-188.
184. Waring MJ, Arrowsmith J, Leach AR, et al. An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nat Rev Drug Discov.* 2015;14(7):475-486.
185. Kaitin KI. Deconstructing the drug development process: the new face of innovation. *Clin Pharmacol Ther.* 2010;87(3):356-361.
186. Kola I. The state of innovation in drug development. *Clin Pharmacol Ther.* 2008;83(2):227-230.

187. Earl J, Kirkpatrick P. Fresh from the pipeline. Ezetimibe. *Nat Rev Drug Discov.* 2003;2(2):97-98.
188. compendium Tem. Summaries of Product Characteristics. Ezetrol 10mg Tablets. https://www.medicines.org.uk/emc/product/9109/smpc#. Published 2021. Accessed.
189. Littlejohns TJ, Holliday J, Gibson LM, et al. The UK Biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions. *Nat Commun.* 2020;11(1):2624.
190. Ahsan F, Oliveri F, Goud HK, et al. Pleiotropic Effects of Statins in the Light of Non-Alcoholic Fatty Liver Disease and Non-Alcoholic Steatohepatitis. *Cureus.* 2020;12(9):e10446.
191. Bhatt HB, Smith RJ. Fatty liver disease in diabetes mellitus. *Hepatobiliary Surg Nutr.* 2015;4(2):101-108.
192. Sigler MA, Congdon L, Edwards KL. An Evidence-Based Review of Statin Use in Patients With Nonalcoholic Fatty Liver Disease. *Clin Med Insights Gastroenterol.* 2018;11:1179552218787502.
193. Paul J, Shihaz AVH. Pancreatic Steatosis: A New Diagnosis and Therapeutic Challenge in Gastroenterology. *Arq Gastroenterol.* 2020;57(2):216-220.